Shaping speech patterns via predictability and recoverability

by

James Doh Yeon Whang

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

New York University

September 2017

Frans Adriaans

Lisa Davidson

Acknowledgments[1]

While writing this acknowledgments section, I found myself returning to the Sino-Korean word 인복 (人福) /inpok/, which roughly translates to “blessings in the form of support and inspiration from friends and acquaintances, which one receives regardless of his deservedness”. I had very little, if any, control over the kinds of people I met over the years, and despite that, I have come to be surrounded by wonderful and lovely people. It would be impossible to fully express my gratitude to everyone, so I will limit myself to those who were in my life during my time at NYU and to just the highlights.

My advisors: It goes without saying that I have become the linguist I am today largely because of my advisors, Lisa Davidson and Frans Adriaans. Professionally, I learned from Lisa the fundamentals of experimental research and especially the importance of spotting implicit assumptions in my own work as well as others’. From Frans, I learned the skills necessary for computational research and also how to break down research questions into smaller, more comprehensible chunks. On a more personal level, it was always a pleasure to speak with Lisa and Frans, not just about research but life in general. I learned from them that my work should not be all-consuming, which helped me enjoy research all the more.

[1]This dissertation is based upon work supported by the National Science Foundation under Grant No. BCS-1524133.

I would like to thank Shigeto Kawahara, whom I consider my third advisor. I doubt there was ever another outside committee member as dedicated as Shigeto, and I cannot thank him enough for his generosity. I would also like to thank Kie Zuraw, who encouraged me to pursue a doctorate in linguistics. It is thanks to Kie that I started graduate school at all.

NYU faculty: Besides my advisors, I would like to thank the rest of my committee members, Gillian Gallagher and Maria Gouskova, who provided me with invaluable feedback on my work. Before this dissertation, Gillian also helped me take my first crack at academic writing as her coauthor. I am grateful for her steady guidance, which made the experience less terrifying. My very first class at NYU was Phonology with Maria, and it was Maria who taught me how to read academic papers. The reading method I learned from Maria made me a more efficient reader, and it also made me a better and hopefully more comprehensible writer.

I would also like to thank Anna Szabolcsi for her "expert contribution to the advancement of phonetics", which she left for me in the fridge, in the form of an apple pie. Her delicious contribution is fully responsible for at least a chapter in this dissertation.

Friends and cohort: When I decided to come to NYU for graduate school, I had no idea I would end up making such wonderful friends. I would like to thank in particular Dylan Bumford, Itamar Kastner, and Vera Zu for their friendship. It is sad to think that we will be strewn all across the globe after sharing an office together for five years. I miss you all.

Family: I would like to thank Clara, my lovely wife. You have been my anchor through all the ups and downs, and I am eternally grateful for your unwavering love and support. Lastly, none of this would have been possible if it were not for my parents. You raised me to have an inquisitive mind and encouraged my pursuit of knowledge. I dedicate this work to you.

Abstract

Recoverability refers to the ease of recovering the underlying form—stored mental representations—given a surface form—actual, variable output signals (e.g., [ðæt˺, ðæt^{h}] → /ðæt/ ‘that’). Recovery can be achieved from phonetic cues explicitly present in the acoustic signal or through prediction from the context. However, recoverability can be compromised when the information in the signal is insufficient or the predictability in a given context is not high enough. The purpose of this dissertation is to investigate through experiments and computational modeling how phonotactic predictability in a given context affects (i) the amount of information present in the signal during production, (ii) the attention paid to the information during perception, and (iii) how reliance on phonotactics and phonetic cues might be learned. The language in focus is Japanese, a language well-known for its CV phonotactic preference, and the process of high vowel reduction, which often results in consonant clusters that violate this phonotactic restriction. The results suggest that language users prioritize predictability during speech processing, with phonetic-cue interpretation applying when predictability is not reliable enough.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.0 Introductory remarks

The primary goal of this dissertation is to investigate how signal-based and knowledge-based mechanisms interact during speech production and perception through experiments and computational modeling. For the purposes of this dissertation, signal-based mechanisms refer to processes that affect low-level phonetic implementation in the acoustic signal during production and perception. Knowledge-based mechanisms encompass lexical knowledge as well as sublexical grammatical knowledge such as syllable structure and phonotactic probability. This dissertation focuses mainly on phonotactic predictability of a target segment and how phonotactic knowledge affects the production and perception of the target segment's phonetic cues. Japanese high vowel reduction is used as the test case for this interaction, because only one high vowel is phonotactically legal in certain environments, whereas two are allowed in others, resulting in different levels of phonotactic

predictability. I also present a computational model that investigates the mechanisms that might be at play to yield the observed empirical results.

Before proceeding, the terms *lexical* and *sublexical* should be defined. Throughout the dissertation, *lexical* is used to refer to the levels of representation that correspond to words and processes that involve word-sized units, such as lexical competition (Vitevitch and Luce, 1998, 1999). In contrast, *sublexical*[1] is used to refer to lower levels of representation and processes that involve units that are smaller than words, such as phonotactics (Daland and Pierrehumbert, 2011), phoneme inventories (Näätänen et al., 1997; Wedel, 2012), as well as low-level phonetic cues (Mattys et al., 2005).

1.1 Language-specific phonetic processes

1.1.1 Articulation

While there are universal factors that constrain phonological processes stemming from the limits of human speech production and perception systems, a phonological grammar is necessarily language-specific, involving language-specific phonetic implementations. For example, the voice onset times (VOT) of velar obstruents are generally longer than alveolar obstruents due to physiological and aerodynamic factors (Stevens, 1998). Velar stops are produced by creating a constriction near the soft palate with the tongue dorsum. Alveolars on the other hand are produced by creating a constriction in the alveolar region with the tongue tip. The tongue body is heavier and consequently slower moving than the tongue tip, and coupled with the fact that the constriction is being made in the soft palate, the result is a longer contact between articulators in velar stops than in alveolars. Furthermore, the longer channel of airflow created by a velar constriction causes the articulators to

[1] Although not the sense in which the term is used here, *sublexical* is also used to refer to phonological processes that occur in subsets of a lexicon (Becker and Gouskova, 2016).

be sucked together with more force than in alveolars due to the Bernoulli effect, further delaying the voice onset time of velars.

The slowness of the tongue body and the Bernoulli effect are universal factors that affect all human speech, but the longer VOT in velars than in alveolars is not a universal property of all sound systems. In Dahalo, the VOT of velar stops are in fact significantly shorter than alveolar stops by 15 ms (27 ms vs. 42 ms; Cho and Ladefoged, 1999). In the same 1999 study, Cho and Ladefoged also show that languages that contrast similar classes of sounds can still have different targets for VOT. For example, the western dialect of Aleut has longer VOT across all places of articulation compared to the eastern dialect, and the VOT of aspirated stops in Navajo are nearly twice as long as the aspirated stops in Jalapa Mazatec. Cross-linguistic differences have also been reported in other classes of sounds (/j, w/; Maddieson and Emmorey, 1985), tone distribution (Zhang, 2004), and even the resting position of articulators (Gick et al., 2004). These studies show that although there are universal factors that affect low-level phonetic details, languages can choose targets that are specific to the language.

1.1.2 Coarticulation

Given that languages can choose different phonetic implementations of speech segments, it is unsurprising that languages also differ on how segments are coarticulated and also that listeners of different languages are sensitive to different phonetic cues. This language-specific coarticulation and how they are modulated by higher-level knowledge of the target language during production and perception are the focus of this dissertation.

Coarticulation is a universal process in speech production where a phonological segment becomes more like an adjacent or nearby segment (Farnetani and Recasens, 2010), such as when an oral vowel preceding a nasal consonant becomes nasalized (e.g., /pæn/ $\rightarrow$ [pæ̰̃n] ‘pan’). Although some anticipatory nasalization of vowels preceding nasal consonants is unavoidable, the degree

of nasalization varies from language to language. Take English and French for example. French phonemically contrasts nasal and oral vowels, and thus the degree of anticipatory nasalization of oral vowels preceding a nasal consonant (i.e., /CVN/ sequence) is tightly controlled to begin late in the vowel, to preserve the oral-nasal contrast (e.g., /CVN/ vs. /CṼ/). English, on the other hand, does not have the same contrast, and allows nasalization to begin early in the oral vowels (Cohn, 1993; see also Beddor et al., 2013).

The degree of coarticulation, however, does not always follow from contrast preservation. Spanish, like English, does not phonemically contrast nasal and oral vowels, but pre-nasal vowels are consistently less nasalized than in English (Solé, 1992), presumably because unlike in the case of English where the lack of contrast is regarded as a reason for less control over nasalization, Spanish speakers simply choose to preserve the gestural targets for oral vowels across all contexts. Solé further argues that English speakers may be using vowel nasalization phonologically to explicitly signal an upcoming nasal consonant. The proposal seems to be supported by Lahiri and Marslen-Wilson (1991), who found that the presence of nasality during a vowel is interpreted as signaling an upcoming nasal consonant by English listeners (i.e., [CṼN]), but as a nasal vowel in a non-nasal context by Bengali listeners (i.e., [CṼC]), a language that has a nasal-oral contrast like French.

The above example of nasalization is of anticipatory coarticulation from an immediately following segment. Coarticulation, however, can also be carried over from a preceding segment (progressive coarticulation) and result from segments further away. For example, in an EPG (electropalatography) study that investigates the coarticulation of /V_1kV_2/ sequences in Catalan, English, French, German, Italian, and Swedish, Gibbon et al. (1993) found that the constriction location for /k/ shifted more towards V_2 for Catalan, French, Italian, and Swedish speakers, showing a tendency towards anticipatory coarticulation with a vowel two segments away (i.e., /ikla/ → [i$\underline{\text{k}}$la] and /akli/ → [a$\underline{\text{k}}_{+}$li]). English and German speakers, on the other hand, showed more coarticulation with V_1, a tendency towards progressive coarticulation (i.e., /ikla/ → [i$\underline{\text{k}}_{+}$la] and /akli/ → [a$\underline{\text{k}}$li]). Furthermore, the phasing of the /k/ and /l/ in the cluster differed across the languages. Catalan speakers showed

the most overlap, with the lateral gesture beginning simultaneously or even before the release of /k/. Swedish speakers showed the least overlap, where the lateral gesture never began until after the release of /k/. The authors propose that the extreme degree of gestural overlap in Catalan speakers may stem from the fact that /l/ is more velarized in Catalan, making coarticulation with the velar stop less effortful. To summarize, coarticulation can vary from language to language depending on a given language's gestural targets for the segments being coarticulated (e.g., Catalan vs. Swedish /l/), but also for higher-level reasons such as contrast preservation (e.g., French vs. English vowel nasalization) or gestural target preservation (e.g., Spanish vs. English vowel nasalization).

Different production targets mean that the phonetic cues in the resulting acoustic signal also differ, making it necessary for perceptual systems to also be language-specific. In other words, listeners of different languages are sensitive to different phonetic cues based on their first language (L1) experience. For example, Korean coda obstruents are obligatorily unreleased while they are optionally released in English (Kang, 2003). So in order to recover the place of a coda obstruent in a sequence such as /mɛt/, it is more important for the Korean listener to pay attention to formant transitional cues from [ɛ] to [t̚] than for the the English listener, who has the option of waiting for the release of the coda obstruent. Hume et al. (1999) found that Korean listeners are indeed more sensitive to V-to-C transitional cues than English listeners.

While listeners are sensitive to phonetic cues that are contrastive in their native language, conversely they are also often insensitive to phonetic cues that are not contrastive in their native language. For example, French listeners have difficulty contrasting short versus long vowels (Dupoux et al., 1999), English listeners have difficulty perceiving tonal contrasts (So and Best, 2010), Japanese listeners have difficulty contrasting /l/ versus /ɾ/ because neither are phonemes of the language (Flege et al., 1996), and the list goes on.[2]

[2]This is not to say that perception is entirely language-specific. For example, although Japanese listeners are unable to accurately categorize /l, r/ in isolation, they nevertheless seem to compensate for the coarticulatory effects of /l, r/ on adjacent segments, much like English listeners (Mann, 1986).

1.1.3 Recoverability

Recoverability is a notion that refers to the ease in which the underlying form—stored mental representations—can be accurately accessed from a given surface form—actual, variable output signals (e.g., [ðæt̚, ðætʰ] → /ðæt/ 'that'; Chitoran et al., 2002; Mattingly, 1981; McCarthy, 1999). Recoverability is inherently gradient, since a target linguistic unit can be less recoverable for various reasons without it being absolutely unrecoverable. Likewise a unit can be more recoverable without it becoming absolutely recoverable. Previous works on recoverability often utilize the framework of Articulatory Phonology (Browman and Goldstein, 1989, 1992a,b) because it captures gradience more efficiently than frameworks that operate on SPE-style features. For example, returning to the example of anticipatory nasalization of oral vowels, the process can be presented as the following SPE-style rule:

[+vocalic, -nasal] → [+nasal] / __ [+nasal]

But what of cases where the vowel is only partially nasalized? Instead of features, Articulatory Phonology treats the laryngeal and supra-laryngeal gestural targets associated with a segment as primitives on which phonological processes operate. Gradient phonological changes result in part because although there are spatial and temporal gestural targets, the targets can be modified both spatially (e.g., reduction) or temporally (e.g., coarticulation) as needed by language-specific details. The gestural score representation of Articulatory Phonology allows varying degrees of coarticulation to be represented rather elegantly, as shown below in Figure 1.1 for vowel nasalization.

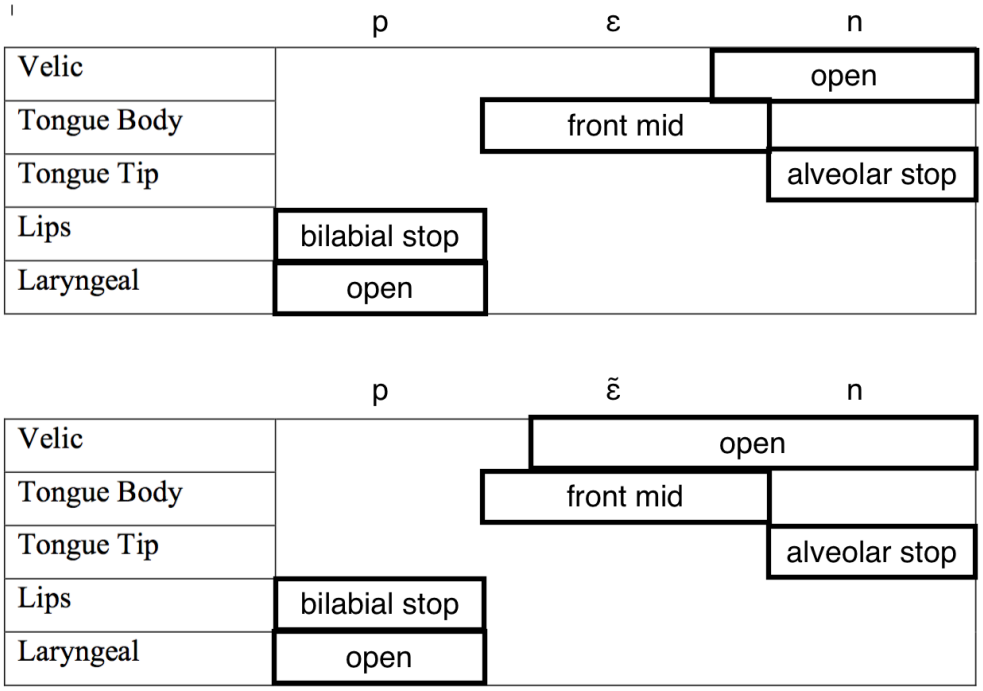

Figure 1.1: Gestural scores of /pɛn/.

Recoverability is also often synonymous with perceptibility, under the view that gestures are phased in order to make linguistic units more perceptible, and hence more recoverable. For example, Bladon (1986) argues that pre-aspirated stops are less common cross-linguistically than post-aspirated stops because the sharp rise in amplitude that results from aspiration following a stop closure elicits greater activation of auditory neurons than the comparatively gradual decrease in amplitude of aspiration that precedes a stop closure, making post-aspiration more perceptually salient (i.e., easily recoverable). This also means that making pre-aspiration as perceptible as post-aspiration requires more effort for the speaker, making pre-aspiration a less efficient and hence less recoverable class of sounds.

1.2 Recoverability and predictability

Selective sensitivity to phonetic cues in acoustic signals shows that recoverability is affected by expectations stemming from the listener's language experience. Rather than paying attention to all phonetic cues in all contexts, language experience allows listeners to devote their cognitive resources to fewer environments and to fewer phonetic cues, conserving effort while maximizing recognition of relevant phonetic cues. This interplay between effort conservation and maximization of recognition has influenced a diverse line of phonological research, including works on lexical access (Marslen-Wilson, 1987; Norris, 1994; Vitevitch and Luce, 1999), influence of perceptual bias on phonological systems (Boersma, 1998; Hayes et al., 2004), as well as influence of articulatory bias on phonological systems (Archangeli and Pulleyblank, 1994; Browman and Goldstein, 2000). Since these works all have a bearing on the notion of recoverability, they are discussed in more detail below.

1.2.1 Recovery of lexical and nonce words

Previous works on spoken word recognition have shown that lexical and sublexical processes contribute independently to recoverability, where they sometimes conflict with each other. First, both lexical and sublexical (phonotactic) knowledge seem to aid recoverability in noise. For example, Brown and Hildum (1956) presented participants with monosyllabic lexical items, phonotactically legal nonce words, and phonotactically illegal nonce words, testing how accurately these items are identified in noise. The results showed that lexical items were identified most accurately, followed by phonotactically legal nonce words, then illegal nonce words, suggesting that lexical knowledge facilitates recoverability. Higher recovery rate of phonotactically legal nonce words additionally suggests that higher phonotactic probability aids recoverability as well. Facilitatory effects of phonotactic probability were also shown in a study by Pitt and McQueen (1998), where participants

exhibited a bias towards identifying phonetically ambiguous segments as segments with higher phonotactic probability.

Second, although lexical items and phonotactically legal sequences are recovered more easily, a number of studies have also shown that phonotactic probability has contradictory effects in lexical and nonce words. For example, high phonotactic probability in lexical items means that there are also that many more similar lexical items (i.e., neighbors), and numerous models of spoken word recognition such as Cohort (Marslen-Wilson, 1987), TRACE (McClelland and Elman, 1986), and Shortlist (Norris, 1994) have proposed that lexical activation is inhibited as lexical neighborhood density rises, slowing down lexical access (Vitevitch and Luce, 1998). By contrast, nonce words seem to be processed more quickly when they consist of sequences with high phonotactic probability (Vitevitch et al., 1997; Vitevitch and Luce, 1999). The question then is, which process takes precedence? The answer to this question seems to depend on the task. In general, listeners seem to prioritize the use of lexical knowledge, relying on their sublexical phonotactic grammar only when lexical activation fails. In other words, a phonotactic grammar is useful mostly for nonce word processing (Shademan, 2006; Vitevitch and Luce, 1999).

Additionally, a study by Mattys et al. (2005) investigated whether participants pay more attention to lexical and sublexical (segmental and prosodic) segmentation cues when they are in conflict. The results showed again that lexical cues are prioritized, and that participants rely on sublexical cues when lexical information cannot be accessed due to noise or absence. This dissertation adds to this line of work by investigating the interaction between two sublexical cues, namely phonotactic predictability and very fine-grained phonetic cues, via a production experiment in Chapter 3 and a perception experiment in Chapter 4. Chapter 5 then presents a preliminary computational model that investigates how the various levels of lexical and sublexical knowledge are learned and combined into a grammar.

1.2.2 Lexical effects on recoverability

Recent works on exemplar-based approaches to phonology have noted the critical role of the lexicon on sound change (Bybee, 2006; Ernestus, 2011; Pierrehumbert, 2001), proposing that it is often the most frequent lexical items that are targeted for reduction because they require fewer cues for recovery. For example, the common English phrase "going to" often functions as a single unit, and is targeted for reduction as in the contracted form "gonna". This reduction of "going to" is sensitive to lexical considerations, as it only applies to cases of volitional use. Therefore, "Are you gonna (going to) eat?" is legal, but "Are you gonna (going to) the store?" is not. The lexicalization of the reduced/contracted form perhaps is further supported by the fact that it can be reduced even further as in "I am going to" → "I'm gonna" → "I'ma". Highly frequent lexical items can reduce, often significantly, because they are more predictable and by extension recoverable without the aid of extra phonetic cues.

Building on this line of research, recent work by Hall et al. (forthcoming) argues that lexical considerations can be applied more broadly to phonological research. The thrust of their argument is that phonological systems tend to reduce segments in predictable and/or perceptually weak positions because enhancing weak cues would require additional effort while contributing little to successful lexical access. For example, word-final codas are neutralized for two reasons. First, during lexical access, segments become more predictable as the listener processes more and more of the target item. This means that word-final codas contribute less to identifying the target lexical item. Second, codas are perceptibly weaker than onset consonants. Rather than enhancing the weak cues of an already predictable segment, phonological systems choose to enhance cues of segments in perceptually strong positions instead. Conversely, it is the phonetic cues of segments in unpredictable and/or perceptually strong positions that are enhanced, such as the aspiration of word-initial obstruents in English (e..g, /pik/ → [p^{h}ik] vs. /spik/ → [spik]).

1.2.3 Sublexical effects on recoverability

Although Hall et al. (forthcoming) explicitly argue that lexical predictability is crucial to account for the reduction and enhancement of phonetic cues, there are a number of works that also suggest that phonetic cues can be enhanced or reduced due to sublexical considerations. Silverman (1997) investigates the sound systems of a wide range of languages and reports that when a sound system contains seemingly inefficient contrasts, it is to preserve other efficient contrasts. For example, Mazatec is a language that uses aspiration contrastively with unaspirated, post-aspirated, and pre-aspirated stops in its phoneme inventory. The language additionally has a complex vowel system with modal, nasal, breathy, and creaky vowels. With the observation that breathy vowels often start out breathy then become modal towards the end, surfacing approximately as [a̤a], Silverman notes that sequences such as [ta̤a] where unaspirated stops are followed by (partially) breathy vowels are unattested in the language, presumably because it is perceptually too similar to attested sequences such as [t^{h}a], where post-aspirated stops are followed by modal vowels. Following the argument of Bladon (1986), Silverman proposes that the attested sequences like [t^{h}a] are preferred because post-aspiration yields maximal perceptibility of the laryngeal abduction that corresponds to aspiration. Silverman further argues that the vocalic laryngeal contrasts in Mazatec make it necessary for consonantal laryngeal contrasts to be realized in other ways. Pre-aspirated stops, although dispreferred cross-linguistically due to their articulatory inefficiency, is the way the full range of consonantal laryngeal contrasts are retained.

Adding to the discussion of sublexical effects on phonetic implementation are Chitoran et al. (2002), who compare the gestural overlap of stop clusters in Georgian using EMMA (electromagnetic midsagittal articulometer). They test specifically for the effects of position in word and the ordering of place of articulation on degree of gestural overlap. One thing to note here is that although increased gestural overlap of CV sequences would increase the efficiency of cue transmission which also aids recoverability, CC gestural overlap would actually decrease recoverability (Browman and

Goldstein, 2000). For example, in a stop C_1C_2 cluster, C_2 closure would mask the release of C_1, which is what often happens in American English in words such as /ækt/ $\rightarrow$ [ækˀt] 'act'. The results from Chitoran et al. (2002) showed a decrease in overlap in word-initial clusters, and also more overlap when the stop clusters are ordered front-to-back in terms of place (e.g, /pt, pk, tk/) compared to the clusters ordered back-to-front (e.g, /tp, kp, kt/). Both of these results are interpreted to show that gestural overlap is decreased when recoverability is at risk – word-intially because the only cues for C_1 comes from its release, and for a similar reason in back-to-front ordering of place since early closure of C_2 in a sequence like /tp/ would mask the release of C_1.

Preference for maximally perceptible cues (Silverman, 1997) and enhancement of word-initial obstruent clusters (Chitoran et al., 2002) are both compatible with the lexical account proposed by Hall et al. (forthcoming) as well. Perceptually robust cues aid lexical access, and likewise, word-initial position is precisely where segments are less predictable. However, the fact that speakers are sensitive to place when coordinating the gestures for consonant clusters clearly shows that there are phonetic/sublexical considerations at play.

1.2.4 Summary

To summarize, manipulation of low-level phonetic cues during production and perception is affected by a variety of higher level processes including lexical knowledge (Bybee, 2006; Hall et al., forthcoming), but also sublexical knowledge of a language's entire sound system (Silverman, 1997) and ordering effects in a word or cluster (Chitoran et al., 2002). It was also shown that when lexical and sublexical information are in conflict with each other, language users prioritize lexical information unless lexical information is absent or inaccessible (Mattys et al., 2005). Adding to the large body of work, this dissertation investigates how two types of sublexical information—phonotactic predictability and fine-grained coarticulatory cues—affect speech processing. If phonotactics has a similar "top-down" effect on low-level phonetic cues, the expected finding is that when phonotactic

predictability is low, speakers should encode more phonetic cues in the acoustic signal during production to aid the recovery of a target segment. Likewise, listeners should show heightened sensitivity to the encoded phonetic cues in environments where phonotactic predictability is low, but effectively ignore the available phonetic cues in environments where phonotactic predictability alone is sufficient for recovery of the target segment. Stated differently, when sublexical contextual cues and phonetic cues are in conflict, listeners should rely more on the contextual cues, except when the context is insufficient for successful recovery of the target segment.

1.3 Japanese high vowel reduction and recoverability

Japanese high vowel reduction is an ideal test case for the interplay of phonotactic and phonetic cues during production and perception. First, in the case of production, Japanese high vowel reduction is a highly productive process that is essentially categorical in most environments. Additionally, the environments in which high vowel reduction occurs have differing levels of phonotactic predictability because only one of two high vowels (i.e., /i, u/) are phonotactically legal after certain consonants, while both are legal after others. These two properties of the reduction process allow for relatively simple control over the effects of context and phonotactic predictability on the degree of high vowel reduction. Second, in the case of perception, the high vowels that are targeted for reduction during production also happen to be the same vowels that are often epenthesized between consonant clusters in loanword adaptations as well as online perception tasks. While the processes of reduction during production and epenthesis during perception are well-documented, the contextual effects of phonotactic predictability on both phenomena are relatively poorly understood, an issue which this dissertation aims to remedy.

1.3.1 Production

High vowel reduction is considered to be an integral feature of standard modern Japanese (Imai, 2010), so much so that dictionaries with explicit instructions on the reducing environments exist (Kindaichi, 1995; NHK, 1985). The phenomenon is commonly described as involving high vowels /i/ and /u/, which are "devoiced" in C_1VC_2 sequences when they are (i) unaccented and (ii) both C_1 and C_2 are voiceless obstruents. For example, while the /u/ in /kúʃi/ 'free use' and /kuʃi/ 'skewer' are both between two voiceless obstruents, only /kuʃi/ 'skewer' undergoes devoicing because the vowel is unaccented. Likewise, the /u/ is unaccented in both /kuki/ 'stem' and /kugi/ 'nail', but only /kuki/ 'stem' undergoes devoicing because the /u/ is flanked by two voiceless stops, namely /k/. High vowels also undergo devoicing at the ends of words when preceded by a voiceless fricative, but this case will not be discussed for the purposes of this dissertation.

Of all the voiceless consonants that can precede a high vowel in Japanese – [p, k, t͡s, ʧ, ɸ, s, ʃ, ç] – the literature on Japanese high vowel reduction focuses primarily on [k, ʃ], after which both high vowels can follow. Varden (2010) states that since certain obstruents can only occur before one of the two high vowels (in non-loanwords; see Table 3.1 below), the vowel is entirely predictable after these obstruents, and thus can be deleted altogether with little impact on recoverability. In other words, both [ɸu] with a voiced vowel and [ɸu̥] with a devoiced/deleted vowel will be analyzed as /hu/, since [ɸ] only occurs as an allophone of /h/ preceding /u/.

		i	u
Unpredictable	p	✓	✓
	k	✓	✓
	ʧ	✓	✓
	ʃ	✓	✓
Predictable	ts	–	✓
	ɸ	–	✓
	s	–	✓
	ç	✓	–

Table 1.1: Voiceless obstruents of Japanese and possible following vowels. "–" means that the vowel is not phonologically possible in this context.

What Varden is arguing essentially is that recoverability-based processes also take into account phonotactic predictability. If it is true that the degree of high vowel reduction is dependent on the predictability of the vowel in a given context, this also means that if predictability from context alone is not reliable enough, the remaining reduced vowel should retain sufficient acoustic cues to disambiguate between the possible vowels. However, whether high vowels in contexts with differing levels of predictability show corresponding degrees of reduction in Japanese has not been tested, and this is the focus of Chapter 3.

1.3.2 Pilot study

I conducted a pilot experiment to determine what factors play the largest role in influencing the amount and type of cues available in the acoustic signal of reduced high vowels in Japanese. There is debate regarding how much high vowels reduce (i.e., devoicing vs. deletion) stemming in part from conflicting definitions of the terms. In the pilot study, *devoicing* was defined as the loss of voicing but retention of coarticulatory cues of the target vowel that color the burst/frication noise of C_1, and *deletion* was defined as the loss of both voicing and coarticulatory cues. The major findings of the pilot study were that the predictability of a vowel determines the degree of reduction.

Data for the pilot study were obtained from eight native Japanese speakers (four women and four men) who were born and raised in the Tokyo area but are currently residing in New York City as students. The participants were recorded in a sound-attenuated booth reading 40 lexical tokens (20 target and 20 control) embedded in meaningful and unique carrier sentences. Target and control tokens consisted of native and Sino-Japanese words and were divided into 10 stop-stop and 10 fricative-stop C_1-C_2 combinations. For all tokens, C_1 was /k, ʃ/, after which either high vowel can occur, or /ɸ, s, ç/, after which only one of the two is possible. The bilabial stop /p/ was not investigated in the pilot study due to its extreme rareness in the lexicon outside of onomatopoetic words and loanwords. Furthermore, the affricates [tʃ, t͡s], which are allophones of /t/ before [i, u] respectively, were also not included for the sake of simplicity. C_2 was /k, t/ for target tokens and /b, d, g/ for control tokens.

Center of gravity (COG), the amplitude weighted mean of frequencies present in a signal (Forrest et al., 1988), was measured for the first half (COG1) and the second half (COG2) of C_1 burst/frication noise to determine the degree of coarticulation with a following vowel. The prediction here was that greater coarticulation with a vowel would lower the amplitude of higher frequencies due to a weakening of the oral constriction. COG results revealed that the cues available in the acoustic signal depend on the predictability (i.e., ease of recovery) of the vowel in a given context. For [ɸ, s, ç], after which the high vowel is highly predictable, COG2 was significantly lower than COG1 when an unreduced vowel followed, suggesting an increased overlap towards the end of C_1. No such effect was found for reduced tokens, however, suggesting that there is no intervening oral vowel gesture between C_1 and C_2. In other words, the vowel seems to have deleted, leaving no cue for the vowel because any such cues are not necessary for recovering an already predictable vowel. In contrast, for /k, ʃ/ after which the vowel is less predictable, COG1 was significantly lower for both reduced and unreduced tokens when the vowel was /u/ regardless of reduction, and this difference was also evident in COG2. Being a back vowel, /u/ has a larger front oral cavity than /i/ and thus has a lowering effect on high-frequency burst/frication noises. The significant effect of vowel type

suggests that the vowels did not delete but devoiced, where the retained vowel gesture overlaps with and colors the burst/frication noise of C_1. Complete deletion of the vowel in these cases would jeopardize the recoverability of the vowel. By devoicing the vowel instead (i.e., reducing the glottal gesture), maximal recoverability is obtained while also minimizing gestural effort. The results are consistent with Chitoran et al. (2002) and Silverman (1997) who argued that overlap of gestures are coordinated in order to preserve recoverability.

Although the results show that Japanese speakers may in fact produce cluster-like sequences as a result of vowel deletion in certain environments, it must be noted that the participants in this study were native speakers of Japanese currently studying in the United States. L2 experience has been shown to significantly affect L1 production even in the short term (Chang, 2010), and since the participants have had extended exposure to English, a language that allows a large number of consonant clusters, they may have been more inclined to produce more cluster-like sequences than monolingual speakers. The production experiment in Chapter 3 therefore revisits this issue with expanded data from monolingual speakers recorded in Japan.

1.3.3 Perception

Because both phonotactic knowledge and phonetic implementation are highly attuned to be language-specific, listeners often make errors when processing L2 speech. I discussed previously that listeners are generally very sensitive to phonetic cues that are contrastive in their native language. Sometimes, this sensitivity can lead to a phenomenon called illusory epenthesis, where L1 listeners report perceiving something that is actually not present in the acoustic signal in contexts where the absence of that element would be phonotactically illegal. For example, when English listeners are presented with a phonotactically illegal sequence such as the word initial cluster in a nonce word like /ptamo/, the /p/ burst is interpreted as signaling the presence of a schwa (Davidson and Shaw, 2012). Similarly, Spanish speakers report perceiving a prothetic /e/ before word-initial /sC/ clusters,

which are prohibited in Spanish (Hallé et al., 2008). Japanese listeners also report perceiving a vowel between phonotactically illegal consonant clusters, even when presented with stimuli that did not originally have an intervening vowel (Dehaene-Lambertz et al., 2000; Dupoux et al., 1999, 2011). This is taken to be the result of a strong CVCV preference in Japanese, where the phonotactic knowledge biases Japanese listeners towards epenthesizing a vowel in C_1-C_2 sequences.

Some crucial questions remain unanswered in the studies by Dupoux and colleagues, however, which are addressed in Chapter 4 via a perception experiment. First, it is unclear whether the illusory vowel perceived by Japanese listeners is driven by phonotactic violations as the authors argue or phonetic cues that were not strictly controlled. Beckman and Shoji (1984) showed that Japanese speakers indeed are sensitive to such coarticulatory information. The results from Dupoux et al. (2011) in particular also show that when vowel coarticulatory cues are in conflict with the default epenthetic vowel, Japanese speakers show sensitivity to the acoustic information while Brazilian Portuguese speakers showed significantly less sensitivity. Although the reason for this difference in sensitivity is not discussed in detail, a likely explanation is experience with high vowel reduction. Brazilian Portuguese lacks a phonological high vowel reduction process, and thus coarticulatory cues would be underutilized in Brazilian Portuguese in comparison to Japanese. Native German listeners have also been shown to be insensitive to coarticulation of devoiced vowels on C_1 burst/frication noise because German also does not have vowel devoicing (Zimmerer et al., 2013).

Second, these studies were not concerned with how high vowel reduction may affect perception. High vowel reduction rarely occurs between two voiced obstruents. A survey of the stimuli used in the studies reveal that obstruent combinations that can condition high vowel reduction in Japanese were mixed with non-reducing combinations (e.g., [ebdo] and [epto]; Dupoux et al., 2011). If it is the case that high vowels are reduced to the point of complete deletion in certain environments but not others, are Japanese listeners more inclined to "recover" the reduced vowel in these environments than others, and thus perceive an illusory vowel? A recent study by Hsieh

(2013) found that reducing environments have little effect on high rates of illusory vowel epenthesis when the C_1 is /k, p/. However, these are both low-predictability environments where vowels are expected to devoice rather than delete, so it is still unclear as to whether the lack of effect was really due to phonotactic repair or hypersensitivity to acoustic cues that drive false recovery.

1.3.4 Summary

Although how reduced high vowels are manifested acoustically in Japanese is still debated, there seems to be some agreement that the degree of reduction can range from simple loss of voicing to complete deletion. My pilot study tested the idea that the variation between devoicing and deletion is conditioned by the vowel's recoverability in a given context (Varden, 2010). When both /i, u/ are possible after a given voiceless obstruent, the vowel is devoiced, leaving coarticulatory cues that aid recovery. When only one of the high vowels can occur, and thus predictability is high, the vowel is deleted since coarticulatory cues are unnecessary for successful recovery. However, since the participants in my pilot study were L2 speakers of English residing in the United States, their tendency to delete vowels in certain cases may have been the result of familiarity with clusters in English. Whether monolingual speakers of Japanese also tend to delete vowels in high-predictability contexts at the expense of violating a CVCV preference remains to be answered. This will be the focus of Chapter 3. Furthermore, while Japanese speakers seem to have difficulties perceiving clusters accurately, whether experience with high vowel reduction allows more accurate perception of certain clusters over others remains unanswered. This will be the focus of Chapter 4.

1.4 Modeling Japanese high vowel reduction

In Chapter 5 of this dissertation, I present a computational model that is trained on the Corpus of Spontaneous Japanese to investigate what mechanisms might be necessary to reproduce the production and perception results of Chapters 3 and 4. While reduction processes are often unin-

tended phonetic consequences of fast or casual speech, where the articulators do not have enough time to reach the spatial gestural targets, Japanese high vowel reduction is a phonological process, where voicing is categorically controlled. Although there are phonetic aspects of Japanese high vowel reduction as well, the model focuses on the phonological aspect of the process, combining a statistically-driven phonotactic learning mechanism and a lexically-driven alternation rule learning mechanism, to test the claims that phonotactic knowledge plays a lesser role once a lexicon of sufficient size is acquired and that lexical alternation rules can fine-tune phonotactic knowledge (Pater and Tessier, 2003; Tesar and Prince, 2007). The model assumes that phonotactic learning occurs in tandem with lexical acquisition, and that the two processes inform each other. The learning process can be represented as the box chart in Figure 1.2 below. Given training data, the lexical learning mechanism induces input-to-output conversion rules based on alternations detected in the lexicon. Simultaneously, the phonotactic learning mechanism calculates the observed/expected ratios for all biphones in the input, which are then used to induce phonotactic constraints. The conversion rules and constraints make up CON, which is used to evaluate the test data.

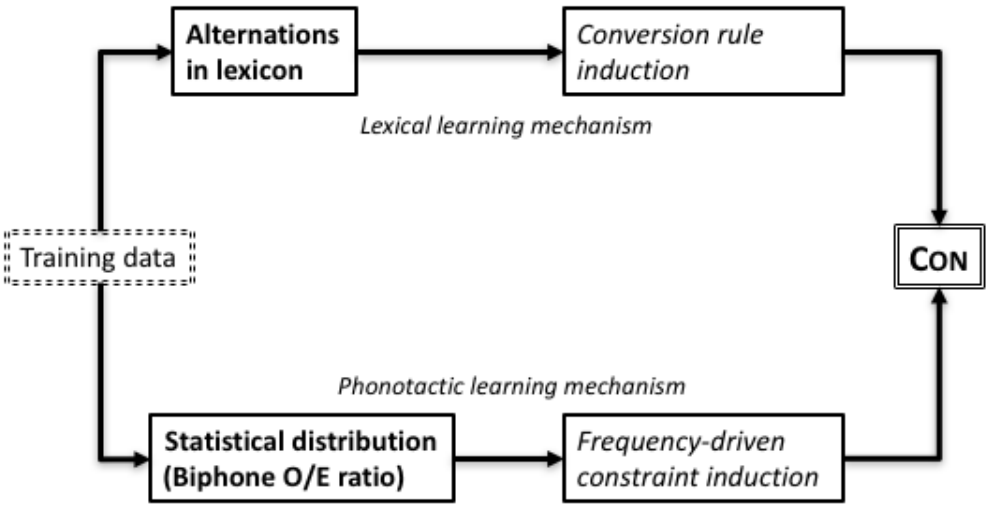

Figure 1.2: Learning mechanism of the statistical and lexical phonotactic model.

1.5 Outline of the dissertation

There are two experiments and one computational modeling component to this dissertation, with a chapter devoted to each. Both experiments and the computational model also utilize information from the Corpus of Spontaneous Japanese. Before presenting the experimental and computational works, presented in Chapter 2 is an overview of works that discuss the phonotactic structure of Japanese. Japanese is well-known for its strong CV preference, but high vowel reduction often results in consonant clusters that seemingly violate this very preference. I propose that Japanese phonotactics and high vowel reduction apply at different phonological levels, thereby reconciling the apparent conflict.

Presented in Chapter 3 is a production experiment that tests the effects of phonotactic predictability on how much high vowel coarticulatory information is preserved or enhanced in the acoustic signal. The experimental design is similar to the pilot experiment discussed above, but with monolingual participants and a larger, more balanced stimulus set. In addition to Yamato and Sino-Japanese stimuli in reducing and non-reducing contexts, English loanwords that originally contain *s*-initial consonant clusters are included in the stimuli to test whether explicit knowledge of underlying clusters affects the production of clusters that have presumably undergone high vowel reduction like similar non-loan items (e.g., /skiː/ → [skiː] 'ski' vs. /suki/ → [s̥ki] 'to like').

The second experiment is a perception experiment, which is presented in Chapter 4. The experiment asks the following question: if it is the case that the amount of phonetic cues of a reduced vowel are modulated depending on the predictability of the vowel, is it also the case that listeners are more attentive to the same high vowel cues in certain environments and not others? Since high vowel reduction requires Japanese listeners to recover reduced vowels in certain contexts in their native language, the perception experiment is designed to investigate how phonotactic predictability from the context and coarticulatory cues present in the acoustic signal interact when they are in conflict. For example, the stimulus [e̥sipo] would be perceived as /esipo/ based on the coarticulatory

cues of the medial [i̥], but based on phonotactic knowledge, the same stimulus would be perceived as /esupo/ instead because /u/ is the most common vowel after /s/ and /si/ is also a phonotactically illegal sequence in Japanese.

Chapter 5 presents a computational learner that attempts to model how the process of high vowel reduction and its effects on the perception of high vowels can be captured through data-driven constraint induction. The model is trained on data from the CSJ-Core, a richly annotated subset of the Corpus of Spontaneous Japanese, and builds a lexicon and learns alternation rules. The model also learns phonotactic constraints, learning biphone distributions strictly from the surface forms of segmented words. The main question the model aims to answer is how seemingly contradictory properties of a phonological grammar can be learned from the same input. More specifically, Japanese phonotactics has a strong preference for CVCV structure which biases listeners to perceive a vowel between clusters; but high vowel reduction is also a highly productive phonological process in the language that often results in consonant clusters.

Chapter 6 concludes the dissertation by summarizing the findings from Chapters 2 through 5 and making suggestions for future work.

CHAPTER 2

Phonological structure of Japanese

2.0 Introduction

It is often taken as given that Japanese has a strong CV preference. Evidence for this phonotactic preference is apparent in all lexical strata of Japanese, which is the focus of §2.1 below, as well as in numerous perception studies, which are discussed in Chapter 4. However, high vowels are known to often delete as a consequence of reduction, which seemingly violates the phonotactic preference. Whether the degree of high vowel reduction can be predicted is the focus of Chapter 3. The purpose of the current chapter instead is—under the assumption that high vowel reduction can lead to consonant clusters—to propose a way to reconcile this apparent conflict by separating the phonological levels in which phonotactic and reduction processes apply (§2.2).

2.1 Japanese phonotactics

Before proceeding, it should be noted that Japanese has a five-vowel system (i.e., /i, e, a, o, u/), and although the high back vowel falls somewhere between a high back unrounded vowel [ɯ] and a high central rounded vowel [ʉ] phonetically, the symbol for a high back rounded vowel will be used used throughout for the sake of simplicity.

As stated earlier, Japanese is well-known for having a strong surface preference for CV structure. Japanese surface forms can only contain two types of codas: the moraic nasal coda and the first half of a geminate. The nasal and geminate codas are traditionally represented as /N/ and /Q/, respectively, presumably to reflect their orthographic counterparts in Japanese (i.e., /N/ =<ん>; /Q/ =<っ>). Both coda types are argued to be underlyingly underspecified for place (Akamatsu, 1997; Kubozono, 2015), and they necessarily assimilate with [+consonantal] segments that follow. Examples are shown below in (1) and (2).

(1) /jamaNba/ → [jamamba] 'mountain witch'
/maNto/ → [manto] 'cloak'
/aNko/ → [aŋko] 'red bean paste'
/haNmaː/ → [hammaː] 'hammer'

(2) /kiQto/ → [kitto] 'surely'
/kaQpa/ → [kappa] 'a mythical water-dwelling monster'
/aQsari/ → [assari] 'simply'

When followed by a vowel, glide, or a word boundary, /N/ surfaces as nasalization on the preceding vowel and/or a uvular nasal (Okada, 1991) as shown in (3).

(3) /hoNja/ → [hõɴja] ‘bookstore’

/ʤuNi/ → [ʤũɴi] ‘rank’

/akikaN/ → [akikãɴ] ‘empty can’

Because /Q/ forms the first half of a geminate, it cannot occur before a vowel, a glide, or a word boundary, none of which can be geminated. Orthographically, /Q/ is written as <っ> and is often used in these environments to signify a glottal stop-like pause in utterance. Unlike /N/, however, the presence vs. absence of /Q/ is not contrastive. For example, <なに> [nani] and <なにっ> [naniʔ] both mean ‘what’.

The Japanese lexicon can be divided largely into four “strata” that have different historical sources (Ito and Mester, 1999; Moreton and Amano, 1999). The four strata are *Yamato* (or native), *Sino-Japanese*, *foreign*, and *mimetic*. The strata can be identified in three ways. First, morphemes from different strata seldom co-occur in compound formation (Shibatani, 1990). Second, the inventory of sounds differ across strata. For example, although vowel length is contrastive in all strata, [aː] occurs in the Foreign stratum but not in the Sino-Japanese stratum (Martin, 1952). Third, the phonology in each strata differ. For example, CV preference is observable across all strata, but when underlying morpheme concatenation results in CC sequences, the preferred repair strategy in the Yamato stratum is deletion or assimilation, while the preferred strategy in the Sino-Japanese stratum is epenthesis. This phonological difference is the focus of the following section.

Ito (1986, *et seq.*) proposed that the surface CV preference in Japanese can be captured by the following markedness constraints, which this dissertation also follows:

- NoCoda: assign a violation for every coda consonant in the surface form.
- Onset: assign a violation for every syllable that lacks an onset.
- CodaCondition: assign a violation for every coda segment that has its own place feature.
- *Complex: assign a violation for every tautosyllabic cluster in the surface form.

Of the four constraints above, only CODACONDITION and *COMPLEX are undominated in modern Japanese, and the effects of NOCODA and ONSET are most apparent in the Yamato stratum. The remainder of this section will provide examples of how the four phonotactic constraints shape the surface forms of Japanese, first focusing on Japanese verbal morphology from the Yamato stratum (§2.1.1), followed by patterns observed in Sino-Japanese roots and compounds (§2.1.2), and lastly, patterns observed in loanwords (§2.1.3).

2.1.1 Japanese verbal morphophonology

Japanese verbal stems can be divided largely into two groups: vowel-final (V-stem; *tabe-* 'eat') and consonant-final (C-stem; *nom-* 'drink'). V-stems always end in front vowels /i/ or /e/, while C-stems can end in /k, s, t, n, m, r, w, g, b/. Verbal suffixes also come in two varieties. The first is mnemonically called C/V-suffixes, which have C-initial and V-initial allomorphs, and the second is called T-suffixes, which always begin with /t/ or /d/.

For a C/V suffix, the choice of allomorph depends on whether the verbal stem it follows is a C-stem or a V-stem. Using the negation suffix *-ana, -na* as an example, the V-initial allomorph is chosen after C-stems (4a), and the C-initial allomorph is chosen after V-stems (4b).

(4) a. *hanas-ana* 'not speak'
aw-ana 'not meet'
nom-ana 'not drink'
kak-ana 'not write'
b. *tabe-na* 'not eat'
tari-na 'not suffice'

Choosing the wrong suffix allomorph results in a violation of ONSET (e.g., **tabe-ana*) or CODA-CONDITION (e.g.,**nom-na*). In the latter case, the surface form could be resyllabified as **no.m-na*

to satisfy both CODACONDITION and ONSET, but the resulting output violates *COMPLEX instead, an undominated constraint that prohibits tautosyllabic clusters.

T-suffixes have /t/-initial and /d/-initial allomorphs but no V-initial allomorphs. This is not problematic when T-suffixes follow V-stems (e.g., *tabe-ta* 'eat.PAST'; *tari-ta* 'suffice.PAST), but faithful realization of both C-stems and T-suffixes would result in a violation of NOCODA (e.g., **nom-da* 'drink.PAST') or *COMPLEX (e.g., **no-mda* 'drink.PAST'). When structural violations result from C-stem + T-suffix cases, the repair method of choice is stem allomorphy as shown in (5) below. /s/-final stems are the only exception to the stem allomorphy, where the repair strategy of choice is *i*-epenthesis instead as shown in (5a). [i] is not the epenthetic vowel of choice in other strata, as will be discussed in the following section. The full pattern of stem allomorphy in Japanese is rather complicated due to historical reasons, but stem allomorphy results in geminates (5b), place-assimilated nasal codas (5c), or diphthongs (5d). As mentioned above, geminates and nasal codas are the two coda types that are allowed in Japanese. Although they violate NOCODA, they conform to CODACONDITION, suggesting additionally that CODACONDITION is ranked higher than NOCODA in Japanese.

(5) a. Epenthesis (/s/ only)

/hana**s**-ta/ → [hana**ʃi**-ta] 'speak.PAST'

/ke**s**-ta/ → [ke**ʃi**-ta] 'erase.PAST'

b. Geminate coda

/a**w**-ta/ → [a**t**-ta] 'meet.PAST'

/ke**r**-ta/ → [ke**t**-ta] 'kick.PAST'

/u**t**-ta/ → [u**t**-ta] 'shoot.PAST'

c. Place-assimilated nasal coda

/no**m**-da/ → [no**n**-da] 'drink.PAST'

/jo**b**-da/ → [jo**n**-da] ‘call.PAST’

/si**n**-da/ → [ʃi**n**-da] ‘die.PAST’

d. Diphthong formation

/ka**k**-ta/ → [ka͡i**i**-ta] ‘write.PAST’

/tu**g**-da/ → [tsu͡i**i**-da] ‘pour.PAST’

Japanese verbal morphophonology shows that while the surface forms of Japanese obey the structural constraints NoCoda, Onset, CodaCondition, and *Complex, there are no such restrictions on the underlying forms. Additionally, underlying clusters are repaired through assimilation or allomorphy. The next section discusses Sino-Japanese roots, which also show that consonant clusters often occur in underlying forms, but are repaired at the surface level using epenthesis, a repair strategy that is seldom used in the Yamato stratum.

2.1.2 Sino-Japanese compounds

Sino-Japanese (SJ) words make up roughly 60% of the Japanese lexicon and have a long history that stretches back to the 6th century. Despite their long history, the Sino-Japanese stratum exhibits phonological characteristics that are distinct from the native stratum (McCawley, 1968; Ito and Mester, 1996). SJ roots are maximally bimoraic and can take the following forms with the initial C being optional: CV, CVV, CVN, and CVCV. This restriction can be explained by the fact that these roots were originally monosyllabic in Chinese and were maximally CVC. Also, SJ roots rarely occur in isolation and most commonly form bimorphemic compounds. In compounds, CV and CVV roots, where VV is either a long vowel or a diphthong, already conform to the surface preference of Japanese and thus exhibit no surprising patterns. Some examples are shown in (6) below. For all examples, periods (.) indicate across-morpheme syllable boundaries, whereas dashes (-) indicate within-morpheme syllable boundaries. The first two examples also show that Onset violations can often result from concatenation of morphemes that lack consonants in the Sino-Japanese stratum.

(6) /u+i/ → [u.i] ‘signs of coming rain (rain+intent)’

/ʃuː+i/ → [ʃuː.i] ‘surroundings (around+area)’

/ki+ki/ → [ki.ki] ‘crisis (danger+chance)’

/e͡i+juː/ → [e͡i.juː] ‘hero (brilliant+magnificent)’

/soː+zoː/ → [soː.zoː] ‘imagination (thought+image)’

CVN roots all end in the moraic nasal coda /N/ and follow the patterns described in (1) and (3). The nasal coda assimilates in place before consonants (e.g., /haN+tai/ → [han.tai] ‘opposition (anti+comparison)’) or surfaces as nasalization on the preceding vowel and/or a uvular nasal elsewhere (e.g., /ʤuN+i/ → [ʤũɴ.i] ‘rank (order+level)’) .

More complex patterns are observed in CVCV roots. The second consonant (C_2) is always either /t/ (*t*-roots) or /k/ (*k*-roots), and the second vowel (V_2) is always a high vowel. The vowel is almost always a high back vowel with a handful of exceptions, where the the vowel is a front high vowel instead. Due to the predictable nature of V_2, the vowel is argued to be epenthetic and not underlyingly represented as part of the root (e.g., /tok/ ‘special’ and /ket/ ‘tie up’; Ito, 1986; Tateishi, 1989) with CVCV allomorphs listed only for cases where a choice between /u, i/ is available (e.g., /sit, siti, *situ/ ‘seven’ vs. /sit, situ, *siti/ ‘loss’; Ito and Mester, 2015; Kurisu, 2001). As high vowel epenthesis is a productive repair strategy in loanwords as well, this dissertation also assumes that V_2 in Sino-Japanese CVCV roots is generally epenthetic and will refer to these roots as CVC roots hereafter.

In compounds with an initial *t*-root, the final /t/ of the root behaves like /Q/ before voiceless obstruents, fully assimilating with the following onset to form a geminate, as shown in (7) below.

(7) /sit+teN/ → [ʃit.tẽɴ][1] ‘loss of points (loss+point)’

/sit+hai/ → [ʃip.paiː][2] ‘failure (loss+defeat)’

/sit+koː/ → [ʃik.koː] ‘lapse (loss+efficacy)’

/sit+siN/ → [ʃiʃ.ʃĩɴ] ‘to faint (loss+spirit)’

On the other hand, *k*-roots only form geminates with another /k/.

(8) /kak+ko/ → [kak.ko] ‘firm/determined (certain+solid)’

/kak+toː/ → *[kat-toː], [ka-ku.toː] ‘definitive answer (certain+answer)’

/kak+ho/ → *[kap-po], [ka-ku.ho] ‘secure (certain+hold)’

In all of the geminating cases, NoCoda is violated but CodaCondition is obeyed.

When CVC roots occur in environments where the final consonant cannot form a geminate (e.g., word-finally, prevocalically, before voiced obstruents, etc.), a high vowel is epenthesized because CodaCondition prohibits codas with independent place features. For *t*-roots, the epenthesized vowel is always [u] with only a handful of exceptions.[3]

(9) /kiN+hat/ → [kim.pa-ts**u**] ‘blond (gold+hair)’

/sit+gen/ → [ʃi-ts**u**.gẽɴ] ‘improper remark (error+speech)’

(cf., /sit**i**+gatu/ → [si-ʧ**i**.gatsu] ‘July (seven+month)’)

/ket+i/ → [ke-ts**u**.i] ‘determination (tie up+will)’

For *k*-roots, the epenthesized vowel is [u] when V_1 is a non-front vowel, [i] when V_1 is /e/, and either [i, u] when V_1 is /i/ as shown in (10), (11), and (12), respectively.

[1] /s/ surfaces as [ʃ] before a high front vowel. Other obstruents that undergo allophony include /t/ → [ʧi, tsu], /d, z/ → [ʤi, dzu], and /h/ → [çi, ɸu]. Note that /d, z/ neutralize before high vowels.

[2] /h/ surfaces as [p] post-consonantally in Sino-Japanese words due to historical reasons.

[3] The only *t*-roots where the final vowel is not [u] are as follows: /iti/ ‘one’, /siti/ ‘seven’, /hati/ ‘eight’, /niti/ ‘sun’, /kiti/ ‘good luck’.

(10) /kak+ho/ → [ka-k**u**.ho] ‘secure (certain+hold)’

/tok+gi/ → [to-k**u**.gi] ‘specialty (special+skill)’

/bok+ʃi/ → [bo-k**u**.ʃi] ‘a minister (pastoral+teacher)’

(11) /sek+taN/ → [se-k**i**.tãɴ] ‘coal (rock+charcoal)’

/tek+i/ → [te-k**i**.i] ‘hostility (enemy+intent)’

(12) /diki/ → [ʤi-k**i**] ‘soon’

/diku/ → [ʤi-k**u**] ‘axle’

The analysis of Sino-Japanese compounds further illustrates a strong preference for CVCV structure. CVC roots, in particular, reveal that heterorganic clusters and word-final consonants that result from compounding are allowed underlyingly but are repaired on the surface. To summarize, violations of ONSET seem unproblematic in the Sino-Japanese stratum unlike the Yamato stratum. However, like the Yamato stratum, CODACONDITION must be obeyed in the Sino-Japanese stratum.

2.1.3 Loanword repairs

Roughly 80% of loanwords in modern Japanese comes from English (Sibata, 1994 (as cited in Kubozono, 2015), which often contain codas that are prohibited in Japanese. The most productive repair strategy for loanwords is vowel epenthesis and primarily high vowel epenthesis. The default epenthetic vowel in loanwords is [u] with three exceptions. First, in a handful *k*-final loanwords, the epenthetic vowel shows front vowel harmony much like Sino-Japanese *k*-roots. This is shown in the first three examples of (13) below. Second, the epenthetic vowel tends to be [i] after palatal consonants. Third, the epenthetic vowel is usually [o] after coronal stops, presumably to retain consonant identity because coronal stops in Japanese become affricated before high vowels (i.e., /ti, tu, di, du/ → [ʧi, tsu, ʤi, ʣu]; Kubozono, 2015). Regardless of the epenthetic vowel, the resulting

structure after repair is CVCV. With this in mind, shown below are examples of how NoCoda and CodaCondition violations are repaired.

(13) NoCoda violations:

Source	*Japanese*	*Gloss*
keɪ͡**k**	ke:**ki**	'cake'
stɹaɪ͡**k**	sutorai**ki**	'strike/protest'
stɹaɪ͡**k**	sutorai**ku**	'strike (baseball)'
pi**tʃ**	pi:**tʃi**	'peach'
eɪ͡**t**	ei**to**	'eight'
ʃe**f**	ʃe**ɸu**	'chef'
eɪ͡**s**	ei**su**	'ace'
tɹʌbə**l**	torabu**ru**	'trouble'
æ**m**nɛsti	a**mu**nesuti	'amnesty'

(14) CodaCondition violations

Source	*Japanese*	*Gloss*
nɛ**k**taɪ͡	ne**ku**tai	'necktie'
tæ**k**si	ta**ku**si	'taxi'
æ**t**ləs	a**to**rasu	'atlas'

Unlike Sino-Japanese roots, which are maximally CVC in the original Chinese, English loanwords often contain tautosyllabic clusters that violate *Complex as well, both in onset and coda positions. Again, such cases are repaired through vowel epenthesis.

(15) *Complex violations:

Source	*Japanese*	*Gloss*
tɹʌbəl	**to**raburu	'trouble'
kɹeɪ͡p	**ku**re:pu	'crepe'
plʌs	**pu**rasu	'plus'
staɪ	**su**ta:	'star/celebrity'
kɹʌ**st**	kura**su**to	'crust'
be**lt**	be**ru**to	'belt'
æ**lps**	a**rupus**u	'the Alps'

It should be noted that although vowel epenthesis is now the most productive repair strategy for loanwords, phonotactic violations were often repaired by deletion in older, pre- and early-20th century loanwords. Examples from Smith (2006) are shown in (16)-(18) below.

(16) NoCoda violations:

Source	*Japanese*	*Gloss*
pɑkɪ**t**ˀ	pokke∅	'pocket'
lɛmənei**d**ˀ	ramune∅	'lemonade'
ʤɪrəˑbʌ**g**ˀ	ʤiruba∅	'jitterbug'

(17) CodaCondition violations

Source	*Japanese*	*Gloss*
wa͡i**t**ˀʃɜˑt	wai∅ʃatsu	'white/dress shirt'
bi**f**ste͡ik	bi∅suteːki	'beefsteak'
hɛ**p**ˀbəˑn	he∅boɴ	'Hepburn'

(18) *Complex violations

Source	*Japanese*	*Gloss*
glɪsəɹin	∅**r**isuriɴ	'glycerine'
sɪmɜ**nt**ˀ	seme**ɴ**∅	'cement'
kɹæ**ŋk**	kara**ɴ**∅	'crank'

Regardless of the repair strategy, however, the examples in (13)-(18) above all show a strong preference for CVCV structure, allowing only nasal and geminate codas. The change in preferred repair strategy in loanwords presumably parallels a change in how borrowing occurred: perception-based older loans vs. orthography-based newer loans (Smith, 2006). Coda obstruents are optionally released in English, which would explain why the final obstruents in (16) were deleted, while the final consonants in words like 'white shirt' and 'beefsteak' in (17) show epenthesis instead.

2.1.4 Summary

This section provided evidence of CVCV preference in Japanese across Yamato, Sino-Japanese, and loanword strata. Of the four structural constraints—ONSET, NOCODA, *COMPLEX, CODACONDITION—only the last two were shown to be inviolable in Japanese. Although ONSET affects all strata of Japanese, its effects seemed relatively weak outside of the Yamato stratum, and NOCODA was violated as long as CODACONDITION could be obeyed in all strata. Stated plainly, the maximal syllabic structure allowed in Japanese is CVC, and the coda C cannot have an independent place feature. The next section will discuss the phenomenon of high vowel reduction in Japanese, which can result in surface consonant clusters that violate the very structures mandated by the structural constraints.

2.2 High vowel reduction

High vowel reduction (simply 'reduction' hereafter) is a highly productive, near-obligatory process that occurs over 90% of the time in most reducing contexts in standard modern Japanese. Reduction in Japanese is a popular topic of study, but there still are some disagreements regarding whether the process is phonetic (i.e., gradient and due to gestural mistiming) or phonological (i.e., categorical and planned). Currently, empirical evidence suggests that although high vowel reduction might have begun as a fast-speech, phonetic process (Hasegawa, 1979; Kuriyagawa and Sawashima, 1989), it has been phonologized in most contexts.

The focus of this dissertation is on the phonologized contexts, but it should be noted that there are two contexts, where high vowel reduction remains a phonetic process: between two fricatives (e.g., /su͡ʃi/ 'sushi') and between an affricate and a fricative (e.g., /t͡ʃi͡seː/ 'intelligence'). High vowels reduce much less often in these environments (Fujimoto, 2015; Tsuchida, 1997; Varden, 1998). Articulatory studies suggest that when reduction does occur in the two non-phonologized

environments, it is largely due to an unsuccessful attempt to close the glottis in time to phonate the vowel. In phonologized environments, speakers make no attempt to close the glottis, and the glottal opening during such phonological reduction is actually much wider than it normally is for voiceless consonants (Sawashima, 1971; Yoshioka et al., 1982).

The traditional description of reduction in Japanese is that high vowels become 'devoiced' when preceded by a voiceless obstruent and followed by a voiceless obstruent (e.g., /tukue/ → [tsu̥kue] 'desk'). Depending on the speaker, devoicing can also occur word finally at the end of an utterance (e.g., /kasi#/ → [kaʃi̥#] 'lyrics'). Although the prevalent assumption is that a high vowel that has undergone devoicing loses only its phonation while retaining its oral gestures (Faber and Vance, 2000; Jun and Beckman, 1993; Tsuchida et al., 1997; Varden, 2010), a number of studies have shown that high vowels in these environments can, in fact, lose both phonation and supralaryngeal gestures (Ogasawara and Warner, 2009; Pinto, 2015; Vance, 2008). This means that the process of reduction can lead to vowel deletion (e.g., /tukue/ → [tskue] 'desk') and potentially extremely long heterorganic clusters (e.g., /kikuti+saN/ → [k̥kʧsãɴ] 'Mr./Ms. Kikuchi'). The question of what factors determine the likelihood of vowel devoicing vs. deletion is the focus of chapter 3. What is important for the current discussion is that complete deletion as a consequence of reduction conflicts with the supposed strong CV phonotactic preference in Japanese.

2.2.1 Preliminary phonological analysis of high vowel reduction

As the name suggests, high vowel reduction/devoicing assumes that there is a high vowel that is targeted for the process. Reduction applies to both underlying and epenthesized high vowels, suggesting that high vowel reduction applies *after* phonotactic repairs are made, late in the grammar. This can be most clearly seen in Sino-Japanese roots. Take the two compounds in (19) below, for example. The word in (19a) begins with a CVC root, which does not have an underlying high vowel.

The word in (19b), on the other hand, begins with a CV root that does have an underlying high vowel. Regardless of their underlying status, however, the high back vowel is reduced in both words.

(19) a. /kak+toː/ → [kak̥toː] ‘definitive answer (certain+reply)’

b. /ku+toː/ → [k̥toː] ‘hard fight (difficult+battle)’

Furthermore, phonotactic constraints prohibit heterorganic clusters on the surface, but high vowel reduction can result in deleted vowels, creating surface clusters as in [kaktoː]. Allowing phonotactic constraints to evaluate the output post-deletion would result in yet another round of repairs through epenthesized vowels, but these vowels would then be targeted for deletion, and so on *ad infinitum*. The fact that reduced outputs are allowed in Japanese suggests that phonotactic constraints do not evaluate the output after reduction applies.

In order to reconcile Japanese phonotactics and high vowel reduction into a single framework, I propose that structural processes and phonetically-driven phonological processes (i.e., phonotactic restrictions and high vowel reduction, respectively) apply at different levels (Boersma, 2009; Hayes, 1999; Zsiga, 2000) and will assume a version of the representational levels proposed in the OT learnability literature (Apoussidou, 2007; Boersma, 1998 *et seq.*; Tesar and Smolensky, 1996) as shown in Figure 2.1 below. In this representation, the |lexical| level is where morphemes necessary for word formation are called upon, which yields the /underlying/ form, on which phonological grammars (of production) operate. The “surface” level is broken down into ⟨surface⟩, [auditory], [articulatory] levels. The ⟨surface⟩ form is where phonotactic evaluations occur and is equivalent to the output of a phonological grammar. The [auditory] and [articulatory] forms are phonetic forms, representing the acoustic signal and the articulatory implementations necessary to produce the auditory target, respectively. The auditory and articulatory forms together correspond roughly to the “overt” form in Tesar (1997), which is also the term I will use throughout this dissertation when a distinction between auditory and articulatory forms is unnecessary.

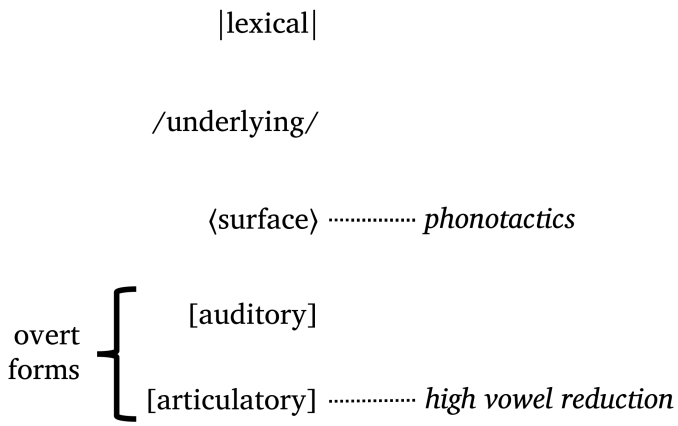

Figure 2.1: Separate levels for phonotactic and reduction processes.

I will address in more detail the theoretical motivations behind representing high vowel reduction in this way in Chapter 6 after all the relevant data are presented in the chapters to follow. For now, I would like to point out that since reduction applies at the overt/articulatory level, both underlying and epenthesized surface vowels can be targeted, and the data in (19) can be reanalyzed as in (20) below. In (20a), vowel epenthesis occurs at the ⟨surface⟩ level, which is then reduced at the [overt] level. In (20b), no epenthesis occurs because there are no phonotactic violations, and the underlying high vowel is reduced at the [overt] level.

(20) a. 'definitive answer (certain+reply)'

|kak+toː| → /kaktoː/ → ⟨ka.ku.toː⟩ →[kak̥toː, kak̚ toː]

b. 'hard fight (difficult+battle)'

|ku+toː| → /kutoː/ → ⟨ku.toː⟩ → [k̥utoː, k̚ toː]

2.2.2 Determining underlying vs. surface high vowels

Because high vowel reduction targets high vowels as represented at the surface level, it does not matter whether the high vowel was underlying or epenthesized. It is worth asking, however, how a Japanese learner might figure out that a reduced vowel was underlying or epenthetic and whether

such efforts are even necessary. In the following sections, I provide examples from each of the lexical strata of Japanese and argue that reducible vowels are all underlying in Yamato words, while reducible vowels can be either underlying or epenthesized depending on the root for Sino-Japanese words. Finally, I discuss why the underlying status of reducible vowels in loanwords might be speaker-dependent.

2.2.2.1 Rendaku and Yamato morphemes

This dissertation assumes that reducible high vowels in Yamato words are underlying without exception. Evidence for this assumption comes primarily from *rendaku*-induced alternations. *Rendaku* is a morphophonological phenomenon in Japanese, where, in a polymorphemic or reduplicated word, the initial voiceless consonant of a non-initial morpheme becomes voiced (Ito and Mester, 2003; Kawahara and Sano, 2016; Vance, 2015). Rendaku is relevant to the discussion of high vowel reduction in that reducing environments can be lost as a result of the process, providing cases of alternation between reduced and unreduced vowels in the same morpheme. Below are some examples of compounds and reduplicated words, where the second morpheme shows alternation as a result of rendaku. Reduced vowels are shown as deleted for the sake of simplicity.

morphemes		*in isolation*	*rendaku*	*gloss*
waru+kuti	(bad+mouth)	k_ʧi	waruguʧi	'insult'
siro+kisu'	(white+sillago)	k_su	ʃirogisu	'Japanese whiting (fish)'
baka+tikara	(absurd+strength)	ʧ_kara	bakaʤikara	'unbelievable strength'
ma+huta	(eye+lid)	ɸ_ta	mabuta[4]	'eyelid'
neko+sita	(cat+tongue)	ʃ_ta	nekoʤita	'inability to handle hot food'
tuki+tuki	(moon+moon)	ʦ_ki	ʦ_kiʣuki	'every month'
hito+hito	(person+person)	ç_to	ç_tobito	'people'
suki+suki	(like+like)	s_ki	s_kizuki	'matter of taste'

Rendaku is considered to be a lexical process that involves an autosegmental voicing feature (Ito and Mester, 2003). Rendaku is largely, although not exclusively, limited to the Yamato

[4] /h/ alternates with [b] as a consequence of rendaku due to a well-known historical change in Japanese.

stratum. Outside of the Yamato stratum, rendaku applies only to the most frequent and oldest (i.e., nativized) Sino-Japanese and loan morphemes. For example, the Sino-Japanese word *kaiʃa* 'company' undergoes rendaku in compounds like *kabuʃikigaiʃa* 'public company (stock+company)' and *juːgeŋgaiʃa* 'limited company (finite+company)' (Kindaichi, 1995). Likewise, the earliest loans in Japanese are of Portuguese origin, and old Portuguese loanwords like *karuta* 'playing cards' also undergo rendaku in compounds like *irohagaruta* and *hanagaruta*, both of which are types of Japanese card games (Vance, 2015).

Even in the Yamato stratum, rendaku applies rather irregularly. It can be blocked by a number of factors such as the presence of a voiced consonant elsewhere in the morpheme (Lyman's law) and the morphological structure of a compound (right-branch condition). In addition, rendaku has semantic consequences. For example, in the compound *jama+kawa* → *jamagawa* 'river in a mountain (mountain+river)', the initial morpheme modifies the second. However, *rendaku* is blocked in dvandva compounds like *jamakawa* 'mountains and rivers (mountain+river)', and consequently do not have the same subordinating relationship (Ito and Mester, 1986).

Being a lexical process, rendaku applies at the |lexical| level, resulting in underlying alternations of morphemes. Shown below in Figure 2.2 is how rendaku, phonotactics, and high vowel reduction fit into the representational levels shown above in Figure 2.1 using the reduplicated word 'every month':

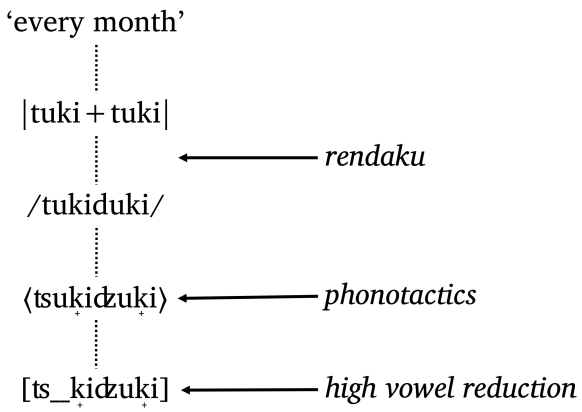

Figure 2.2: Putting it all together.

The reducible vowel must be present underlyingly for Yamato words for two reasons. First, unlike in the Sino-Japanese and loanword strata, phonotactic repair through epenthesis is dispreferred in the Yamato stratum. Second, even if underlying clusters were repaired through epenthesis, the derivation would still be wrong in many cases. Consider the examples in (21) below.

(21) a. ‘every month’

|tki+tki| → /tkidki/ → *⟨tokidoki⟩ → *[tokidoki] (cf., [ts_kidzuki])

b. ‘people’

|hto+hto| → /htobto/ → *⟨ɸutobuto⟩ → *[ɸ_tobuto] (cf., [ç_tobito])

In (21a), the derivation is wrong because the epenthetic vowel of choice after coronal stops is [o]. Even if the default epenthetic vowel [u] were used, cases like (21b) would still be problematic. In order to derive the correct output, reducible vowels in Yamato words must be lexical/underlying.

2.2.2.2 Sino-Japanese roots

The issue of determining whether a reducible vowel is underlying or epenthetic is rather straightforward for CV and CVC Sino-Japanese roots. CVV and CVN roots are irrelevant because long vowels

do not reduce, and a coda nasal creates a non-reducing environment. For CV roots, the vowel must be underlying because there is no restriction on what the vowel could be in these roots, and the choice of vowel would otherwise be unrecoverable. The situation is the opposite for CVC roots, since CVC roots alternate with their CVCV forms in a predictable manner as discussed in §2.1.2 .

2.2.2.3 Loanwords

For loanwords, it is actually unclear whether a reducible vowel is underlying or not. Diachronically speaking, a vowel is underlying if the vowel is present in the source word and epenthetic if the vowel is not. For example, in the loanword /piːʧi/ 'peach', the first vowel /iː/ is presumably underlying since there is a corresponding vowel in the English source word /piʧ/, and the final vowel /i/ is presumably epenthetic since there is no corresponding vowel in the source. However, this does not mean that epenthetic segments are represented as epenthetic in the synchronic grammars of Japanese speakers. Loanwords rarely undergo rendaku, and reducible words reduce nearly 100% of the time, exhibiting no alternations comparable to Yamato and Sino-Japanese roots. For example, both the 'epenthetic' vowel in /staː/ → ⟨su̲taː⟩'star' and the 'underlying' vowel in /ʧi̲kiN/ → ⟨ʧi̲kĩɴ⟩ 'chicken' are regularly reduced in the overt form. This lack of alternation may lead a Japanese learner to represent all reducible loanwords with heterorganic clusters underlyingly (i.e., /staː/ and /ʧkiN/). A Japanese learner might alternatively infer, especially after learning to read, that surface forms are equivalent to the underlying form (i.e., /sutaː/ and /ʧikiN/). The potential effects of orthography are discussed further in §2.2.2.4 below.

The choice between these two extremes seems to depend on the speaker and possibly individual lexical items. I provide here an anecdote involving the Canadian city name 'Toronto' to illustrate this point (Shigeto Kawahara, personal communication). The city name is ⟨toronto⟩ in Japanese, but when asked to spell the city name in English, Japanese speakers provide a range of answers from <Tront>, <Toront>, <Tronto>, and <Toronto>, revealing a confusion regarding whether the initial and final syllables ⟨to⟩ are the result of phonotactic repair or are underlyingly

present. Although the epenthetic vowel in question is not a high vowel, what such cases show is that Japanese speakers may treat vowels often used in phonotactic repairs as epenthetic or underlying depending on the speaker. Whether Japanese speakers treat the 'epenthetic' vowels in loanwords differently from underlying vowels in similar contexts is further explored in the Chapter 3.

2.2.2.4 Effects of orthography

It is likely that orthography would have an effect on how Japanese speakers represent lexical items underlyingly (in literate adults) and consequently on the apparent preference of CV structure. Young children are often shown to have poor phoneme awareness before being taught to read (Stuart and Coltheart, 1988; Chaney, 1992, 1994; Byrne and Fielding-Barnsley, 1995). In literate adults, however, phonemic awareness seems closely tied to orthography. For example, English speakers judge the vowel in words such as <demonic> to be homophonous with the vowel in <dem> but not in <dim>, presumably because of the orthographic form <e> despite the fact that the vowel in <demonic> can reduce to either [ɛ] or [ɪ] (Taft and Hambly, 1985). In a related, more recent study, Taft (2006) also found that speakers of Australian English (a non-rhotic dialect) consider pairs of lexical and nonce words such as <soak>–<soke> (i.e., [so͡ʊk]) as homophonous but not pairs such as <corn>–<cawn> (i.e., [kɔːn]), suggesting that the rhotic consonant is represented underlyingly despite its absence in overt forms. In short, these studies suggest that orthographic knowledge affects how segments and words are represented phonologically.

Additionally, speakers of a language with a phonographic system (e.g., alphabetic = Spanish, English; syllabary = Japanese) show a stronger effect of orthography on phonological processing than speakers of a language with a logographic system like in Chinese (Koda, 1988). This finding complicates the issue of how Japanese orthography affects phonological knowledge. Japanese uses a mix of logographic and phonographic scripts. Content words are written in *kanji*, which are logographic, Chinese characters. Function words, inflectional affixes, and loanwords are written in *kana* syllabary. Children, however, are first taught the *kana* syllabary and are introduced to more and

more *kanji* characters throughout their education, further complicating the issue of whether lexical items are more strongly associated with their respective logographic characters or the syllabic symbols.

Additionally, it is not possible to write coda consonants other than the two placeless codas /N/ and /Q/ with the *kana* syllabary. Apparent attempts at representing codas with independent place can be found in transliterations of Ainu words and Korean pop song lyrics in Japanese karaoke machines, where codas are represented with subscript /Cu/ *kana* symbols (e.g., [maʤimak̚] 'last' (Korean) → <マジマ$_{ク}$> (as in <ma.ʤi.ma.$_{\underline{ku}}$>)). However, the use of subscript *kana* is not a universal convention, and Japanese speakers disagree on how to read them (Whang, 2016).

For the reasons discussed above, it is difficult to precisely tease apart the effects of orthography and phonological knowledge in Japanese speakers, as noted by Otake et al. (1993). Some clearer evidence of orthographic effects in Japanese, however, can be found in phonological processes like *zuuja-go*, a reversing argot in Japanese that originated from jazz musicians and later spread to the broader lingo of the entertainment industry (Ito et al., 1996). Although simplifying a bit, the argot has a maximal prosodic structure of two bimoraic feet ($\mu\mu$)($\mu\mu$), and the first foot must be bimoraic while the second foot must be at least monomoraic. Generally the last two moras of the source word is moved to the front as shown in (22a). If the source word is too short, the last syllable is moved to the front and lengthened to meet the bimoraic requirement as shown in (22b). In fact, the term *zuuja-go* itself contains an argot of the word 'jazz'. Lastly, if the penultimate mora is a coda or the second half of a long vowel, the argot results in a non-exhaustive reversal as shown (22c).

(22) a. (pi)(jano) → (jano)(pi) 'piano'
(koː)(çiː) → (çiː)(koː) 'coffee'

b. (ʤa)(zu) → (zuː)(ʤa) 'jazz'
(a)(i) → (iː)(a) 'love'

c. (ba)n(do) → (do)m(ba) ‘band’

(be)ː(su) → (su)ː(be) ‘bass’

What is important for our purposes here is that when a syllable containing a geminate coda ends up as a result of the process in word-final position or before a consonant where gemination is dispreferred, the geminate is pronounced as [tsu] as shown in (23) below.

(23) a. (bi<u>k</u>)(kuri) → (kuri)(bi<u>tsu</u>) ‘surprised’

b. (ra)<u>p</u>(pa) → (pa)<u>tsu</u>(ra) ‘trumpet’

c.f. (ka)<u>p</u>(pa) → (pa)<u>k</u>(ka) ‘water imp’

The only plausible explanation for the pattern in (23) is that the geminate coda /Q/ is represented with a small *kana* symbol for <tsu>. When the coda becomes stranded in a position where there is no longer a following obstruent to gain a place feature from, it is pronounced as the full version of the symbol. This is shown in (24) below.

(24) a. (bi<u>k</u>)(kuri) → (kuri)(bi<u>tsu</u>) ‘surprised’

(び<u>っ</u>)(くり) → (くり)(び<u>つ</u>)

b. (ra)<u>p</u>(pa) → (pa)<u>tsu</u>(ra) ‘trumpet’

(ら)<u>っ</u>(ぱ) → (ぱ)<u>つ</u>(ら)

Similarly, reducible vowels can be moved to a non-reducing context as the consequence of the argot as well. For example, the word ‘medicine’ *(ku)(suri)* becomes *(suri)(ku)*, moving the vowel in *ku* from an obligatorily reducing context (between two obstruents, where neither or only one is a fricative) to an optionally reducing context (word-final position). When unreduced, the vowel always surfaces as a high back vowel, showing again that the orthography clues in Japanese speakers on the identity of the reduced vowel. The utilization of orthography in *zuuja-go* does not necessarily

mean that underlying representations in Japanese are equivalent to orthographic representations, but it does provide evidence that orthographic representations play a role in phonological processes.

2.2.3 Summary

This section proposed that the apparent phonotactic violations that result from high vowel reduction can be reconciled by separating the phonological levels in which phonotactic and reduction processes apply. Based on the observation that high vowel reduction targets both underlying and epenthesized vowels, reduction was argued to apply at the [overt] level after phonotactic repairs are made at the ⟨surface⟩ level. Since vowels need only be represented at the surface level for high vowel reduction to occur, the process is unconcerned with whether the target vowel is underlying or epenthetic.

Additionally, clusters were shown to be repaired differently depending on the stratum a word belongs to. Table 2.1 below summarizes how consonant clusters are generally repaired in Japanese by stratum.

	repair strategy		
stratum	*stem-allomorphy*	*assimilation*	*epenthesis*
Yamato	/kak̲-ta/ → [kai̲ta]	/aw̲-ta/ → [at̲ta]	–[5]
Sino-Japanese	–	/kat̲-ki/ → [kak̲ki]	/kak̲-toː/ → [kakuu̲toː]
Foreign	–	–	/piːʧ̲/ → [piːʧ̲i]

Table 2.1: Summary of cluster repair strategies in Japanese by stratum.

Furthermore, it seems likely that the *kana* syllabary of Japanese might have some effect on how words and segments are represented underlyingly, at least in literate adults. While orthographic representations would have little effect in preliterate children, if it is the case that Japanese children start out with a grammar that allows underlying clusters initially, learning to read *kana*, which explicitly prohibits cluster representation, might result in a reanalysis of their phonological representations (Goswami, 2000).

[5]Only after /s/.

2.3 Conclusion

Having established the levels in which phonotactic and reduction processes apply, the rest of this dissertation will refocus on the issue of recoverability and address the following two issues. First, if it is the case that reduction can result in both devoiced and deleted high vowels, is there a way to predict which? This is the focus of chapter 3, and I propose that it is phonotactic predictability, which would allow varying levels of recoverability from context, that determines the degree of reduction. Second, a Japanese learner must learn that reduced vowels, regardless of whether they are devoiced or deleted, are equivalent to high vowels. As a listener, the choice of high vowel can be recovered either from the context or acoustic cues in the signal depending on whether or not the preceding consonant undergoes allophony (i.e., /t/ $\leftrightarrow$ ⟨ʧi, tsu⟩, /s/ $\leftrightarrow$ ⟨ʃi, su⟩, /h/ $\leftrightarrow$ ⟨çi, ɸu⟩, but /p/ $\leftrightarrow$ ⟨pi, pu⟩, /k/ $\leftrightarrow$ ⟨ki, ku⟩, /ʃ/ $\leftrightarrow$ ⟨ʃi, ʃu⟩), which is the focus of the perception experiment in Chapter 4. Also, a lexicon that links alternations to the same lexical item is required for such equivalence learning to occur, which is the focus of the computational model in Chapter 5.

CHAPTER 3

Predictability-conditioned coarticulation

3.0 Introduction

The aim of this chapter is to investigate the acoustic properties of high vowel reduction in Japanese, and more specifically, what cues in the signal allow for recovery of a reduced vowel and whether predictability from phonotactic knowledge affects the availability of these cues. Although the process of high vowel reduction in Japanese has garnered great interest in the field of phonetics and phonology, is still debated whether the reduced vowels are devoiced (i.e., lose only their phonation) or are completely deleted. The production experiment presented in this chapter asks specifically whether high vowels in contexts of low phonotactic predictability are more likely to devoice than those in high predictability contexts.

3.0.1 Background

High vowel reduction is considered to be an integral feature of standard modern Japanese (Imai, 2010), so much so that dictionaries exist with explicit instructions for reducing environments (Kindaichi, 1995: pp.25–27). The phenomenon is commonly described as involving phonemically short high vowels /i/ and /u/, which are reduced in C_1VC_2 sequences when the vowels are unaccented and both C_1 and C_2 are voiceless obstruents. For example, while the /u/ in /kúʃi/ 'free use' and /kuʃi/ 'skewer' are both between two voiceless obstruents, only /kuʃi/ 'skewer' undergoes reduction because the vowel is unaccented. Likewise, the /u/ is unaccented in both /kuki/ 'stem' and /kugi/ 'nail', but only /kuki/ 'stem' undergoes reduction because the /u/ is flanked by two voiceless stops. Whether the process is categorical depends largely on the manner of the flanking consonants, where reduction rates can be as low as 60% when C_1 is a fricative or affricate and C_2 is a fricative, but can be nearly 100% elsewhere (Maekawa and Kikuchi, 2005; Fujimoto, 2015). Although an unaccented high vowel may also be reduced if the vowel is preceded by a voiceless fricative or affricate and is followed by a pause, as in /káʃu#/ → [káʃu̥#] 'singer)' (Fujimoto, 2015), this last case is speaker-dependent and thus will not be discussed in the current study. In light of the discussion in Chapter 2 regarding the phonological level in which high vowel reduction applies, the input for high vowel reduction is the /surface/ form and the output is the [overt] form.

Despite the productivity of high vowel reduction in Japanese and the amount of interest the phenomenon has received in the field of phonetics and phonology, it is still debated whether the reduced vowels are devoiced or are completely deleted. The current study analyzes the duration and center of gravity measurements of C_1 to investigate the effects of recoverability on coarticulation. Recoverability refers to the ease of accessing the underlying form—stored mental representations—from a given surface form—actual, variable output signals (e.g., [kæt̚, kæt$^{\text{h}}$] → /kæt/ 'cat'; Mattingly, 1981; McCarthy, 1999; Chitoran et al., 2002). Recovery can be achieved by interpreting information explicitly present in the acoustic signal (phonetic interpretability) or by pre-

diction based on context (phonotactic predictability). However, recoverability can be compromised if both phonetic interpretability and phonotactic predictability are insufficient.

Varden (2010) states what seems to be a prevalent assumption in the literature on Japanese high vowel reduction, which is that since high vowels trigger allophonic variation for /t, s, h/ (i.e., /t/ → [ʧi, ʦu]; /s/ → [ʃi, su]; /h/ → [çi, ɸu]), the underlying vowel is easily recoverable even if the vowel were to be phonetically deleted unlike in cases where the vowel is unpredictable. To give a concrete example, [ɸku] 'clothes' would be analyzed as /huku/, since (i) [ɸ_k] is a reducing context where the vowel to be recovered can only be one of /i, u/, and (ii) the mere presence of [ɸ] narrows the choice down to /u/ because [ɸ] can only occur as an allophone of /h/ preceding /u/ in non-loanwords. What Varden assumes can be represented as the gestural scores in Figure 3.1 below. In the case of [ʃi̥ka] 'deer' (top), the alveo-palatal fricative can be followed by both high vowels. Deletion of the target high vowel /i/ would jeopardize the recovery of the vowel, and thus the oral vocalic gesture remains. The glottis, however, remains open through the vowel, resulting in a devoiced vowel. In the case of [ɸku] 'clothes' (bottom), the oral vocalic gesture for /u/ need not be retained since the mere presence of [ɸ] limits the possible vowel that can follow to /u/, and thus the vocalic gesture simply deletes.

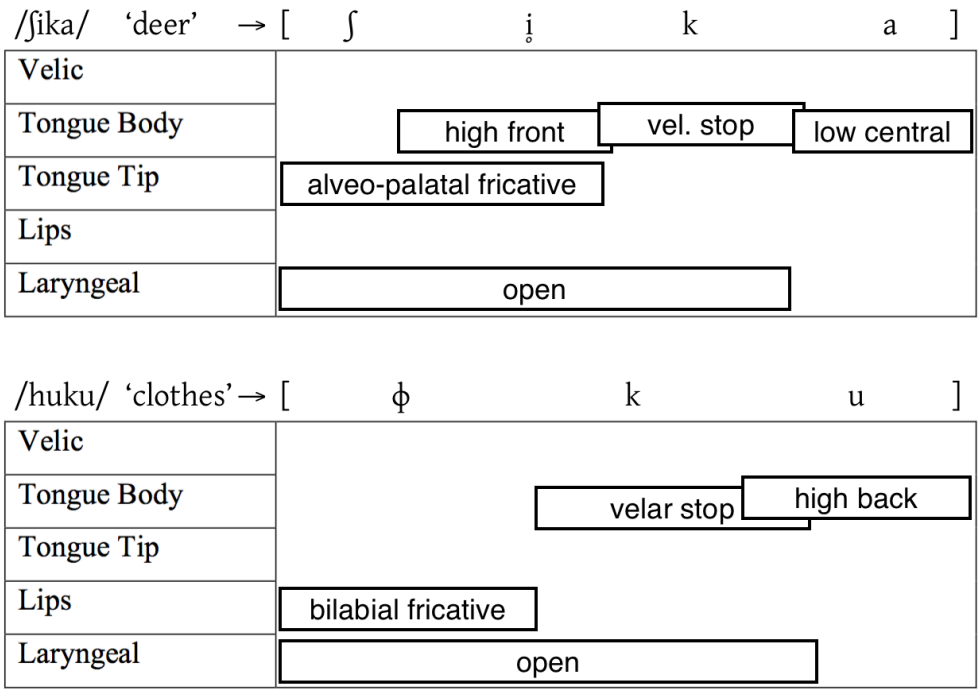

Figure 3.1: Gestural scores of Japanese high vowel devoicing (top) and deletion (bottom).

Rephrased in terms of phonetic interpretability and phonotactic predictability, what Varden is proposing is that the amount of phonetic interpretability decreases when phonotactic predictability is high. If true, this also leads to the prediction that if phonotactic predictability is low, phonetic interpretability should increase. A number of studies have proposed similar recoverability-conditioned coarticulation as well, where speakers seem to preserve or enhance the phonetic cues of a target segment in situations where the target segment would be less perceptible, such as when a phoneme inventory contains acoustically similar phonemes (Silverman, 1997) or in word-initial stop-stop sequences, where the closure of the second stop would obscure the burst of the first (Chitoran et al., 2002). However, whether the amount of C_1V coarticulation is similarly modulated by predictability in Japanese has not been tested systematically. The current study therefore investigates /t, s,

h/, which are targets of allophonic variation before high vowels, and /k, ʃ/[1], which are not. The consonants included in the current study and the vowels that can follow each of the consonants is summarized in Table 3.1 below. It should be noted that a contrast between /s and ʃ/ is allowed before all vowels except /i/, where the contrast is neutralized to /ʃ/. /ʃ/ additionally cannot precede /e/ in non-loanwords. Furthermore, while [ʧ], [s], and [ɸ, ç] are allophones of /t, s, h/, respectively, they are also semi-phonemic and can precede all other non-high vowels in Sino-Japanese words and loanwords. The bilabial stop /p/ is excluded because it almost never occurs word-initially outside of loanwords and mimetic words. When /p/ does occur in Yamato and Sino-Japanese words, it is usually the result of |h| allophony after codas (e.g., |kaN+hai| → /kampai/ 'cheers (dry+cup)') or part of a suffix which begins with a geminate (e.g., |kodomo+ppoi| → /kodomoppoi/ 'childish (child+ish)'). Furthermore, the affricate /ts/, which is another allophone of |t| that occurs before /u/, is also not included to keep the number of stimuli balanced between high and low predictability tokens.

		i	u
High predictability	ʧ	✓	–
	s	–	✓
	ɸ	–	✓
	ç	✓	–
Low predictability	k	✓	✓
	ʃ	✓	✓

Table 3.1: Consonants used in stimuli and high vowels that can follow. "–" means that the vowel is not phonologically possible in this context (in non-loanwords).

3.0.2 Previous studies

There are primarily three ways in which reduced high vowels are argued to be manifested acoustically: (i) by lengthening the burst/frication noise of C_1 which carries coarticulatory cues of a

[1]Although the post-alveolar fricative is more accurately an alveo-palatal fricative /ɕ/, the IPA symbol for the post-alveolar fricative is used throughout the chapter for the sake of readability and to enhance differentiation from the palatal fricative [ç] which is an allophone of /h/.

devoiced vowel (Han, 1994), (ii) by unphonating the vowel and coloring the C_1 burst/frication noise with the retained oral gestures without necessarily lengthening C_1 (Beckman and Shoji, 1984), and (iii) by deleting the vowel altogether (Vance, 2008). Each of the proposed manifestations has contrad8icting evidence in previous literature as discussed below. It should be noted that although high vowel reduction is more commonly referred to as high vowel "devoicing" in Japanese, the term *reduction* is used throughout this dissertation as a general term to refer to a lack of phonation associated with a target vowel since there is disagreement regarding whether the target vowels are devoiced or simply deleted.

Although it is commonly argued that C_1 is longer in reduced syllables than in unreduced syllables, the empirical evidence is not unanimous. Part of the problem in the lack of consensus regarding the effects of vowel reduction on C_1 duration in Japanese is that there are differences in the methodologies and stimuli among the studies. For example, Varden (1998) examines /k,t/ (where /t/ $\rightarrow$ [ʧi, tsu]) and reports that the burst and aspiration of C1 in reduced syllables are significantly longer than the consonant portion of their corresponding unreduced CV syllables. On the other hand, studies that focus on /s/ ($\rightarrow$ [ʃi, su]; Beckman, 1982; Beckman and Shoji, 1984; Faber and Vance, 2000) consistently report that there is no significant difference in duration between /s/ in reduced and unreduced syllables. In other words, studies that investigate fricatives find no lengthening effect while studies that investigate stops and affricates find lengthening effects.

Additionally, studies that report lengthening effects generally assume that Japanese is mora-timed and that moras are roughly equal in duration. Based on these assumptions, the duration results of individual C_1 are often collapsed (Tsuchida, 1997; Kondo, 2005) or C_1 in reduced contexts are compared to different segments in unreduced contexts (Han, 1994). These practices are justified if moras in Japanese are indeed equal in duration, but Warner and Arai (2001a,b) argue that there is actually no compensatory lengthening effect related to mora-timing and that the apparent rhythm in Japanese is epiphenomenal, the result of a confluence of factors that result from the phonological structure of Japanese.

While it is conceptually plausible that the presence of an underlying vowel can be signaled solely by C_1 lengthening, especially if mora preservation is the reason behind it, much of the literature arguing for compensatory lengthening also assumes that reduced vowels are devoiced rather than deleted. A number of articulatory studies looking at /k, t, s/ as C_1 found that the glottis is wider when the vowel in a C_1VC_2 sequence is reduced than when it is not, and that there is only one activity peak for the laryngeal muscles, aligned with the onset of C_1 in reduced sequences, resulting in a long frication or a frication-like burst release for stops (Fujimoto et al., 2002; Tsuchida et al., 1997; Yoshioka et al., 1982). Since there is no laryngeal activity associated with C_2 apart from the carry-over from that of C_1 and the abduction peak for the glottis was found to be larger than the sum of two voiceless consonants, these results are interpreted to mean that the glottal gesture is being actively controlled to spread the feature [+spread glottis] from the first consonant to the second. As a consequence of this spreading, the intervening high vowel is devoiced. Despite the lack of a laryngeal gesture associated with phonation, the presence of formant-like structures in the burst/frication noise of C_1 are often reported, which is taken as evidence of retained oral vowel gestures. For example, a recent acoustic study by Varden (2010) reports visible formant structures apparent in the fricated burst noise of [k̟̥i, kʊ̟̥], which are interpreted to be the result of oral gestural overlap that allows consistent identification of the underlying devoiced vowel.

In contrast, Ogasawara (2013) reports a lack of visible formant structures in the burst/frication noise of /k, t/ in most reduced cases and argues that this provides support for the claim that high vowel reduction results in deletion rather than devoicing (Hirose, 1971; Yoshioka, 1981). The lack of apparent formant structures in the burst/frication noise of C_1, however, seems to be an inadequate criterion for measuring the presence of vocalic oral gestures. While Beckman and Shoji (1984) also report that the presence of formant-like structures on the frication noise of /ʃ/ is inconsistent, spectral measurements of [ʃ] showed a small yet noticeable influence of reduced vowels on the aperiodic noise of the preceding fricative, where the mean frequency of [ʃu̟̥] was lower than [ʃi̟̥] by approximately 400 Hz, suggesting a coarticulatory effect of a reduced vowel. Perceptually, this

difference was enough to aid the listeners in identifying the underlying vowel above the rate of chance (77% for [ʃi̥] and 67% for [ʃu̥]).

3.0.3 Possible effects of predictability on coarticulation

There are three main possibilities with respect to the question of how predictability affects coarticulation. The first is that high vowel reduction is blind to predictability and is driven primarily by Japanese phonotactics, which has a strict CVCV structure that disallows tautosyllabic clusters (Kubozono, 2015). If this is the case, then no difference between low predictability and high predictability C_1 would be found, where the reduced vowel never deletes completely but always devoices instead, coloring the burst or frication noise of C_1 to signal the presence of the target vowel (Beckman and Shoji, 1984; Varden, 2010). The second is that the degree of coarticulation between C_1 and the following vowel is not systematic but rather a consequence of how the reduced vowel happened to be lexicalized for the speaker. Ogasawara and Warner (2009) found in a lexical judgment task that when Japanese listeners were presented with unreduced forms of words where reduction is typically expected, reaction times were longer than when presented with reduced forms. This suggests that the reduced forms, despite their phonotactic violations, can have a facilitatory effect on lexical access due to their commonness, making vowel recovery unnecessary (Cutler et al., 2009; Ogasawara, 2013). The third and last option is that high vowel reduction is constrained by recoverability. In this case, increased coarticulation would be observed either by lengthening or spectral changes of C_1 burst/frication when the predictability of the target vowel is unreliable from a given C_1 to aid phonetic interpretability as in the case of /k, ʃ/, but not when predictability is high, as in the case of /ʧ, s, ɸ, ç/. This last outcome would also be compatible with the idea that reduced forms are lexicalized as such (Ogasawara and Warner, 2009), but with the caveat that the degree of reduction is not entirely random, but dependent on predictability from context.

While this chapter does not explore sociolinguistic factors that affect high vowel reduction, it is worth noting that men have been reported to reduce more than women (Okamoto, 1995; Yuen and Hubbard, 1998) and that reduction rates are higher overall in younger speakers (Varden and Sato, 1996). However, Imai (2010) found that while younger speakers did tend to reduce more, this was only true for men. Young female speakers were actually shown to reduce the least among all age groups. Based on these findings, Imai proposes that high vowel reduction may be being actively utilized as a feature of gendered speech. If high vowel reduction is being utilized as sociolinguistic feature, then the process could not be a purely phonological or a phonetic process, and thus I recruited a balanced number of men and women to investigate any gender-based differences.

3.1 Materials and methods

3.1.1 Participants

Twenty-two monolingual Japanese speakers (12 women and 10 men) were recruited in Tokyo, Japan. All participants were undergraduate students born and raised in the greater Tokyo area and were between the ages 18 and 24. Although all participants learned English as a second language as part of their compulsory education, none had resided outside of Japan for more than six months and have not been overseas within a year prior to the experiment. All participants were compensated for their time.

3.1.2 Materials

The stimuli for the experiment were 160 native Japanese and Sino-Japanese words with an initial C_1iC_2 or C_1uC_2 target sequence. The stimuli were controlled to be of medium frequency (20 to 100 occurrences, which is the mean and one standard deviation from the mean, respectively) based on the frequency counts from a corpus of Japanese blogs (Sharoff, 2008). Any gaps in the data were

filled with words of comparable frequency based on search hits in Google Japan (10 million to 250 million). Since high vowel reduction typically occurs in unaccented syllables, an accent dictionary of standard Japanese (Kindaichi, 1995) was used as reference to ensure that none of the stimuli had a target vowel in an accented syllable.

The stimuli were divided into *low predictability* and *high predictability* groups. Predictability, for the purposes of this study, refers specifically to the predictability of vowel backness, given high vowels. Examples of the reducing stimuli are shown in Table 3.2 below.

stimulus type	C_1	*V*	*example*	*gloss*
low predictability	k	i	kiki	'handedness'
		u	kuki	'twig'
	ʃ	i	ʃikoː	'thought'
		u	ʃukoː	'plan'
high predictability	ʧ	i	ʧikʲuː	'earth'
	s	u	sukuː	'rescue'
	ɸ	u	ɸukoː	'unhappy'
	ç	i	çiteː	'denial'

Table 3.2: Example of reducing stimuli by C_1 and vowel.

As shown above, for the low predictability group, C_1 was either /k, ʃ/ after which both /i, u/ can occur. For the high predictability group, C_1 was one of [ʧ, s, ɸ, ç], after which only one of the high vowels is possible. The two groups were further divided into *reducing* and *non-reducing* contexts. The difference between reducing and non-reducing tokens was that C_2 was always a voiceless stop (i.e., [p, t, k]) for reducing contexts as shown above, but a voiced stop for non-reducing tokens (i.e., [b, d, ɡ]). Since high vowel reduction typically requires the target vowel to be flanked by two voiceless obstruents, it was expected that reduction would not occur in the non-reducing contexts. The C_1 and C_2 combinations resulted in fricative-stop, affricate-stop, or stop-stop contexts. These contexts were chosen for two reasons: (i) these are contexts in which high vowel reduction is reported to occur systematically and categorically (Fujimoto, 2015), and (ii) the C_2 stop closure

clearly marks where the previous segment ends. There were 10 tokens per C_1V combination within each context, for a total of 160 tokens (80 reducing and 80 non-reducing).

It should be noted that while both /i, u/ can follow /ʧ/ in Japanese, only /ʧi/ was included in the stimulus set because /ʧ/ is rarely followed by short /u/ in Japanese. A search of the Shogakukan (2013) dictionary revealed that of the 6,041 entries that begin with /ʧ/, 38% are followed by /i/ compared to only 1% that are followed by the short vowel /u/, of which only 5 words were in potentially reducing contexts. In other words, when /ʧ/ is followed by a reducible high vowel, the phonotactic distribution of the language heavily favors the vowel /i/, making the environment highly predictable.

Ten additional English loanwords with word-initial /sC_2/ clusters were included in the stimuli, where C_2 was a voiceless stop (e.g., schedule → /sukeʤu:ru/). The environment was therefore a reducing environment in Japanese. These loanwords were included to see if epenthetic vowels, which are underlyingly not present in English, are represented differently in Japanese.

3.1.3 Design and procedure

All tokens were placed in the context of unique and meaningful carrier sentences of varying lengths. No tokens were included more than once in the experiment, and no two carrier sentences were identical. All carrier sentences contained at least one stimulus item, and the sentences were constructed so that no major phrasal boundaries immediately preceded or followed the syllable containing the target vowel. An example carrier sentence, which was actually uttered by a weather forecaster in Japan, is given below with glosses.

manatsu-no ʃigaisen-ni-wa k̟i-o-ʦuke-maʃo:
midsummer-LNK ultraviolet rays-DAT-TOP be careful.VOL

'Let's be careful of midsummer's ultraviolet rays'

The carrier sentences were presented one at a time to the participants on a computer monitor as a slideshow presentation. The participants advanced the slideshow manually, giving the participants time to familiarize themselves with the sentences. They were also allowed to take as many breaks as they thought was necessary during the recording. As a fluent speaker of the language, I gave all instructions in Japanese. I remained with the speaker during the recording and prompted the speaker to repeat any sentences that were produced disfluently. All participants were recorded in a sound-attenuated booth with an Audio-Technica ATM98 microphone attached to a Marantz PMD-670 digital recorder at a sampling rate of 44.1 kHz at a 16 bit quantization level. The microphone was secured on a table-top stand, placed 3-5 inches from the mouth of the participant.

3.1.4 Data Analysis

Once the participants were recorded, the waveform and spectrogram of each participant were examined in Praat to (a) code each token for reduction, (b) to measure the duration of C_1 and the following vowel, and (c) to measure the center of gravity of C_1 burst/frication noise. The spectrogram settings were as follows: pre-emphasis was set at +6 dB, dynamic range was set at 60 dB, and autoscaling was turned off for consistency of visual detail. Because visual inspection alone is an inadequate method for determining the presence of vowel coarticulation on C_1 (Beckman and Shoji, 1984), tokens were simply coded for "reduction", a term used throughout this chapter to collectively to refer to devoicing and deletion of the vowel. The criteria used for reduction status are described in the following section.

3.1.4.1 Reduction analysis

Vowels in reducing environments were coded as unreduced if there was phonation accompanied by formant structures between C_1 and C_2. Vowels were coded as reduced when there was no phonation between C_1 and C_2. Below in Figure 3.2 are examples from the same female speaker. On the left is

an unreduced vowel in the word [kuk̟̊i] ‘twig’, which shows clear phonation and formant structures between C_1 and C_2. On the right is a reduced vowel in the word [kʊ̥tɛɴ] ‘period’, where there is neither phonation nor formant structures between C_1 and C_2.

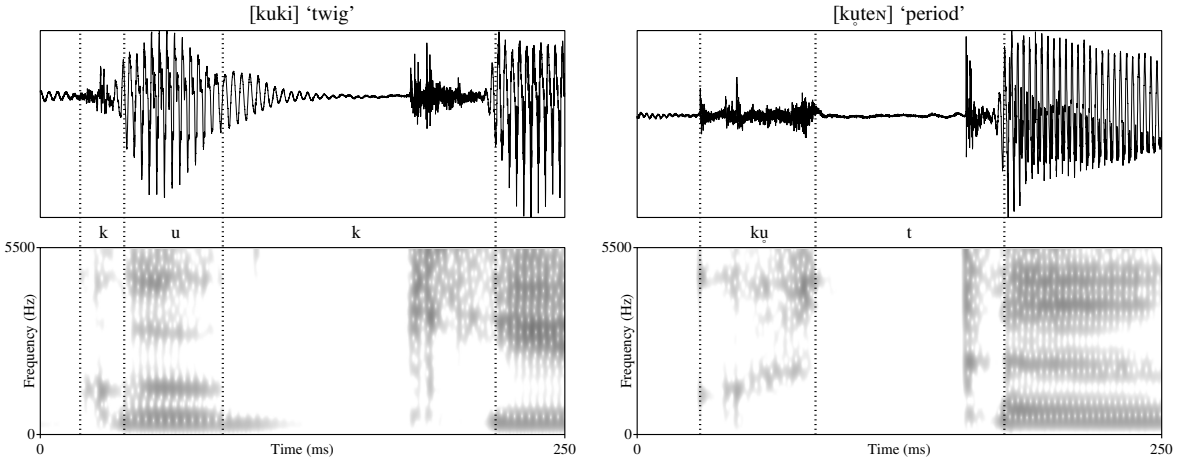

Figure 3.2: Waveform and spectrogram of unreduced (left) and reduced (right) vowels in reducing environments, showing landmarks for C_1, vowel, and C_2 duration.

The coding criteria were similar for non-reducing tokens. A vowel was coded as unreduced if phonation and formant structure were both present between C_1 and C_2. Otherwise, a vowel was coded as reduced. Below in Figure 3.3 are examples from another female speaker. On the left is an unreduced vowel in the word [ʃugeː] ‘handicraft’, where there is a clear formant structure accompanying phonation. On the right is a rare case of a reduced vowel in a non-reducing context, the word [ʃʊ̥daika] ‘theme song’, where there is phonation between C_1 and C_2 but no formant structure.

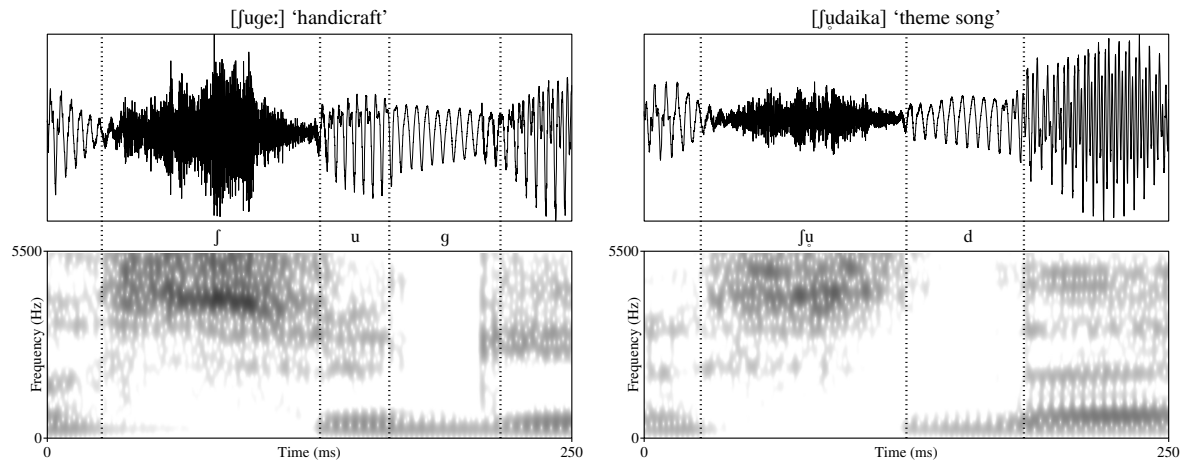

Figure 3.3: Waveform and spectrogram of unreduced (left) and reduced (right) vowels in non-reducing environments, showing landmarks for C_1, vowel, and C_2 duration.

3.1.4.2 Duration analysis

Once all tokens were coded for reduction status, duration measurements were taken to investigate how reduction affects the gestural timing of C_1 and the target high vowel. For [k] and [ʧ], duration measurements excluded the silence from closure and included only the aperiodic burst energy. For fricative C_1, duration measurements included the entire aperiodic frication noise. For tokens coded as reduced, C_1 measurements were assumed to include the reduced vowel because the vowel could not be isolated from C_1 reliably. For unreduced tokens, C_1 was measured from the onset of burst/frication noise to the onset of vowel F2.

3.1.4.3 Center of gravity analysis

Center of gravity (COG), which is the amplitude weighted mean of frequencies present in the signal (Forrest et al., 1988), was also calculated for C_1 to investigate the degree of coarticulation between C_1 and the target vowel. COG measurements are used based on Tsuchida (1994), who found that Japanese listeners rely primarily on C_1 centroid frequency (i.e., COG) to identify reduced

vowels. COG measurements are known to be particularly sensitive to changes in the front oral cavity (Nittrouer and McGowan, 1989), so the effects of increased coarticulation between a vowel and C_1 on COG values are expected to differ by the backness and roundedness of the vowel as well as C_1 place of articulation. The predicted effects of vowel coarticulation on each C_1 are discussed in detail in §3.2.3 together with the results.

Before measuring COG values, the sound files were high pass filtered at 400 Hz to mitigate the effects of f0 on the burst/frication noise. The filtered sound files were then down-sampled to 22,000 Hz. The COG values measured therefore were taken from FFT spectra in the band of 400 to 11,000 Hz (Forrest et al., 1988; Hamann and Sennema, 2005). With the exception of /k/, two center of gravity (COG) measurements were taken from 20 ms windows for each C_1: one starting 10 ms after the beginning of C_1 burst/frication (COG1) and one ending 10 ms before the end of C_1 burst/frication (COG2). The 10 ms buffers were used to mitigate the coarticulatory effects of segments immediately adjacent to C_1. Comparisons of COG1 and COG2 between reduced and unreduced tokens allows inference of how early or late in the consonant vowel coarticulation effects begin. Comparison of ΔCOG (COG2 $-$ COG1) also allows testing of how the trajectory of coarticulatory effects differs between reduced and unreduced tokens.

For /k/, COG measurements were taken from a single 20 ms window at the midpoint of the burst. Two COG measurements could not be taken from /k/ because /k/ duration in unreduced tokens were too short for two measurements. /k/ tokens shorter than 20 ms were excluded from analysis, which resulted in the loss of five tokens, or 0.6% of the /k/ data. Since the vocalic gesture of the following vowel most likely begins during the stop closure for /k/ (Browman and Goldstein, 1992a; Fowler and Saltzman, 1993), the single COG measurement is assumed to be equivalent to the COG2 measurements of other consonants.

3.2 Results

Statistical analyses were performed by fitting linear mixed effects models using the *lme4* package (Bates et al., 2015) for R (R Core Team, 2016). In order to identify the maximal random effects structure justified by the data, a model with a full fixed effects structure (i.e., with interactions for all the fixed effects) and the most complex random effects structure was fit first. If the model did not converge, the random effects structure was simplified until convergence was reached while keeping the fixed effects constant (Barr et al., 2013). The simplest random effects structure considered was one with random intercepts for participant and word with no random slopes.

Once the maximal random effects structure was identified, a Chi-square test of the log likelihood ratios were performed to identify the best combination of fixed effects. A complex model with all interaction terms was fit first, which was then gradually simplified by removing predictors that did not significantly improve the fit of the model, starting with interaction terms. The simplest model considered was a model with no fixed effects and only an intercept term. Because the fixed and random effects were slightly different for each of the analyses performed, the structure of the final model will be described in the respective sections below.

3.2.1 Reduction rate

3.2.1.1 Overall reduction rates and analysis

Reduction rates were at or near 100% for reducing tokens, while non-reducing tokens had significantly lower reduction rates, as shown in Table 3.3 below.

stimulus type	C_1	*V*	*reducing*	*non-reducing*
low predictability	k	i	1.000	0.077
		u	0.959	0.032
	ʃ	i	1.000	0.086
		u	0.973	0.073
high predictability	ʧ	i	1.000	0.191
	ç	i	1.000	0.015
	ɸ	u	1.000	0.042
	s	u	1.000	0.214
overall			0.992	0.091

Table 3.3: Reduction rate by C_1V and context.

A mixed logit model was fit using the *glmer()* function of the *lmer* package for the overall reduction rate with reducing context, predictability, gender, and their interactions as predictors. Vowel was not included as a predictor because it is redundant for high predictability tokens since only one vowel is allowed. Random intercepts for participant and word were added to the model. By-participant random slopes for context and predictability as well as by-word random slopes for gender were also included in the model. The final model retained the full random effects structure. The following predictors were removed from the fixed effects structure of the final model as they were not significant contributors to the fit of the model: three-way interaction ($p = 0.999378$), context:gender interaction ($p = 0.901798$), and predictability:gender interaction ($p = 0.062329$). The function for the final model, therefore, was as follows:

model = glmer(reduction ∼ context + predictability + gender + context:predictability + (1 + context + predictability | participant) + (1 + gender | word), family = binomial(link = 'logit'), data = non-loanwords)

The results of the final model showed that the difference in reduction rates between reducing and non-reducing contexts was significant ($p < 0.0001$) and that men were more likely to reduce than women ($p = 0.0175$). Predictability and context:predictability interaction did not have significant effects ($p = 0.2374$ and 0.7237, respectively).

An additional analysis was performed on just the non-reducing subset of the data because reducing tokens reduced essentially 100% of the time and had no between-participant differences to test statistically. First, a mixed logit model was fit to the low predictability non-reducing tokens with gender, C_1, vowel, and their interactions as predictors. Random intercepts for participant and word were included in the model. By-participant random slopes for C_1 and vowel, and by-word random slopes for gender were also included. /ʃ/ tokens as produced by female participants were the baseline. However, none of the predictors were significant contributors to the fit of the model, and a Chi-square test showed the fit of the intercept-only model was not significant different from more complex models. In other words, /k, ʃ/ had similar reduction rates in non-reducing contexts regardless of vowel or gender.

Second, a mixed logit model was fit to the high predictability non-reducing tokens with gender, C_1, and their interaction as predictors. Random intercepts for participant and word were included in the model. By-participant random slopes for C_1 and by-word random slopes for gender were also included. The interaction term was not a significant contributor to the model ($p = 0.07828$), and thus was removed from the final model. /ʧ/ tokens as produced by female participants were the baseline. The results showed that male participants were more likely to reduce than women ($p = 0.011490$). C_1 did not have a significant effect ($p = 0.171173$, 0.092141, and 0.516625 for /ɸ, ç, s/ respectively).

3.2.1.2 Summary of reduction rate results

The analysis of reduction rates showed that there is an effect of context on overall reduction rates. At essentially 100%, reduction rates are significantly higher in the reducing environments than in non-reducing environments. Male participants were also shown to be more likely to reduce than female participants, but the difference did not come from reducing tokens. Separate analyses of low and high predictability tokens revealed men reduced more in high-predictability environments, where reduction was not actually phonologically conditioned (e.g., /ɸugoːri/ → [ɸ̥ugoːri] ‘unreasonable’).

3.2.2 Duration

The majority of previous studies that report lengthening effects of reduction on C_1 have focused on /k, t/ (Varden, 1998), while studies that report a lack of such effect focus on /s, ʃ/ (Beckman and Shoji, 1984; Vance, 2008). There are two confounded differences between /k, t/ and /s, ʃ/ that may be contributing to the contrary results: manner and inherent duration. /k, t/ are non-continuant obstruents while /s, ʃ/ are continuants, but it is also the case that the burst of the former are inherently much shorter than the frication noise of the latter. This means that the contrary results could be due to either or both of these differences. The allophones of /h/—[ɸ, ç]—are therefore crucial in teasing apart the two factors because [ɸ, ç] are fricatives but are also similar in duration as the frication portion of [ʧ] in Japanese.[2]

3.2.2.1 Overall duration results and analysis

Duration results are shown in Figure 3.4 and Table 3.4 below. The results suggest that overall C_1 burst/frication durations are not different between women and men. Reduction seems to have a lengthening effect only on non-fricative obstruents (i.e., /k̟i, ku, ʧi/). For fricatives, reduction seems to have no effect on /ɸu/ and a shortening effect on others (i.e., /çi, su, ʃu, ʃi/).

[2] An analysis of consonant durations in the Corpus of Spontaneous Japanese revealed that there is no significant duration difference between [ʧ] and [ɸ] in unreduced contexts (∼65 ms; $p = 0.891$), and between [ʧ] and [ç] in reduced contexts (∼75 ms; $p = 0.475$).

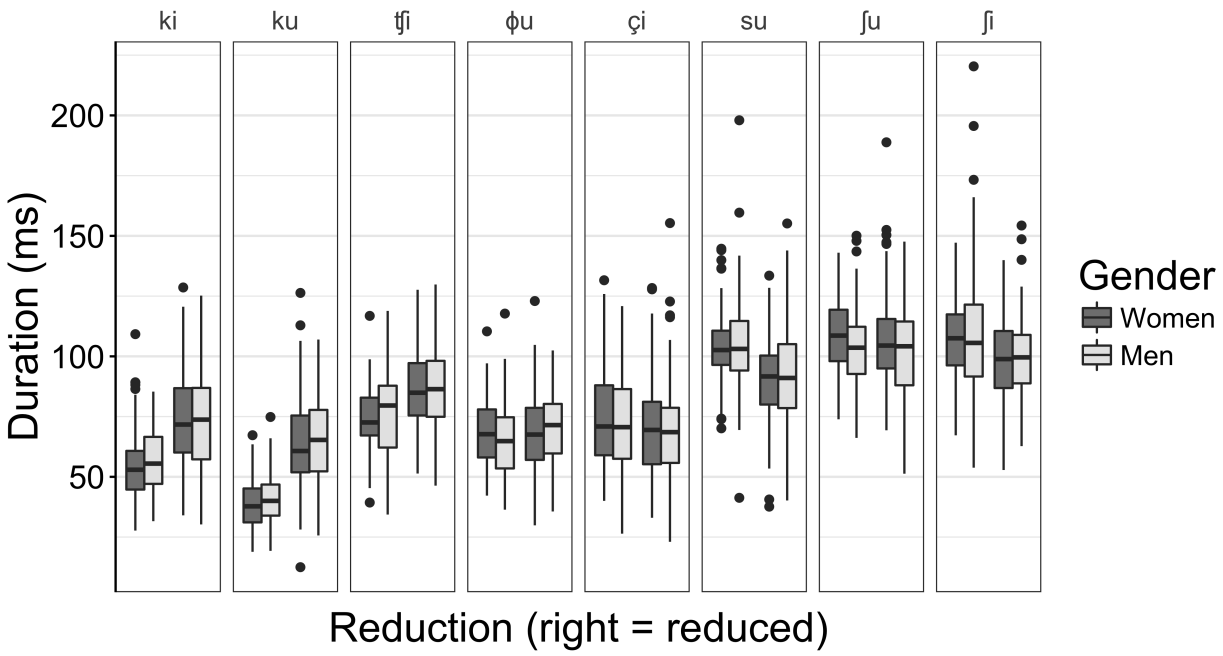

Figure 3.4: C_1 duration in ms by C_1V, gender, and reduction.

	C_1V	unreduced		reduced	
		female	*male*	*female*	*male*
low predictability	**ki**	**55 (*13*)**	**57 (*13*)**	**74 (*20*)**	**74 (*20*)**
	ku	**39 (*11*)**	**41 (*11*)**	**65 (*17*)**	**65 (*17*)**
	ʃi	107 (*16*)	108 (*26*)	99 (*18*)	99 (*17*)
	ʃu	109 (*14*)	104 (*16*)	106 (*19*)	102 (*20*)
high predictability	**ʧi**	**74 (*14*)**	**76 (*17*)**	**86 (*15*)**	**86 (*18*)**
	çi	75 (*21*)	74 (*19*)	71 (*18*)	71 (*19*)
	ɸu	69 (*14*)	66 (*15*)	69 (*15*)	71 (*14*)
	su	104 (*15*)	105 (*21*)	91 (*17*)	92 (*20*)

Table 3.4: C_1 mean duration (*standard deviation*) in ms. Lengthening effect in bold.

A linear mixed effects regression model was fit to the overall duration results with reduction, gender, C_1, and their interactions as predictors. Again, vowel was not included as a predictor because it is only meaningful for /k, ʃ/ tokens. Random intercepts for participant and word were added to the model. By-participant random slopes for context and C_1 were also included in the model,

as well as by-word random slopes for gender. Because there is currently no consensus on how to accurately calculate *p*-values for mixed effects models, absolute *t* values greater than 2 were regarded as significant (Baayen et al., 2008).

The final model retained the full random effects structure. The following non-significant predictors were removed from the fixed effects structure of the final model: three-way interaction ($p = 0.3040$), reduction:gender interaction ($p = 0.9266$), gender:C_1 interaction ($p = 0.6081$), and gender ($p = 0.5797$). The final model therefore retained reduction, C_1, and their interaction as predictors. The function for the final model was as follows:

*model = lmer(duration ~ reduction * C1 + (1 + context + C1 | participant) + (1 + gender | word), control=lmerControl(optimizer="bobyqa"), REML = F, data = non-loanwords)*

The final model's results are summarized below in Table 3.5. Unreduced /k/ tokens are the baseline.

	Estimate	Std. Error	*t*	
(Intercept)	47.365	2.264	20.917	*
reduced	22.068	3.106	7.106	*
ɸ	20.464	3.516	5.819	*
ç	26.808	3.746	7.156	*
ʧ	27.399	3.634	7.539	*
s	55.317	3.751	14.749	*
ʃ	59.454	3.155	18.844	*
reduced:ɸ	-20.396	4.877	-4.182	*
reduced:ç	-25.340	4.964	-5.105	*
reduced:ʧ	-10.514	4.895	-2.148	*
reduced:s	-33.451	4.903	-6.823	*
reduced:ʃ	-27.009	3.983	-6.781	*

Table 3.5: Linear mixed effects regression model results for overall C_1 duration.

The results show that reduction indeed has a lengthening effect of 22 ms on /k/. The intercept estimates for C_1 predictors show that all other C_1 are significantly longer than the /k/ baseline. The negative values of the estimates for the reduction:C_1 interaction predictors also show that reduction has a smaller lengthening effect on all other C_1 relative to the /k/ baseline.

The model above only shows how other C_1 differ from /k/. In order to explore whether reduction actually had significant effects on the individual C_1, differences of least squares means were calculated from the final model using the *difflsmeans()* function of the *lmerTest* package (Kuznetsova et al., 2016). The results showed that reduction had a significant lengthening effect on /ʧ/ (11.6 ms, $p = 0.007$). The fricatives on the other hand showed varying effects. Reduction had a non-significant lengthening effect of 1.7 ms on /ɸ/ ($p = 0.691$) and non-significant shortening effects of 3.3 ms on /ç/ ($p = 0.447$) and 4.9 ms on /ʃ/ ($p = 0.114$). However, reduction had a significant shortening effect of 11.4 ms on /s/ ($p = 0.008$).

A separate linear mixed effects regression model was fit to low predictability tokens (i.e., /k, ʃ/) to investigate the effects of vowel type. Since the overall model above already showed that reduction had a lengthening effect on /k/, the baseline was set to /ʃ/. Reduction status, C_1, vowel type, and their interactions were included as predictors. Random intercepts by participant and word were included. By-participant random slopes for reduction, C_1, and vowel type were also included, as well as by-word random slopes for gender. The final model retained the full random effects structure. The three-way interaction term and the reduction:vowel interaction were not significant contributors to the model ($p = 0.7549$ and 0.1262, respectively) and were removed from the fixed effects structure of the final model.

The results of the final model showed that although reduction had a slight shortening effect of 5 ms and the vowel /u/ had a slight lengthening effect of 3 ms on /ʃ/, neither was significant ($t =$ -1.522 and 1.061, respectively). Also, as was shown in the overall model above, reduction had a significant lengthening effect of 22 ms on /k/ ($t = 6.496$).

3.2.2.2 Summary of duration results

For duration, gender did not have a significant effect. Reduction, on the other hand, had opposite effects depending on manner and C_1 duration. In terms of manner, reduction had a lengthening effect on the two non-fricative C_1 /k/ and /ʧ/. Reduction actually showed a significant shortening effect for

/s/ tokens, and no effect on the rest of the fricatives tested in this study /ʃ, ɸ, ç/. Particularly, the fact that /ʧ/ lengthened but not /ɸ, ç/ despite being similar in length suggests that lengthening is largely dependent on whether the consonant is a continuant or not. However, inherent C_1 duration also had an effect on the magnitude of the duration effects. Both /k/ and /ʧ/ lengthened, but the shorter segment /k/ lengthened more (22 ms vs. 12 ms), a difference that was shown to be significant (t = -2.148). For the fricatives, /s/ shortened while the shorter fricatives did not. Additionally, despite the fact that /s/ and /ʃ/ both have similar durations of $\sim$100 ms, only /s/ shortened significantly, suggesting that there may be an effect of predictability as well.

3.2.3 Center of gravity (COG)

COG is sensitive primarily to changes in the front cavity (Nittrouer and McGowan, 1989), so the effects of gestural changes are expected to differ depending on the coarticulated vowel and the C_1 place of articulation. In general, however, increased coarticulation with the high back vowel /u/ is expected to have a lowering effect for all C_1 due to lip rounding, which would lengthen the front oral cavity. Although the high back vowel of Japanese has traditionally been regarded as unrounded (i.e., [ɯ]), a recent articulatory study by Nogita et al. (2013) has shown that the high back vowel is actually closer to a high central rounded vowel [ʉ], at least in younger speakers. On the other hand, the high front vowel is expected to have different effects depending on how front in the oral cavity the C_1 place of articulation is, making direct statistical comparisons impractical. This section therefore analyzes each C_1 separately. /ʃ/ is analyzed first because it is the only fricative that can be tested for both vowel and reduction effects. The /ʃ/ results are then used as reference for interpreting all other C_1.

A linear mixed effects regression model was fit for all statistical analyses. Unless noted otherwise, the random effects structure for the fully complex model included random intercepts for

participants and words, by-participant random slopes for vowel and reduction, and by-word random slopes for gender.

3.2.3.1 /ʃ/ COG results and analysis

For /ʃ/, the COG values for /u/ tokens are expected to be lower than for /i/ tokens regardless of reduction, similar to Beckman and Shoji (1984). COG2 is also expected to be lower than COG1 in unreduced tokens for both vowels, as lip rounding increases for /u/ articulation or the tongue shifting back towards the palate for /i/, both of which would lengthen the front oral cavity. Given the expected lowering effect of coarticulation for both vowels, there are a three possible effects of reduction. First, reduced tokens may show increased coarticulation between C_1 and the target vowel, resulting in lower COG values. Second, reduced tokens may show decreased coarticulation, leading to higher COG values. Third, reduced and unreduced tokens may show similar degrees of C_1V coarticulation, showing no difference in COG values.

Shown below in Table 3.6 are the COG1 and COG2 values of /ʃ/. The mean COG values are lower when the vowel is /u/ for both COG1 and COG2 as expected. Male participants also have lower COG values overall, which is unsurprising given than men generally have longer vocal tracts than women. Reduction also seems to have a lowering effect on COG1 when the vowel is /i/, but not when the vowel is /u/.

C_1V		unreduced		reduced	
		female	*male*	*female*	*male*
ʃi	COG1	5694 (*622*)	4996 (*488*)	5201 (*862*)	4738 (*619*)
	COG2	5317 (*489*)	4592 (*342*)	5317 (*926*)	4695 (*787*)
ʃu	COG1	4948 (*606*)	4403 (*384*)	4924 (*758*)	4504 (*470*)
	COG2	4469 (*508*)	4060 (*409*)	4555 (*842*)	4311 (*633*)

Table 3.6: COG1 and COG2 mean (*standard deviation*) in Hz for /ʃ/.

The final model fit to the COG1 results of /ʃ/ retained the full random effects structure. The following non-significant predictors were removed from the final model: three-way interaction ($p =$

0.5353) and gender:vowel interaction ($p = 0.3846$). The final model's results are summarized below in Table 3.7, with the baseline set as unreduced /ʃi/ tokens produced by female participants.

	Estimate	Std. Error	*t*	
(Intercept)	5625.1	140.7	39.97	*
/u/	-631.0	126.6	-4.98	*
reduced	-448.3	132.7	-3.38	*
male	-606.3	159.4	-3.80	*
reduced:/u/	358.9	163.5	2.20	*
reduced:male	199.4	73.0	2.73	*

Table 3.7: Linear mixed effects regression results: COG1 of /ʃ/.

The results of the model show that COG1 is significantly lower for /ʃu/ tokens compared to /ʃi/ tokens, suggesting that coarticulation with /u/ begins early in unreduced tokens. Additionally, reduction has a significant lowering effect. Since the model's baseline was unreduced /ʃi/ tokens, lower COG1 suggests that the front oral cavity is larger in reduced [ʃi̥] tokens than in unreduced [ʃi] tokens. In other words, coarticulation with /i/ begins earlier when the vowel is reduced. The lowering effect of reduction is significantly smaller for male participants and when the vowel is /u/, however. Differences of least squares means of the model revealed that the effects of reduction are in fact not significant for the male participants ($p = 0.429$) and when the vowel is /u/ ($p = 0.932$). In other words, reduced tokens do not show evidence of increased coarticulation in male participants and for [ʃu̥] tokens. Lastly, the results also show that COG1 is significantly lower for male participants.

For COG2, the full random effects structure was retained, and the following predictors were removed from the final model as they did not improve the fit of the model: three-way interaction ($p = 0.4223$), reduction:vowel interaction ($p = 0.5073$), reduction:gender interaction ($p = 0.2178$), and reduction ($p = 0.3771$). The results of the final model are summarized below in Table 3.8, with /ʃi/ tokens produced by female participants as the baseline. The fact that reduction was not a significant predictor means that the /ʃi/ tokens show comparable degrees of coarticulation with /i/ by the end

of the consonant. There still was a significant lowering effect of /u/, however, suggesting that /u/ coarticulation begins early as shown in the COG1 results and remains throughout the consonant for both reduced and unreduced tokens.

	Estimate	Std. Error	t	
(Intercept)	5343.6	127.1	42.04	*
/u/	-795.4	156.9	-5.07	*
male	-752.8	116.5	-6.46	*
/u/:male	313.9	118.0	2.66	*

Table 3.8: Linear mixed effects regression results: COG2 of /ʃ/.

Lastly, a linear mixed effects model was fit to the ΔCOG (COG2 $-$ COG1) data to check whether the change from COG1 to COG2 was significantly different between reduced and unreduced tokens. The final model retained the full random effects structure, and the following non-significant predictors were removed from the fixed effects structure of the final model: three-way interaction ($p = 0.3216$), reduction:vowel interaction ($p = 0.8130$), reduction:gender interaction ($p = 0.6935$), and reduction ($p = 0.1653$). The results of the final model are shown below in Table 3.9, with /ʃi/ tokens produced by female participants as the baseline. The intercept of the final model was not significantly different from zero. Gender and vowel were also not significant, suggesting COG1 and COG2 are not significantly different from each other. The interaction term for gender and vowel was significant, suggesting that the fall in COG for /u/ is significantly smaller for male participants than for female participants. However, a separate analysis of the male participants showed that, like the female participants, the change in COG for /u/ tokens were not significantly different from /i/ tokens ($t = 0.564$).

	Estimate	Std. Error	t	
(Intercept)	-124.92	155.24	-0.805	
/u/	-290.05	212.80	-1.363	
male	-100.97	85.07	-1.187	
male:/u/	244.59	92.58	2.642	*

Table 3.9: Linear mixed effects regression results: ΔCOG (COG2 - COG1) of /ʃ/.

The non-significant results of ΔCOG seemingly contradicts the significant effect of reduction on COG1. A separate analysis of reduced and unreduced tokens revealed that while the intercept term for ΔCOG was not significantly different from zero for reduced tokens ($t = 0.403$), unreduced tokens did show a significant fall of 365.01 Hz (t = -5.366).

3.2.3.2 /ʧi, su/ COG results and analyses

COG2 is expected to be lower than COG1 for the affricate /ʧ/ as it begins with an alveolar constriction and moves back towards the alveo-palatal region for an /ʃ/-like frication. The possible effects of reduction are similar to that of /ʃi/ tokens: increased coarticulation would result in further lowering of COG values as the tongue shifts back towards the palate for /i/, while decreased coarticulation would lead to higher COG values. For /s/, increased coarticulation with /u/ would also lead to lower COG values, since lip rounding would lengthen the front oral cavity.

Shown below in Table 3.10 are the COG1 and COG2 values of /ʧ, s/. The overall pattern seems to be that reduction does not have a significant effect, suggesting that the degree of coarticulation is not different between reduced and unreduced tokens. Additionally, male participants have lower values for /ʧ/, but there seems to be no significant gender effect on /s/.

C_1V		unreduced		reduced	
		female	*male*	*female*	*male*
ʧi	COG1	6397 (*687*)	5350 (*588*)	6185 (*751*)	5186 (*587*)
	COG2	5594 (*456*)	4803 (*406*)	5468 (*1102*)	4822 (*898*)
su	COG1	6118 (*1464*)	6154 (*879*)	6032 (*1196*)	5976 (*834*)
	COG2	6125 (*1076*)	6046 (*797*)	6026 (*1347*)	5977 (*1106*)

Table 3.10: COG1 and COG2 mean (*standard deviation*) in Hz for /ʧ, s/.

For /ʧ, s/, the fully complex model included reduction status, gender, and their interaction as predictors. The maximal random effects structure included random intercepts for participant and word, by-participant random slopes for reduction status, and by-word random slopes for gender.

For /ʧ/, the final models for COG1, COG2, and ΔCOG retained the full random effects structure and only gender as a predictor. Neither reduction nor the reduction:gender interaction were significant contributors to the COG1 model ($p = 0.1520$ and 0.9884, respectively), the COG2 model ($p = 0.9773$ and 0.4069, respectively), and the ΔCOG model ($p = 0.6599$ and 0.4337, respectively). The fact that reduction does not play a significant role in the COG results suggest that the degree of coarticulation between /ʧ/ and /i/ does not increase for reduced tokens, unlike /ʃi/. On the other hand, the results of the final models showed that male participants had lower values for both COG1 (-1006 Hz; t = -5.16) and COG2 (-760 Hz; t = -7.0), which was also the case for /ʃi/. The ΔCOG model also had a significant non-zero intercept at -745 Hz (t = -5.033), showing that COG2 is significantly lower than COG1 as predicted, regardless of reduction. Male participants were also shown to have a significantly smaller degree of change, where the drop in COG was 441 Hz. A separate analysis for just the male participants showed that the smaller lowering effect was still significant at t = -3.243.

For /s/, the final models for COG1, COG2, and ΔCOG retained the full random effects structure but none of the predictors. The fact that an intercept-only model fit the data equally well as a model with predictors shows that COG values for /s/ do not change throughout the consonant regardless of gender or reduction.

3.2.3.3 COG results and analyses of /h/ allophones

Although /ɸ, ç/ can contrast with /h/ depending on the lexical stratum (Ito and Mester, 1995; Moreton and Amano, 1999), /ɸ/ and /h/ neutralize to [ɸ] before /u/, while /ç/ and h/ neutralize to [ç] before /i/ across all strata. Because /ɸ, ç/ are essentially identical in place with their respective neutralizing vowels, changes in COG are expected to come primarily from constriction strength rather than change in the length of the front oral cavity[3], where weakening constriction lowers the amplitude of the higher frequencies and results in a lower COG value overall (Hamann and Sennema, 2005; Kiss and Bárkányi, 2006). In other words, for /ɸ/, an increased coarticulation with /u/ would result in more lip rounding and weaker constriction, both contributing to lower COG values. For /ç/, increased coarticulation with the vowel would make the fricative more vowel-like with a weaker constriction, also resulting in lower COG values.

The COG1 and COG2 results are summarized in Table 3.11 below. The overall pattern is that COG falls for unreduced /ɸu/ tokens but rises for unreduced /çi/. On the other hand, COG measures for reduced /ɸu, çi/ tokens both rise.

C_1V		unreduced		reduced	
		female	*male*	*female*	*male*
ɸu	COG1	1926 (*559*)	2014 (*824*)	2703 (*721*)	2872 (*1055*)
	COG2	1879 (*541*)	2044 (*891*)	2931 (*913*)	2816 (*1091*)
çi	COG1	3706 (*931*)	3682 (*833*)	4009 (*701*)	3793 (*760*)
	COG2	3951 (*919*)	3723 (*852*)	4579 (*673*)	4210 (*728*)

Table 3.11: COG1 and COG2 mean (*standard deviation*) in Hz for /h/.

For /ɸ/, the final models for COG1 and COG2 retained the full random effects structure and only reduction as a predictor. Neither gender nor the reduction:gender interaction were significant contributors to the COG1 model ($p = 0.5886$ and 0.6246, respectively) and the COG2 model ($p =$

[3]Although, see Kumagai (1999) whose EPG study found that palatal constriction is more fronted before reduced vowels for [ɸ].

0.4454 and 0.0916, respectively), and thus were removed from the final model. Reduction had a significant raising effect of 813 Hz on COG1 ($t = 5.418$) and 905 Hz on COG2 ($t = 4.097$).

The final model for /ɸ/ ΔCOG retained the fully complex fixed and random effects structures. The results are summarized in Table 3.12 below. Unreduced tokens produced by female participants are the baseline. The fact that the intercept is not significantly different from zero suggests that COG1 and COG2 are not significantly different from each other for female participants in unreduced tokens. Reduction and gender did not have a significant effect on how the COG values change. However, the interaction term shows that COG rises less for male participants in reduced tokens. A separate analysis for the male participants showed that the change in COG is in fact not significant for the male participants (t = -0.502).

	Estimate	Std. Error	t value	
(Intercept)	-46.54	155.40	-0.300	
reduced	275.26	222.93	1.235	
male	82.03	77.49	1.059	
reduced:male	-366.14	122.12	-2.998	*

Table 3.12: Linear mixed effects regression results: ΔCOG of /ɸ/.

For /ç/, the final model for COG1 retained the full random effects structure but only the intercept for fixed effects. This means that /ç/ COG1 values were unaffected by either gender or reduction at the start of the consonant. The final model for COG2, however, retained reduction and gender but not their interaction for its fixed effects structure, which showed that reduction has a significant raising effect of 580 Hz towards the end of the consonant ($t = 3.704$) and that COG2 was lower for male participants by 349 Hz (t = -2.300).

For ΔCOG of /ç/, the model with the full random effects structure failed to converge. The fit of a model with only by-participant random slopes for reduction did not significantly differ from a model with only by-word random slopes for gender ($p = 1.000$). Since reduction was the only significant predictor for both COG1 and COG2, the random effects structure with by-participant

random slopes for reduction was selected for the final model. For the fixed effects structure, the reduction:gender interaction was not a significant contributor to the model ($p = 0.6981$) and was removed from the final model. The intercept for the final ΔCOG model was significantly higher than zero. In other words, COG2 was higher than COG1. Reduction, however, had a significant increasing effect of 794 Hz ($t = 5.033$), showing that COG rose even more in reduced tokens. Additionally, COG rose less for male participants.

	Estimate	Std. Error	t value	
(Intercept)	236.38	70.56	3.350	*
reduced	348.43	90.42	3.854	*
male	-185.76	71.99	-2.580	*

Table 3.13: Linear mixed effects regression results: ΔCOG of /ç/.

The rising COG pattern for /çi/ is somewhat unexpected, since under the prediction stated above, a rising pattern would suggest a weakening coarticulation. A separate analysis of just the male participants revealed that ΔCOG for men actually was not significantly different from zero (41 Hz; $t = 0.400$), but reduction had a raising effect of 376 Hz ($t = 2.404$). In other words, the rising effect in unreduced tokens is present only in female participants. A closer examination of /çi/ tokens as produced by female participants revealed a pattern not observed in male participants or other fricatives. For female participants, the aperiodic noise for /çi/ typically began quite diffuse, with the lower frequencies gradually being lost until the concentration of the aperiodic noise stabilized just before the onset of /i/, contributing to the overall rise in COG. Examples of a typical waveform and spectrogram of unreduced /çi/ tokens for female participants (left) and male participants (right) are shown below in Figure 3.5 below from the word [çidoi] ‘harsh’.

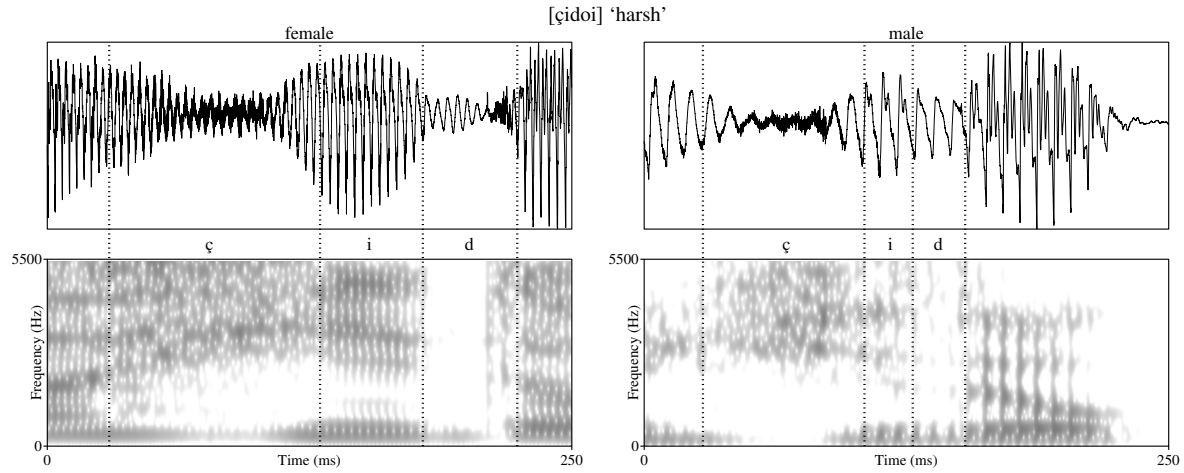

Figure 3.5: Waveform and spectrogram of unreduced C_1V in [çidoi], showing landmarks for C_1, vowel, and C_2 duration.

Although an articulatory study is required to verify the exact nature of /ç/ articulation in Japanese, there are two likely gestural explanations for the observed spectral pattern of /ç/ in female participants. First, the observed pattern is consistent with the lingual gesture for /ç/ beginning more [h]-like, further back in the oral cavity, then transitioning forward into the palatal place of articulation. If true, this suggests that the neutralization between /h/ and /ç/ before /i/ is actually incomplete, at least for female speakers. Second, the pattern is also consistent with /ç/ starting as a weak palatal fricative, whose constriction strengthens over time. A combination of the two explanations is also consistent with the observed spectral pattern, both of which suggest that female and male speakers are using different variants of the /h/ allophone. Regardless of which of the explanations is correct, the fact that the COG of /ç/ for reduced tokens continues to rise significantly beyond that of unreduced tokens, suggests that there is no anticipatory /i/ articulation that intervenes and halts the continued increase in the degree of palatal constriction between C_1 /ç/ and C_2 in reduced tokens.

3.2.3.4 COG results and analysis /k/

For /k/, COG measures were taken from the burst with a 20 ms window centered at the middle o the burst noise. Coarticulation with the high back vowel would lower the COG measures of /k/ due to the lengthened oral cavity through lip rounding. Increased coarticulation would lead to further lowering of COG. For the high front vowel, the COG will be higher as palatalization moves the tongue body forward, shortening the front oral cavity.

	unreduced		reduced	
C_1V	*female*	*male*	*female*	*male*
ki	4707 (*634*)	4353 (*529*)	4727 (*852*)	4378 (*614*)
ku	1871 (*682*)	1708 (*526*)	2270 (*665*)	2277 (*877*)

Table 3.14: COG mean (*standard deviation*) in Hz for /k/.

To analyze the /k/ tokens, a model with the most complex fixed and random effects structures was fit first. The predictors were reduction, vowel, gender, and their interactions. The random effects structure included random intercepts for participant and word, by-participant random slopes for reduction and vowel, and by-word random slopes for gender. The final model retained the full random effects structure. The following non-significant predictors were removed from the fixed effects structure of the final model: three-way interaction ($p = 0.3106$), reduction:vowel interaction ($p = 0.3754$), reduction:gender interaction ($p = 0.4037$), gender:vowel interaction ($p = 0.1042$), and gender ($p = 0.2817$). The results of the final model are summarized in Table 3.15 below. Unreduced /ki/ tokens are the baseline. Reduction had a significant raising effect 272 Hz, and the vowel /u/ had a significant lowering effect of 2505 Hz.

	Estimate	Std. Error	t	
(Intercept)	4431.73	122.00	36.33	*
reduced	272.35	77.73	3.50	*
/u/	-2504.55	117.29	-21.35	*

Table 3.15: Linear mixed effects regression results: COG of /k/.

3.2.3.5 Summary of COG results

The mean COG values for all C_1 are summarized in Table 3.16 below. C_1 that showed significantly more coarticulation for reduced tokens are noted with a '+' after the C_1V label, while C_1 that showed significantly less coarticulation for reduced tokens are noted with a '-' after the C_1V label. Also, /ʃ, k/ both showed significant lowering of COG when followed by /u/ compared to /i/.

C_1V			unreduced		reduced	
			female	*male*	*female*	*male*
ʃi	+	COG1	5694 (*622*)	4996 (*488*)	5201 (*862*)	4738 (*619*)
		COG2	5317 (*489*)	4592 (*342*)	5317 (*926*)	4695 (*787*)
ʃu	+	COG1	4948 (*606*)	4403 (*384*)	4924 (*758*)	4504 (*470*)
		COG2	4469 (*508*)	4060 (*409*)	4555 (*842*)	4311 (*633*)
ʧi		COG1	6397 (*687*)	5350 (*588*)	6185 (*751*)	5186 (*587*)
		COG2	5594 (*456*)	4803 (*406*)	5468 (*1102*)	4822 (*898*)
su		COG1	6118 (*1464*)	6154 (*879*)	6032 (*1196*)	5976 (*834*)
		COG2	6125 (*1076*)	6046 (*797*)	6026 (*1347*)	5977 (*1106*)
ɸu	−	COG1	1926 (*559*)	2014 (*824*)	2703 (*721*)	2872 (*1055*)
		COG2	1879 (*541*)	2044 (*891*)	2931 (*913*)	2816 (*1091*)
çi	−	COG1	3706 (*931*)	3682 (*833*)	4009 (*701*)	3793 (*760*)
		COG2	3951 (*919*)	3723 (*852*)	4579 (*673*)	4210 (*728*)
ki	−		4707 (*634*)	4353 (*529*)	4727 (*852*)	4378 (*614*)
ku	−		1871 (*682*)	1708 (*526*)	2270 (*665*)	2277 (*877*)

Table 3.16: COG1 and COG2 mean (*standard deviation*) in Hz for all C_1. '+' = decreased coarticulation in reduced tokens, '−' = increased coarticulation in reduced tokens, **bold** = significant effect of vowel.

Center of gravity measurements for /ʃ/ tokens showed that coarticulation with /u/ begins early in the consonant regardless of reduction status. These early difference in COG between /ʃi/ and /ʃu/ tokens suggest an attempt to maximize recoverability through increased coarticulation. Furthermore, coarticulation with the following vowel was also shown to begin earlier in reduced /ʃi/ tokens. This effect was not found in /ʃu/ tokens, however, but given that /ʃi, ʃu/ differences are already apparent early in the consonant, further increasing the coarticulation for /ʃu/ perhaps is

not necessary for recovery. In other words, only modifying one of the /ʃ/ contexts is enough to distinguish the two vowel possibilities.

/ʃi/ results can be compared directly with /ʧi/ results, since the two consonants share a place of articulation. Unlike /ʃi/, however, COG results for /ʧi/ did not show any effect of reduction, suggesting that there was no increase of coarticulation to aid recoverability. The same was true of /s/ tokens, which showed no effect of reduction.

An increase in coarticulation was expected to lower COG values for the two allophones of /h/, and both allophones in fact had higher COG values for reduced tokens, suggesting a decrease in coarticulation for reduced tokens. In other words, vowel gestures were not maintained to the same degree as in unreduced tokens perhaps because the allophonic variation was sufficient for recovery.

Lastly, the single COG measurement for /k/ showed that there is a difference of 2500 Hz between /ki/ and /ku/ tokens, which is more than six times greater than the 400 Hz difference that Japanese listeners were shown to be sensitive to for /ʃ/ in Beckman and Shoji (1984) and more than four times greater than the difference of ∼600–800 Hz found in the current study.

3.2.3.6 Loanwords

Comparison of /s/-initial loanwords to /s/-initial non-loanwords showed no difference in terms of reduction rates, duration, or COG. Reduction rates for the loanwords were 100%, like the non-loanword tokens. A linear mixed effects regression model was fit to compare the duration of loan and non-loan tokens with token type, gender, and their interaction as predictors. The random effects structure included random intercepts for participants and words, by-participant random slopes for token type, and by-word random slopes for gender. A Chi-square log likelihood ratio test showed that none of the predictors were significant contributors to the model. The duration of /s/ in reduced non-loanwords and reduced loanwords, therefore, were not different. Similar results were found for COG measurements, where none of the predictors were significant contributors to the models fit for COG1, COG2, and ΔCOG results.

3.3 Discussion

The aim of this chapter was to investigate the acoustic properties of high vowel reduction in Japanese – specifically, what cues in the signal allow the recovery of a reduced vowel and whether gender and predictability from context affect the availability of these cues. The cues specifically tested for were coarticulatory effects of the target vowel on C_1, measured in the form of burst/frication duration and center of gravity (COG) of C_1.

With respect to the issue of lengthening, duration measurements showed that lengthening is observable only in non-fricatives. Reduction generally had no effect on fricatives, with the exception of /s/ which shortened in reduced contexts instead. The fact that C_1 lengthening is dependent on the manner of the consonant suggests that it is not an obligatory process whose goal is to maintain mora-timing (Han, 1994). Furthermore, the fact that [ʧ] lengthened while [ɸ, ç] did not despite similar durations suggests that C_1 lengthening is not a recoverability-conditioned process.

On the other hand, reduced /s/ tokens showed significant shortening while /ʃ/ did not, despite similar durations of ∼100 ms. Based on the COG results, it can be safely assumed that vocalic oral gestures are actively retained for /ʃ/, but less so for /s/. When both duration and COG results are considered together, the reason for shortening in /s/ is likely due to a lack of an intervening vowel gesture, which shortens the gestural timing for C_1 to C_2 transition. In other words, there is perhaps vowel deletion rather than devoicing when a high vowel is reduced after /s/. The question arises as to why it is only /s/ that shortens and none of the other obstruent C_1. One possible explanation is that reduced /sV.C_2V/ sequences are in fact being resyllabified as [sC_2V] with the initial /s/ being treated as part of an onset cluster, whereas other obstruents lengthen or remain unchanged because they are treated as belonging to a separate syllable (e.g., /hu.ku/ → [ɸ.ku]). In light of recent articulatory work by Kawahara et al. (2016), where it was found that Japanese speakers tend to retain the oral gestures of high vowels in reducing contexts either completely or not at all, the formation of clusters seems unproblematic. As to why only /s/ resyllabifies, it may have to do with special cluster-forming

properties of /s/, which, although will not be discussed further here, has long been noted in previous works (Selkirk, 1982; Kaye, 1992; Gierut, 1999; Morelli, 1999; Barlow, 2001).

COG results suggest that there is an effect of predictability on how early vowel coarticulation begins in Japanese high vowel reduction. The high predictability tokens showed either no change in C_1V coarticulation as in /ʧ, s/ or a possible decrease in /ɸ, ç/. For these consonants, the phonetic cues associated with the vowel are not essential for the recovery of the underlying high vowel when C_2 is voiceless because the reducible vowel that can follow is fully predictable. In other words, enhancing coarticulatory cues between C_1 and the target vowel does little to increase the likelihood of recovery.

The results for /ʃ/ on the other hand show that this is not the case for low-predictability contexts. Since both /i, u/ can follow the consonant, complete deletion of the vowel in these cases would jeopardize the recoverability of the vowel. Additional articulatory effort is required to transmit the contrastive information necessary for vowel recovery. As the COG results in this study and the spectral analysis in Beckman and Shoji (1984) have shown, oral gestures alone are enough to color the burst/frication noise of C_1 for reliable recovery. By retaining and overlapping the oral vowel gesture with the preceding consonant, maximal recoverability is obtained even in the absence of phonation. The idea that overlap of gestures are coordinated in order to preserve recoverability has been proposed by Silverman (1997) and Chitoran et al. (2002), and it was also suggested by Varden (2010) for Japanese.

The results for /k/ were less straightforward. First, /u/ had a significant lowering effect compared to /i/ like in the case of /ʃ/. The large spectral difference is most likely due to /k/-fronting that results from coarticulation with the following /i/, and positing the presence of coarticulatory effects even in reduced tokens allows /k/ to be grouped with /ʃ/. However, reduction also had a raising effect, suggesting a weakening coarticulation with the target high vowels, much like the two allophones of /h/, which are high predictability fricatives. A possible explanation as to why /k/ seemingly patterns with the high predictability consonants for reduced tokens is the large COG

difference of 2,500 Hz between the burst noises of /ki/ and /ku/. This difference is nearly four times the differences of ∼600–800 Hz observed for /ʃ/ in the current study and nearly six times the 400 Hz spectral difference reported in Beckman and Shoji (1984), which Japanese speakers were shown to be sensitive to. Given such a large spectral difference, loss of some coarticulatory cues are unproblematic for recovery, since the difference is already quite obvious.

Lastly, gender did not seem to have an effect on the acoustic results, although male participants were shown to reduce more than the female participants, which confirms what Imai (2010) also found in younger speakers. What is interesting from the reduction results, however, is where the observed difference between men and women came from. With tokens in reducing environments having reduction rates of essentially 100%, the difference in reduction rates was clearly from the non-reducing tokens. An analysis of just the non-reducing tokens showed that reduction rates were significantly different for high predictability environments but not low predictability environments. In other words, predictability also seems to affect reduction rates, although only in men.

3.4 Conclusion

The results of the production experiment suggest that speakers provide more coarticulatory information on C1 burst/frication when the target vowel is unpredictable (i.e., after /k, ʃ/), supporting the results of previous studies which showed that the degree of coarticulation between segments are controlled to aid recoverability (Silverman, 1997; Chitoran et al., 2002). The results also provide novel insight into recoverability-driven coarticulation in that speakers not only increase the perceptibility of coarticulatory information when recoverability is in jeopardy (i.e., after /ʃ/) but that they also do the opposite, where speakers provide less or an unchanged amount of coarticulatory information when the normal amount of coarticulation is already highly perceptible (i.e, after /k/) or the vowel is highly predictable (i.e., after /ʧ, ɸ, s, ç/) and additional coarticulatory cues are unnecessary for recovery. Chapter 4 will look at whether these cues that the speakers seem to be

actively manipulating depending on their predictability during production are similarly used by listeners during perception.

CHAPTER 4

Phonotactic effects on sensitivity to phonetic cues

4.0 Introduction

If it is the case that speakers are varying the amount of coarticulatory information depending on contextual predictability as shown in Chapter 3, the question that naturally follows is, "Do listeners utilize the extra information during perception?" Put differently, is it the case that listeners are more attentive to coarticulatory cues in low-predictability contexts than in high-predictability contexts? Previous studies on recoverability have largely focused on gestural timing and the resulting changes in the acoustic signal that allow easier recovery of a target segment, with assumption that this is done for the benefit of the listeners. With new insight that phonotactic predictability also affects whether the perceptibility of phonetic cues are enhanced or weakened, this chapter presents a perception experiment that specifically controls predictability and investigates how much more or less sensitive to coarticulatory cues Japanese listeners are depending on predictability from context.

4.0.1 Perceptual recovery in Japanese

Before discussing the experiment itself, some overview on what is known about recoverability and perception in Japanese should be discussed. In the now well-known study commonly referred to as the "*ebzo* test" (Dupoux et al., 1999), French and Japanese speakers were presented with acoustic stimuli with the high back rounded vowel [u] of varying lengths occurring between two consonants, ranging from 0 ms to 90 ms (e.g., [ebzo] → [ebu:zo]). The stimuli were designed so that when there is no vowel in the stimuli, the result is a phonotactically legal sequence in French but illegal in Japanese. The task was to identify whether the vowel [u] was present in the stimuli. In the same study were three other experiments which utilized a similar set of stimuli, where without the medial vowel, the resulting sequence is legal in French but illegal in Japanese. The first of the three experiments was a gating task similar to the one described above, but with additional stimuli that sought to control for any effects of coarticulation on misperceiving a vowel. The remaining two experiments were match-to-sample (ABX) tasks, where the participants had to decide whether the third stimulus was the same as the first or second. The two ABX tasks were designed to test, in addition to vowel-less and vowel-ful distinctions, how vowel length is perceived. The results of all the experiments showed that while French speakers could accurately distinguish the vowel-less from the vowel-ful tokens, Japanese speakers were essentially "deaf" to such differences, erring heavily towards misperceiving—or for the purposes of this chapter, *recovering*—what the authors call an "illusory" vowel. On the other hand, French speakers were unable to accurately perceive vowel length, with which the Japanese participants had little trouble perceiving.

The authors propose that the results are due to the different phonotactics of French and Japanese. The argument is that Japanese listeners perceive a non-existent vowel between two consonants in vowel-less tokens because Japanese phonotactics disallows heterorganic consonant clusters; French listeners, on the other hand, are insensitive to vowel length because it is not contrastive in French. In other words, the authors argue, one's native phonotactics strongly biases

the listener towards mapping a non-native, phonotactically illegal sequence to one that is legal from the earliest stages of perception rather than at a higher, abstract phonological level.

A more recent, follow-up study (Dupoux et al., 2011) aimed to further bolster this claim by also investigating Portuguese speakers. There were two experiments in this paper, a gating task and an ABX task, much like in the 1999 work. The participants in this study were monolingual speakers of Brazilian Portuguese, European Portuguese, and Japanese. The reason for choosing the two dialects of Portuguese was that while European Portuguese allows the same types of clusters as French, Brazilian Portuguese does not, leading to the expectation that their perception would be similar to that of Japanese speakers. The crucial difference between Brazilian Portuguese and Japanese is that in the former, the default epenthetic vowel is /i/ as opposed to the Japanese /u/. Since the quality of the epenthetic vowels are different in the two epenthesizing languages, the experiments were modified slightly to enable identification of the perceived illusory vowels in the results. Like the French listeners in the original 1999 study, European Portuguese listeners did not have trouble distinguishing vowel-less from vowel-ful tokens. Japanese listeners, again, showed a tendency towards perceptually recovering /u/ between illegal consonant clusters. The results, however, additionally showed that Japanese listeners were also sensitive to [i]-coarticulation, which conflicts with default /u/. By comparison, Brazilian Portuguese listeners tended to perceptually recover /i/ between illegal consonant clusters by default, but did not show the same degree of sensitivity to [u]-coarticulation. The difference is likely due to Brazilian Portuguese listeners have little experience with a systematic high vowel reduction process, and thus underutilizing coarticulatory cues relative to Japanese listeners.

In addition to the two behavioral studies, Dehaene-Lambertz et al. (2000) also tested the illusory vowel epenthesis effect in an event-related potential (ERP) study. In this study, Dehaene-Lambertz et al. carried out experiments similar to that of Näätänen et al. (1997), where electrophysiological responses have been shown to be sensitive to phoneme categories. Dehaene-Lambertz et al. looked at how mismatch negativity (MMN) responses in Japanese and French speakers differ

in the absence versus presence of a vowel in the same kind of sequences as those in Dupoux et al. (1999). The experiments followed an oddball paradigm where in one trial a sequence that is legal in both languages is presented as the standard (e.g., [igumo]) and one that is illegal only in Japanese as the deviant (e.g., [igmo]). The reverse is presented in a separate trial. Although the results reported collapsed the trials, the ERP results showed that Japanese speakers are insensitive to the differences between the vowel-ful and vowel-less items, while French speakers are, supporting the behavioral results from the original study by Dupoux et al. (1999). A related fMRI study by Jacquemot et al. (2003), also report similar but slightly weaker results. Jacquemot et al. report that in an AAX task (*A*-stimulus presented twice before *X*-stimulus) neural activity increased whenever the X stimulus was different from the A stimulus for both Japanese and French participants. This was true regardless of whether or not the acoustic difference was phonologically contrastive in the language, although neural activation was significantly greater when the acoustic contrasts were also phonologically contrastive.

4.0.2 Problems and solutions

The studies discussed above paint a seemingly convincing picture, where "illusory" vowels are the result of the listeners' inability to represent illegal sequences accurately. There are, however, a number of refinements that the experiments would benefit from. The experiment presented in this chapter addresses the following two issues with the original studies: the acoustic stimuli used might have biased Japanese listeners toward perceiving a vowel, and the stimuli collapsed reducing and non-reducing contexts.

First, the waveform and spectrogram examples of the stimuli used in the studies by Dupoux and colleagues reveal that the burst of C_1 (e.g., [b] in [ebzo, ebuzo]) were rather long, potentially biasing the participants to perceive a vowel. For example, Dupoux et al. (2011) show that in a sequence like [agno], the voiced stop had a burst of at least 50 ms and contained formant-like

structures. Japanese voiced stops, on the other hand, typically have a burst duration of less than 20 ms (Kong et al., 2012). In addition, Japanese high vowels are inherently short, with an average duration of approximately 40 ms, but they can be as short as 20 ms (Han, 1994; Beckman, 1982). Taking the short burst and vowel durations of Japanese together, an atypically long burst can be interpreted as containing a vowel, possibly confounding the independent effects of acoustic cues and phonotactic violations (Wilson et al., 2014). Furthermore, the stop closure of the stop is also nearly 100 ms, which is closer to the geminate range than the singleton range in Japanese (Kawahara, 2006). Geminates are not known to affect high vowel reduction in C_1 position, but geminate consonants in C_2 position have been shown to increase the likelihood of preceding vowels being phonated in Japanese regardless of voicing feature of the consonant (Maekawa and Kikuchi, 2005; Fujimoto, 2015). This means that in stimuli with obstruents in both C_1 and C_2 positions (e.g., [igba]) could have further biased Japanese participants toward expecting a vowel in the target context. While this is also a tendency that is language-specific and phonotactically driven, it is unclear whether the primary driving force behind perceptual epenthesis is the heterorganic clusters or the geminate-like phonetic cues. It seems likely that although phonotactic constraints may certainly be playing a role in perceptual repair, this effect was also confounded by acoustic cues in the stimuli that have been overlooked.

Second, the stimuli used in the *ebzo* tests included a mix of environments in which high vowel reduction is expected to occur in Japanese (i.e., when between two voiceless obstruents) as well as non-reducing environments. The results reported in these studies, however, make no distinction between the two types of environments. This is a peculiar decision since high vowel reduction, a process in which the vowel is often reduced to nothing, is extremely relevant to the issue of L1 phonotactic effects on illegal sequence perception. Because high vowel reduction is highly productive, it is very likely that this phonological process had an effect that is independent of phonotactic constraint violations in creating an expectation for a vowel. An obvious solution to this issue is to test and analyze the two environments separately (e.g., [ezpo] vs. [espo]). In addition, the

stimuli group in which reduction is expected to occur can be further divided into high-predictability and low-predictability sub-groups (e.g., [eʃpo] and [espo], respectively) based on the findings of the previous chapter.

4.1 Materials and methods

The experiment presented in this chapter tests how Japanese listeners perceive obstruent clusters and how phonotactic predictability and recovery from coarticulatory cues interact during perception. There are three possible ways in which phonotactics and phonetic cues can interact: (i) listeners rely on phonotactics when phonetic cues are uninformative, (ii) listeners rely on phonetic cues when phonotactics is uninformative, and (iii) listeners stick to either phonotactics or phonetic cues. The stimuli are in the form $V_1C_1(V_T)C_2V_2$, where V_T is the target vowel and C_1 and C_2 are determined based on the stimulus group the token belongs to. The stimuli were divided into three groups: no reduction (No-Reduce) where vowel reduction is not expected, low predictability (Lo-Predict) where coarticulatory cues should be essential for recovery of a reduced vowel, and high predictability (Hi-Predict) where coarticulatory cues should play a lesser role in recovery of the target vowel. Below in Table 4.1 are the stimuli.

No-Reduce	eb_ko	ez_po	eg_to	ob_ke	oz_pe	og_te
Lo-Predict	ep_ko	eʃ_po	ek_to	op_ke	oʃ_pe	ok_te
Hi-Predict	eɸ_ko	es_po	eç_to	oɸ_ke	os_pe	oç_te

Table 4.1: Stimuli for Experiment 2.

There were 252 stimulus items in total. The stimulus forms shown in Table 4.1 were first recorded by a trained, non-Japanese-speaking, Hungarian-English bilingual phonetician in a sound-attenuated booth with stress on the initial vowel and with /i, u, a/ as target vowels (V_T). Attempts were made to record the stimuli with two native Japanese speakers, but both speakers had difficulties keeping high vowels unreduced in reducing contexts, and even when they were successful in

producing unreduced high vowels in reducing contexts, either the burst durations were too short to manipulate or the target vowel was stressed. Furthermore, /a/ was included as a target vowel because it is a low vowel that typically does not reduce in Japanese, and also because its inclusion allows investigation into whether Japanese listeners are sensitive to coarticulatory cues of all vowels or just high vowels.

For each recording, the target vowels were manipulated by inserting or removing whole periods to achieve a duration of $\sim$40 $\pm$ 5 ms. From each of the recordings, four additional tokens were created by removing right to left, half of V_T (splice-1), the remaining half of V_T (splice-2), half of the C_1 burst/frication noise (splice-3), then the remaining half of the C_1 burst/frication noise leaving only the closure for stops and $\sim$15 ms for fricatives (splice-4). An example of how the splicing was done is shown in Figure 4.1 below with the token [ekuto].

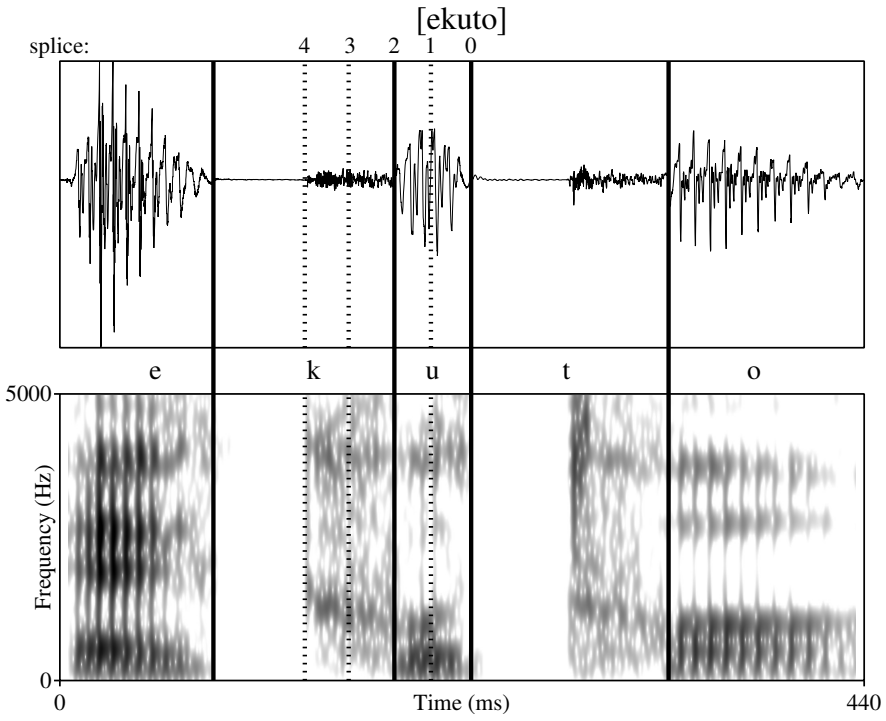

Figure 4.1: Example of token splicing: [ekuto].

The result of the splicing process is a gradual decrease of vowel coarticulatory information available in the burst/frication noise of C_1. The manipulation of stop bursts in particular will show whether it is phonotactic predictability or interpretation of phonetic information that drive illusory vowel epenthesis, since sensitivity to and interpretation of stop bursts as signaling the presence of a vowel is reported not just in Japanese (Furukawa, 2009; Whang, 2016) but in Korean (Kang, 2003) and English (Davidson and Shaw, 2012; Hsieh, 2013) as well. In addition, a naturally produced vowel-less token was also recorded for each stimulus form, where C_1 was released if a stop, to test how it differs in perception from the spliced vowel-less stimuli (splice-2), which may have traces of coarticulation from the target vowel on the surrounding consonants.

4.1.1 Participants

Twenty-nine monolingual Japanese listeners (16 women, 13 men) were recruited for the perception experiment in Tokyo, Japan. The participants of the experiment were similar in profile as the participants of the production experiment in Chapter 3, but no participant took part in both experiments. All participants were undergraduate students born and raised in the greater Tokyo area and were between the ages 18 and 24. Although all participants learned English as a second language as part of their compulsory education, none had resided outside of Japan for more than six months and have not been overseas within a year prior to the experiment. All participants were compensated for their time.

4.1.2 Procedure

The experiment follows the forced-choice vowel labeling task from Dupoux et al. (2011). The participants were told that they will be listening to foreign words over headphones and that they would have 5 seconds to choose a spelling choice that best matches the word they heard. The stimuli were presented through noise-isolating headphones, and answer choices that give the vowel-less and

various vowel-ful spellings of the stimulus that just played were presented on screen simultaneously (e.g., [epuko] → <epko>, <epako>, <epiko>, <epuko>). Participants selected their answer choices by using arrow keys on a keyboard (i.e., ↑ ↓ ← →). A typical answer-choice screen is shown below in Figure 4.2.

Figure 4.2: Answer choice screen for [epVko], where V = /a, i, u, ∅/.

While it is true that Japanese orthography is a syllabic system, most Japanese speakers are quite comfortable with the Latin alphabet, not only because of frequent exposure to loanwords but also because of the keyboards used for word processing. There are currently two main input methods—direct input (one key = one syllabic character) and conversion (QWERTY keyboard used to input CV combinations which are then converted to the corresponding syllabic character)—and the conversion method is commonly more preferred, and thus participants are expected to be comfortable with answer choices presented in the Latin alphabet. The experiment was designed to continue as soon as the participant makes an answer choice.

4.1.3 Analysis and predictions

As with the production study, all statistical analyses were performed by fitting linear mixed effects models using the *lme4* package (Bates et al., 2015) for R (R Core Team, 2016). The statistical

analyses assess vowel detection and vowel identification. Detection refers to how often participants report perceiving any vowel at all both in the presence and absence of vocalic segments. Identification refers to whether the vowel the participants perceive is in agreement with the acoustic vocalic information contained in the stimuli.

In the case of detection, accuracy is expected to be higher for the *No-Reduce* group (e.g., [ɛz_po]) and lower in the *Lo-Predict* and *Hi-Predict* groups (e.g., [ɛʃ_po] and [ɛs_po], respectively). Since the phonological process of high vowel reduction is nearly obligatory (Vance, 1987), it could bias Japanese speakers toward mistakenly "recovering" a high vowel that is expected to be present between two voiceless consonants even when it is acoustically absent. Since the current experiment uses nonce-words, there is no underlying or lexical form to access. Reduced and unreduced sequences involving two voiceless obstruents in Japanese would map to the same phonotactically legal surface form (e.g., [esupo] ≡ [es̥po] ≡ [espo] → ⟨esupo⟩). The reduced and unreduced sequences would all be regarded as legal, and the actual presence or absence of the vowel in the signal is readily ignored. It should be noted that if the stimuli being used were lexical items, reaction times would be predicted to differ depending on reduction status of the stimuli. Ogasawara and Warner (2009) found that Japanese listeners are quicker to identify reducible lexical items when presented with reduced overt forms. The interpretation of the results was that the most frequent overt form is considered to be the lexical form, and thus a direct mapping is possible from the most frequent overt form to the lexical form. For example, assuming that the lexical form of the loanword 'star' is |staː|, identification of the word would be the quickest when presented with the overt form [staː] (→ |staː|). Conversely, rare but nevertheless equivalent forms would undergo repairs, leading to slower lexical access. For example, when presented with the rare, unreduced overt form [sutaː], it must be first mapped to the surface form ⟨sutaː⟩, then to the underlying form /staː/, then to the lexical form |staː|.

While reduction is possible in the *No-Reduce* environments, it is extremely rare as was shown in Chapter 3 and also previous works (Maekawa and Kikuchi, 2005). Since only the vowel-ful

token is legal in the language in non-reducing contexts, the reduced or vowel-less counterpart is not in an equivalence relationship (e.g., [sude] ≢ *[sụde] ≢ *[sde] 'barehand') . Thus Japanese speakers are expected to be more sensitive to the presence versus absence of a medial vowel. Furthermore, regardless of the stimulus group, higher accuracy is expected in recognizing that there is no vowel as the burst/frication noise gets shorter, especially when there is no burst present.

In the case of identification, high accuracy is expected for the *Lo-Predict* group and lower accuracy for the *Hi-Predict* group. Japanese speakers have been shown to be sensitive to high vowel coarticulation in /ʃ/ (Beckman and Shoji, 1984), but this sensitivity is only useful when the vowel is unpredictable after a given C_1 (i.e., *Lo-Predict* group). Japanese listeners, therefore, should be sensitive to coarticulatory cues of at least [i, u]-coarticulation in the *Lo-Predict* group but biased towards a single high vowel that most frequently follows C_1 *Hi-Predict* group regardless of coarticulation. Since there are four answer choices <i, u, a, ∅>, identification rates are expected to be at least 50% in the *Lo-Predict* group and approximately 25% in the *Hi-Predict* group. Furthermore, since /a/ rarely reduces in Japanese, <a> responses should be relatively low even for [a]-coarticulated tokens, defaulting instead to the most phonotactically probably vowel. The *No-Reduce* group is expected to show some effects of coarticulation, as was the case in Dupoux et al. (2011), but like the *Hi-Predict* group, /a/ should show little effect.

These predictions contrast with the account given by Dupoux and colleagues. According to Dupoux and colleagues, there are two mechanisms at play during illusory vowel epenthesis. First, perceptual repair is a one-step process where phonotactically illegal sequences are perceived as their repaired counterparts rather than being perceived accurately first then repaired to their phonotactically legal counterparts. What this means is that listeners do not have access to the source language's underlying form, making heterorganic C_1C_2 sequences and their repaired C_1VC_2 sequences equivalent for Japanese listeners. If this is correct, the prediction in terms of detection is that the rate of vowel detection between C_1C_2 and C_1VC_2 sequences should be statistically the same since the two sequences are equivalent.

Second, although Dupoux and colleagues argue that perceptual repair is triggered by phonotactic violations, the repair strategy employed is not a purely phonotactic one but also a phonetic one, where L2 listeners make phonetically minimal repairs to the phonotactically illegal sequence. In Japanese, /u/ is epenthesized because it is the shortest vowel in the language, whereas the epenthesized vowel is /i/ in Brazilian Portuguese for the same reason (Dupoux et al., 2011). What this means is that the choice of the epenthesized segment is not because it is the most phonotactically probable vowel between any given consonant cluster but because it results in the smallest possible phonetic change. Also, based on the results of Chapter 3, high vowels can delete in Japanese, making many C_1C_2 and C_1i/uC_2 sequences equivalent. If the choice of the epenthetic segment is indeed based on the magnitude of phonetic change rather than phonotactic probability, no observable effect of phonotactic predictability is expected, since the phonotactic knowledge merely flags repair sites but is not involved in the repair itself. Vowel identification rates, therefore, are predicted to suffer across all contexts whenever the coarticulated vowel is not /u/. What this chapter aims to show, instead, is that minimizing phonetic changes are more relevant in low-predictability contexts and that phonotactic probability plays a more decisive role in high-predictability contexts. For example, given [ekto], the surface form that involves a minimal phonetic change is /ekito/, since /i/-epenthesis retains velar-fronting. Mapping the same stimulus to /ekuto/ results in a larger degree of phonetic change despite /u/ being the "default" epenthetic vowel because /u/-epenthesis also removes velar-fronting. On the other hand, given the high-predictability token [ecto], the surface form it maps to is /ecito/, not because of phonetic change but because /çu/ is phonotactically less probable.

4.2 Results

Shown in Figure 4.3 below are the overall results of the experiment. Figure 4.3.A shows results for all C_1 and Figure 4.3.B for stop C_1 only, which consequently also results in the exclusion

of all high-predictability tokens, since /ɸ, s, ç/ are all fricatives. The colors indicate the target vowels, and the solid and dashed lines indicate vowel detection and successful vowel identification rates, respectively. Vowel detection rates simply collapse all non-zero responses, whereas vowel identification rates only include cases where participant responses matched the coarticulated vowels in the stimuli (e.g., respond <epuko> for [epu̥ko]). The smaller the distance between two lines of the same color, the higher the proportion of successful vowel identification.

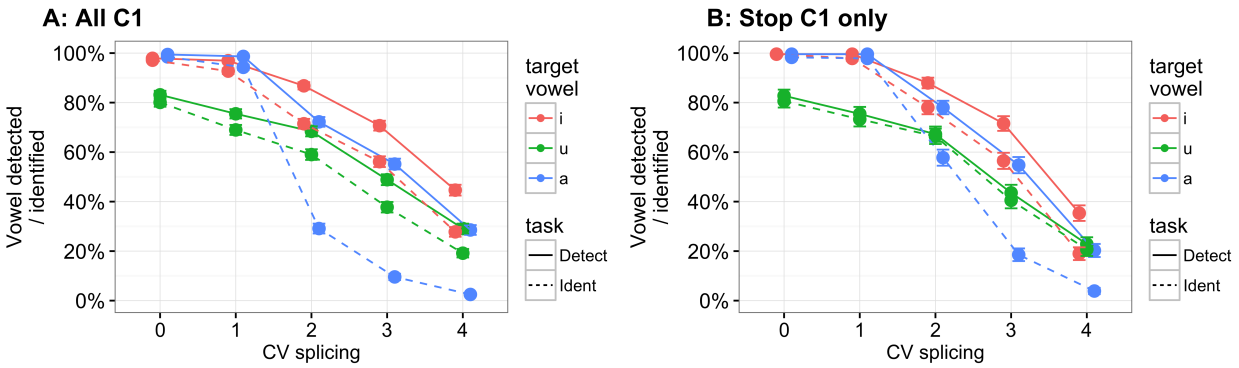

Figure 4.3: Vowel detection and identification rates with error bars by degree of splicing.
CV splicing: 0 = full-CV, 1 = full-C half-V, 2 = full-C zero-V, 3 = half-C zero-V, 4 = zero-CV.

Figures 4.3.A and 4.3.B are qualitatively similar, where detection and identification rates fall as more of the C_1V_T information is spliced, and the most noticeable effect of including fricatives in 4.3.A is that identification rates are driven lower. In both figures, there are three things that stand out. First, detection rates for /u/ never reach 100% even when there is a full vowel of 40 ms present in the stimuli, suggesting that there is confusion between the presence and absence of /u/. Second, vowel detection rates never quite reach 0%, remaining above 20% even in the absence of any C_1 burst noise (Figure 4.3.B, splice-4), suggesting an overall confusion between vowel-fulness and vowel-lessness. Third, /a/ identification rates (blue dashed line) fall the most dramatically and are the lowest in tokens where the medial target vowel is spliced out, suggesting that only high vowels are potentially available for recovery.

Because the results of splice-1 and splice-3 tokens show no surprising trends, the rest of this chapter will focus on the splice-0 (full-vowel), splice-2 (no vowel), and splice-4 (no vowel and no C_1 burst/frication) results. The splice-2 results will also be compared against naturally produced vowel-less tokens to test how the presence of coarticulatory cues affect the responses.

4.2.1 Tokens with full medial vowel

Shown below in Figure 4.4 are vowel identification rates for tokens with a full target vowel of 40 ms, broken down by context and by C_1. As shown previously in Figure 4.3, the identification rates are surprisingly low for /u/.

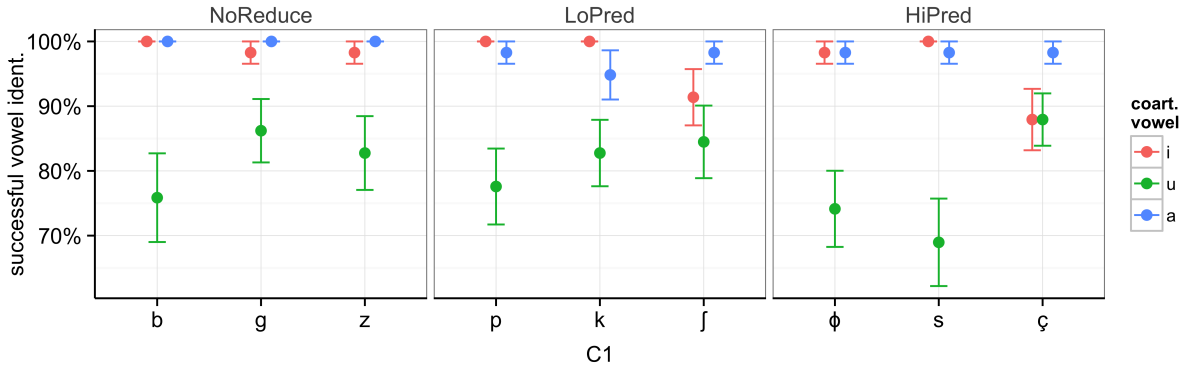

Figure 4.4: Successful vowel identification in VC_1VC_2V tokens with full medial vowel.

The most common wrong response by the participants for [u] identification was ∅ for all C_1 as shown in Figure 4.5, meaning that the participants either heard the vowel accurately or confused [u] with ∅, but rarely confused the vowel with another vowel. The confusion specifically between /C_1C_2/ and /C_1uC_2/ sequences suggests two things. First, the overt sequence [C_1uC_2] can be mapped to both /C_1uC_2/ and /C_1C_2/, although there is a bias towards the former. Second, fact that there is confusion between /C_1uC_2/ and /C_1C_2/ even when a vowel of 40 ms is fully present suggests that

the distinction between the two sequences is weak. Stated differently, C_1C_2 and C_1uC_2 sequences are treated as more or less equivalent by Japanese listeners.

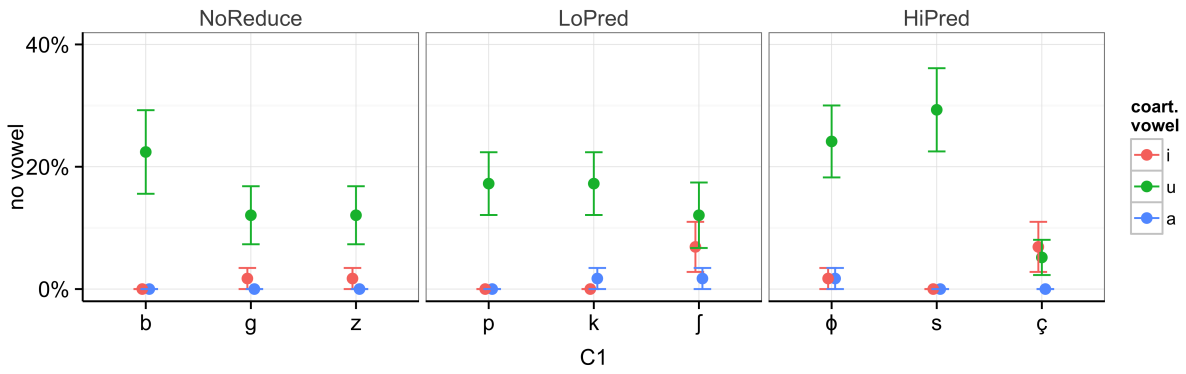

Figure 4.5: "No vowel" responses for VC_1VC_2V tokens with full medial vowel.

While this provides some support for the account presented by Dupoux and colleagues, the participants also exhibit some confusion between /i/ and ∅ after /ʃ, ç/. The reason for this additional confusion possibly stems from the phonotactics of Japanese. Presented below in Table 4.2 are the observed/expected ratios for all pertinent C_1V biphones, calculated from the Corpus of Spontaneous Japanese. What the O/E ratios show is that /u/ is highly overrepresented in Japanese after most consonants. The exceptions are /ʃ, ç/ after which /i/ is the most common vowel, and /g/ after which /a/ is the most common vowel.

	NoReduce			LoPred			HiPred		
	b_	g_	z_	p_	k_	ʃ_	ɸ_	s_	ç_
_a	1.63	**3.44**	0.93	1.78	1.80	0.27	0.11	0.92	0.43
_i	0.79	0.31	0.00	0.65	1.12	**6.28**	0.10	0.04	**6.28**
_u	**4.14**	0.78	**4.67**	**2.86**	**2.24**	0.33	**9.01**	**5.42**	0.002
_e	1.24	0.75	2.30	0.49	0.97	0.003	0.12	0.90	0.006
_o	0.75	1.33	0.99	0.43	1.33	0.42	0.07	1.16	0.01

Table 4.2: Observed/expected (O/E) ratio of C_1V from CSJ.

The results of the full-vowel tokens suggest that stimuli such as [epko, epu̥ko, epuko] are possibly all being treated as equivalent to /epuko/. Because they all map to the same phonotactically legal structure, there is bidirectional repair, although with a bias towards vowel recovery. The fact that there is confusion for /u/ across the board, even for /g/ despite /a/ being the most common vowel that follows, provides some support to the phonetically minimal repair hypothesis presented by Dupoux and colleagues. However, the fact that there is also confusion for /i/ after /ʃ, ç/ additionally suggests that phonotactic probability affects perception as well. It should be noted that the confusion between vowel-ful and vowel-less tokens might be the result of a task effect. Japanese listeners might not have even considered the vowel-less answer as a possibility were it not for the fact it was given as an answer choice. Given the choices <a, i, u, ∅>, [u] is not similar to either [i, a], and [o] was not an option, leaving ∅ as the most confusable choice.

4.2.2 Tokens with no medial vowel

This section compares the results of naturally vowel-less tokens and the splice-2 tokens where the medial, phonated vocalic material has been completely removed but C_1 burst/frication noise fully remains. Acoustically, the difference between these tokens is that the naturally vowel-less tokens contain no obvious coarticulatory information, unlike the spliced tokens.

4.2.2.1 Naturally vowel-less tokens

The prediction in terms of vowel detection was that the rate of <∅> responses should be highest for non-reducing contexts since high vowel reduction is rare in these contexts making Japanese listeners more sensitive to the presence versus absence of a medial vowel. Conversely, the rate of <∅> responses was expected to be low in contexts where high vowel reduction is expected, since high vowels can delete in these contexts, leading to a bias towards recovery. This bias should be

especially high in high-predictability contexts because C_1 in these contexts are allophones that precede specific high vowels (e.g., *ç* → *hi*).

Presented first below in Table 4.3 are the responses for naturally produced VCCV tokens. Bold numbers indicate the most frequent responses for a given C_1. A chi-square test was performed using the *chisq.test()* function in R, to test whether the observed response rates were significantly different from chance. /a/ responses were excluded under the assumption that /a/ is not a candidate for recovery and also because /a/ responses were at or near 0% in most contexts. The results showed that the observed responses were significantly different from chance at $p < 0.01$ with the exception of /p/ ($p = 0.4909$).

	NoReduce			LoPred			HiPred		
	ebko	egto	ezpo	epko	ekto	eʃpo	eɸko	espo	eçto
a	0.14	0.02	0.03	0.10	0.02	0.00	0.00	0.00	0.00
i	0.10	0.05	0.09	0.24	0.02	**0.55**	0.07	0.07	**0.76**
u	0.34	0.43	**0.50**	0.29	**0.59**	0.26	**0.60**	**0.60**	0.14
∅	**0.41**	**0.50**	0.38	**0.36**	0.38	0.19	0.33	0.33	0.10

Table 4.3: Responses for naturally produced VC_1C_2V tokens.

Overall, the results show that <∅> responses are 50% or lower across all contexts, revealing an overall bias towards illusory epenthesis. However, the rate of <∅> responses are highest for NoReduce environments, suggesting that there indeed is an effect of high vowel reduction. Additionally, <∅> responses are lowest for HiPred environments, suggesting that predictability has an effect on the rate of repair as well.

The responses for the naturally vowel-less tokens also suggest that there is an effect of phonotactics that drives the choice of vowel that is recovered by Japanese listeners. The vowel recovered after [ʃ, ç] is, again, /i/ rather than /u/, further strengthening the account that the choice of the vowel used for phonotactic repair is not just merely a default, minimal vowel but rather chosen based on phonotactics. This is also in line with a recent finding by Durvasula and Kahng (2015), who also found in Korean listeners that the choice of recovered vowel is better predicted by

the phonological alternations observed in the language rather than a phonetically minimal repair strategy.

4.2.2.2 Spliced vowel-less tokens (Splice-2)

Another prediction was that participants should be more sensitive to high vowel coarticulatory cues in contexts where high vowel reduction is expected, and especially so in low-predictability contexts. A mixed logit model was fit using the *glmer* function of the *lme4* package of R, with successful vowel identification rates as the dependent variable. The statistical analysis compares the rate of correct identification of spliced vowels from coarticulatory cues, so naturally produced VCCV tokens, which should contain no vowel coarticulatory cues, are not included in the analysis. The fixed effects structure of the model consisted of target vowel, context, and their interaction as predictors. The model with a fully-crossed, maximal random effects structure failed to converge, hence the final random effects structure included by-participant and by-stimulus random intercepts as well as by-participant random slopes for target vowel and C_1. The interaction term was shown to be a non-significant contributor to the fit of the model (p = 0.466024), and thus was excluded from the final model. The results are shown below in Table 4.4 with spliced [i] tokens in LoPred contexts as the baseline. The diacritic for devoicing (i.e., V̥) is used throughout the rest of this chapter to indicate vowels that have been spliced out.

	Estimate	Std. Error	z	Pr(>|z|)	
(Intercept)	2.3179	0.5048	4.591	4.40e-06	***
[u̥]	-0.9420	0.5879	-1.602	0.10906	
[ḁ]	-2.9936	0.5460	-5.483	4.19e-08	***
NoReduce	-0.9332	0.5121	-1.822	0.06844	.
HiPred	-1.6675	0.5146	-3.240	0.00119	**

Table 4.4: Mixed logit model results comparing successful vowel identification rates across difference predictability contexts.

The results show that the rates of high vowel identification were statistically comparable, but the rate of /a/ identification was significantly lower. Identification rates were also significantly lower in HiPred contexts than in LoPred contexts as predicted. Identification rates were also lower in NoReduce contexts, although not significantly so. These results are also shown graphically in Figure 4.6 below.

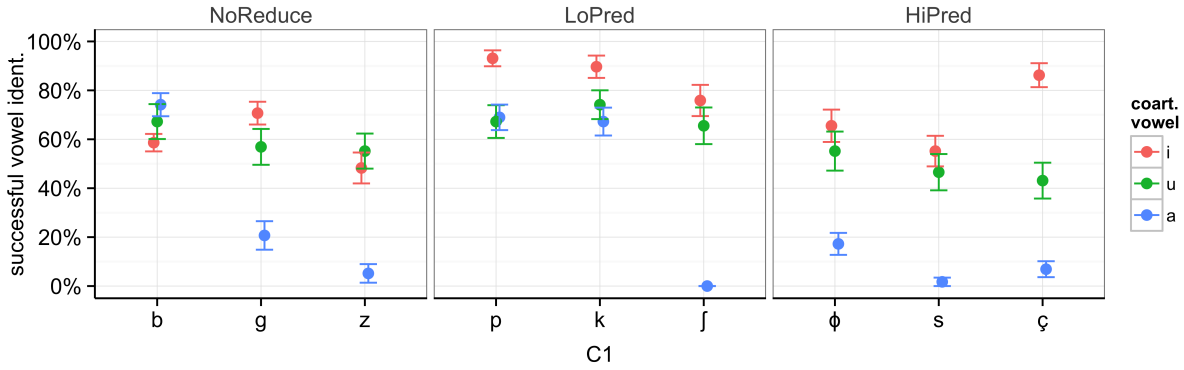

Figure 4.6: Successful identification rate of target vowel for spliced VCVCV tokens.

Since /a/ is a low vowel that is typically not targeted for reduction in Japanese, another mixed logit model with the same full fixed and random effects structures was fit to the data but with [a̰] tokens excluded in order to test how sensitive the participants were to high vowel cues specifically. Target vowel and its interaction with context were shown to be non-significant contributors to the fit of the model (p = 0.07804 and 0.36211, respectively), and thus only context was retained as a predictor in the final model. Shown below in Table 4.5 are the results, with LoPred context as the baseline.

	Estimate	Std. Error	z	Pr(>|z|)	
(Intercept)	1.7789	0.3597	4.945	7.6e-07	***
NoReduce	-1.1143	0.4382	-2.543	0.01100	*
HiPred	-1.2712	0.4376	-2.905	0.00367	**

Table 4.5: Mixed logit model results comparing successful vowel identification rates across difference predictability contexts, excluding [a].

The results show that there is a clear effect of predictability on how successful Japanese listeners are in identifying coarticulated vowels when only high vowels are considered. Both NoReduce and HiPred contexts have significantly lower identification rates, with HiPred being the lowest as predicted.

4.2.2.3 Comparison of naturally vowel-less and spliced vowel-less tokens

Naturally vowel-less tokens and spliced tokens by themselves tell only part of the story. Another prediction was that Japanese listeners should be able to recover high vowels from the coarticulatory information in spliced tokens, leading to differences between splice-2 and naturally vowel-less tokens. If it is the case that phonotactic violation alone is responsible for vowel epenthesis and that the choice of vowel is the phonetically minimal segment, namely /u/, then the presence of vowel coarticulatory information should do little to affect the choice of vowel.

Shown in Table 4.6 below are the results of a mixed logit model that compares detection rates for spliced tokens compared to a naturally vowel-less baseline. The results show that [i] coarticulation drives up the vowel responses significantly, and nearly so for [a] coarticulation. The lack of an effect for [u] coarticulation again shows increased confusion in these contexts. Additionally, vowel responses are significantly higher for HiPred tokens compared to a NoReduce baseline.

	Estimate	Std. Error	z	Pr(>|z|)	
(Intercept)	0.340915	0.361828	0.942	0.346089	
[i̥]	1.691540	0.480801	3.518	0.000435	***
[u̥]	0.634394	0.470474	1.348	0.177525	
[ḁ]	0.789575	0.455128	1.735	0.082768	.
LoPred	0.673063	0.457778	1.470	0.141485	
HiPred	1.032983	0.454347	2.274	0.022993	*
[i̥]:LoPred	0.221310	0.700610	0.316	0.752092	
[u̥]:LoPred	-0.233197	0.646576	-0.361	0.718350	
[ḁ]:LoPred	0.004041	0.648511	0.006	0.995028	
[i̥]:HiPred	-1.126838	0.670755	-1.680	0.092966	.
[u̥]:HiPred	-0.811062	0.650404	-1.247	0.212393	
[ḁ]:HiPred	-1.552257	0.639344	-2.428	0.015187	*

Table 4.6: Mixed logit model results comparing vowel detection between VCCV and spliced VCVCV tokens.

For identification rates, spliced tokens are compared separately to naturally vowel-less tokens to make the effects of coarticulation for each vowel clearer. Presented below in Table 4.7 below are the responses for spliced [u] tokens. The rate of <u> responses is higher compared to naturally vowel-less tokens (Table 4.3 above).

	NoReduce			LoPred			HiPred		
	ebu̥ko	egu̥to	ezu̥po	epu̥ko	eku̥to	eʃu̥po	eɸu̥ko	esu̥po	eçu̥to
a	0.02	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.02
i	0.02	0.00	0.12	0.00	0.00	0.09	0.03	0.05	**0.47**
u	**0.67**	**0.57**	**0.55**	**0.67**	**0.74**	**0.66**	**0.55**	0.47	0.43
∅	0.29	0.43	0.31	0.33	0.26	0.26	0.40	**0.48**	0.09

Table 4.7: Responses for VC_1(u)C_2V tokens with medial vowel spliced out.

A mixed logit model was fit to the data with the rate of <u> responses as the dependent variable. <u> was chosen since it is regarded as the default epenthetic segment. The predictors were target vowel (i.e., /∅, i, u, a/), C_1, and their interactions. C_1 was used as a predictor rather than context because the epenthetic vowel does not seem to be uniform across all contexts but rather depend on C_1. By-participant and by-stimulus random intercepts were included. By-participant random

slopes for target vowel and C_1 were also included. All predictors were significant contributors to the fit of the model. The results for <u> responses are shown below in Table 4.8, with ∅ tokens (i.e., naturally vowel-less) tokens as the baseline.

	Estimate	Std. Error	z	Pr(>|z|)	
(Intercept)	-0.84033	0.42766	-1.965	0.049418	*
[u̥]	1.89897	0.55110	3.446	0.000569	***
[g]	0.51123	0.51257	0.997	0.318574	
[z]	0.84678	0.53856	1.572	0.115880	
[p]	-0.39739	0.54914	-0.724	0.469270	
[k]	1.31600	0.53419	2.464	0.013757	*
[ʃ]	-0.45277	0.56776	-0.797	0.425184	
[ɸ]	1.58576	0.59324	2.673	0.007516	**
[ç]	-1.34398	0.68487	-1.962	0.049718	*
[s]	1.51309	0.57901	2.613	0.008969	**
[g]:[u̥]	-1.21609	0.73349	-1.658	0.097328	.
[z]:[u̥]	-1.62714	0.72768	-2.236	0.025347	*
[p]:[u̥]	0.50075	0.77601	0.645	0.518741	
[k]:[u̥]	-0.63586	0.78019	-0.815	0.415068	
[ʃ]:[u̥]	0.27845	0.76058	0.366	0.714290	
[ɸ]:[u̥]	-2.24928	0.78144	-2.878	0.003997	**
[ç]:[u̥]	-0.03906	0.78253	-0.050	0.960191	
[s]:[u̥]	-2.85498	0.77463	-3.686	0.000228	***

Table 4.8: Mixed logit model results comparing <u> responses between VCCV and spliced VC(u)CV tokens.

The model shows that there indeed is a significant raising effect of the coarticulated vowel. The mere presence of [k, ɸ, s] also drive up the rate of <u> responses significantly, while [ç] significantly lowers the rate of <u> responses, presumably because the expected vowel after [ç] is /i/. Interestingly, the expected vowel after [ʃ] is also /i/, but no lowering effect of <u> responses are observed. The interaction terms also show that the raising effect of [u] coarticulation is mitigated significantly after [z, ɸ, s] and nearly so after [g]. This is perhaps because the rates of <u> responses after these C_1 were already high in the naturally vowel-less tokens. These observations are also shown graphically in Figure 4.7 below.

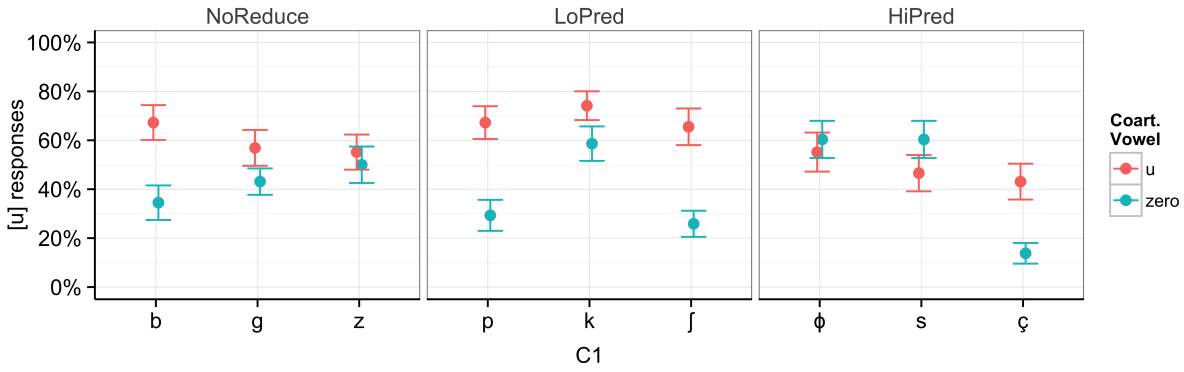

Figure 4.7: <u> responses for naturally vowel-less vs. spliced [u] tokens.

<i> responses are also driven up by [i] coarticulation. Shown below in Table 4.9 is a summary of the responses for spliced [i] tokens.

	NoReduce			LoPred			HiPred		
	ebi̯ko	egi̯to	ezi̯po	epi̯ko	eki̯to	eʃi̯po	eɸi̯ko	esi̯po	eçi̯to
a	0.09	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
i	**0.59**	**0.71**	**0.48**	**0.93**	**0.90**	**0.76**	**0.66**	**0.55**	**0.86**
u	0.10	0.10	0.41	0.03	0.03	0.10	0.19	0.24	0.03
∅	0.22	0.16	0.10	0.03	0.07	0.14	0.16	0.21	0.10

Table 4.9: Responses for $VC_1(i)C_2V$ tokens with medial vowel spliced out.

As was the case with [u] tokens, vowel coarticulation has a significant effect on which vowel participants report to hearing. A similar model as in Table 4.8 was fit, with the same predictors and random effects structure. The dependent variable was <i> responses with naturally vowel-less tokens as the baseline. The results are shown in Table 4.10 below.

	Estimate	Std. Error	z	Pr(>|z|)	
(Intercept)	-2.6791	1.0210	-2.624	0.00869	**
[i̥]	3.6475	1.4304	2.550	0.01077	*
[g]	-1.2952	1.5484	-0.836	0.40290	
[z]	-0.9058	1.4768	-0.613	0.53962	
[p]	0.3428	1.4628	0.234	0.81470	
[k]	-8.3719	4.7751	-1.753	0.07956	.
[ʃ]	2.9869	1.4244	2.097	0.03600	*
[ɸ]	-0.9942	1.5300	-0.650	0.51584	
[ç]	4.1493	1.4175	2.927	0.00342	**
[s]	-1.2364	1.5057	-0.821	0.41157	
[g]:[i̥]	1.9034	2.0904	0.911	0.36252	
[z]:[i̥]	-0.1457	2.0147	-0.072	0.94233	
[p]:[i̥]	2.6507	2.0799	1.274	0.20250	
[k]:[i̥]	16.1672	5.3645	3.014	0.00258	**
[ʃ]:[i̥]	-2.2153	1.9577	-1.132	0.25780	
[ɸ]:[i̥]	0.9448	2.0502	0.461	0.64491	
[ç]:[i̥]	-2.3320	1.9793	-1.178	0.23871	
[s]:[i̥]	0.6056	2.0328	0.298	0.76576	

Table 4.10: Mixed logit model results comparing <i> responses between VCCV and spliced VC(i)CV tokens.

In addition to the raising effect of [i] coarticulation on the rate of <i> responses, [ʃ, ç] also drives up the rate of <i> responses in naturally vowel-less tokens. This is also shown graphically in Figure 4.8 below. As discussed previously, the most common vowel after these consonants is /i/ in Japanese. The interaction term [k]:[i̥] also shows that the raising effect of [i] coarticulation is significantly higher after [k]. This is most likely due to the fact that /k/ is fronted before /i/, surfacing as [k^{j}], which was also shown to be the case in the previous chapter.

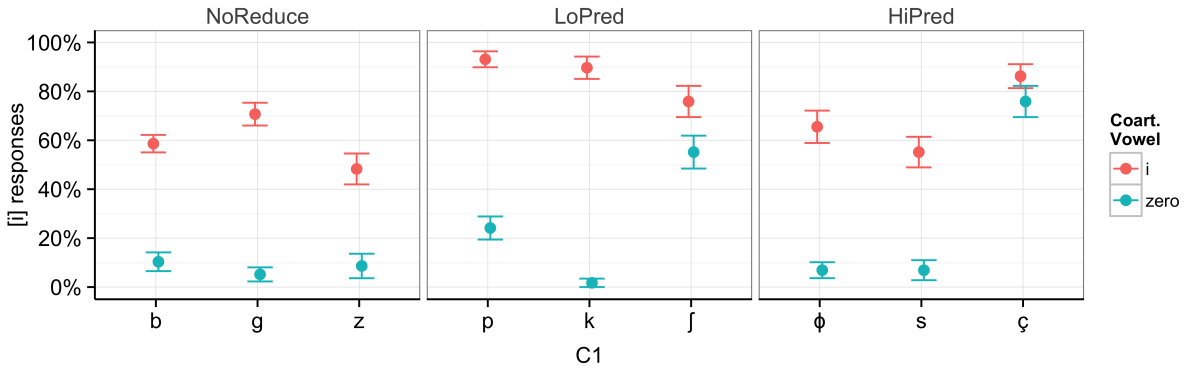

Figure 4.8: <i> responses for naturally vowel-less vs. spliced [i] tokens.

Thus far, the results suggest that the choice of epenthetic vowel for Japanese listeners is not simply a default /u/, but rather that the choice of vowel is sensitive to the acoustic cues in the signal. /u, i/ are both high vowels that are targeted for reduction in Japanese, so this is perhaps not surprising. Japanese listeners have had a lifetime of practice attending to subtle coarticulatory cues to recover reduced high vowels. Then what about a vowel like /a/, which rarely undergoes reduction? The responses to spliced [a] tokens are shown in Table 4.11 below.

	NoReduce			LoPred			HiPred		
	ebo̥ako	ego̥ato	ezo̥apo	epo̥ako	eko̥ato	eʃo̥apo	eɸo̥ako	eso̥apo	eço̥ato
a	**0.74**	0.21	0.05	**0.69**	**0.67**	0.00	0.17	0.02	0.07
i	0.00	0.03	0.09	0.00	0.00	**0.57**	0.10	0.03	**0.52**
u	0.12	*0.38*	**0.55**	0.09	0.19	0.26	0.21	0.47	0.28
∅	0.14	*0.38*	0.31	0.22	0.14	0.17	**0.52**	**0.48**	0.14

Table 4.11: Responses for $VC_1(a)C_2V$ tokens with medial vowel spliced out.

Although limited to post-stop environments (i.e., [b, g, p, k]), the results show that participants can recover the spliced [a] vowel at relatively high rates. Bilabial place also seems to have a facilitatory effect. Given that <a> responses were generally low in naturally vowel-less tokens, the raising effect even in the limited environments is surprising. The fact that /a/ identification is limited to stops may be due to the articulatory differences between stops and fricatives. Because

stops have a portion in which there is no airflow, coarticulation with the following vowel can be more complete by the time the stop burst/aspiration occurs. This is also true of bilabial place, where the lack of lingual gesture allows the following vowel to be coarticulated earlier. This is less true of fricatives where the transition into a fricative is more gradual, and coarticulation with the following vowel occurs towards the end of the segment rather than throughout. Since /a/ is a low vowel that a Japanese listener does not often have to recover, it may be that the beginning of the fricative already leads to the listener anticipating a high vowel and ignore to the low vowel cue towards the end.

A similar mixed logit model with the same predictors and random effects structure as in Tables 4.8 and 4.10 was fit. The interaction between target vowel and C_1 was not a significant contributor to the fit of the model and thus was excluded in the final model. The results are shown below in Table 4.12. The dependent variable was <a> responses with naturally vowel-less tokens as the baseline. Responses to [ʃ] tokens were removed from the model because <a> responses are at 0% for both the naturally vowel-less and spliced [a] tokens, resulting in no meaningful difference. When included in the model, [ʃ] tokens had an extremely low intercept of -27, but an absurdly high standard error of 22,246, both of which are most likely errors stemming from an absolute lack of difference between participants.

	Estimate	Std. Error	z	Pr(>|z|)	
(Intercept)	-3.0933	0.8330	-3.713	0.000205	***
[ḁ]	4.6969	0.7104	6.612	3.80e-11	***
[g]	-5.5653	1.2935	-4.303	1.69e-05	***
[z]	-9.7734	2.2923	-4.264	2.01e-05	***
[p]	-0.2094	0.9610	-0.218	0.827515	
[k]	-1.2300	0.9920	-1.240	0.215034	
[ɸ]	-3.8215	1.1929	-3.204	0.001357	**
[ç]	-8.7425	2.2688	-3.853	0.000116	***
[s]	-12.6671	2.6739	-4.737	2.17e-06	***

Table 4.12: Mixed logit model results for <a> responses.

The results confirm that indeed [a] coarticulation does have a significant raising effect on the rate of <a> responses. With [b] as the baseline C_1, the rate of <a> responses are significantly lower for [g, z, ɸ, ç, s]. The results are also shown graphically in Figure 4.9 below. The figure shows that the rate of <a> responses is indeed much higher in the spliced tokens than in the naturally vowel-less tokens.

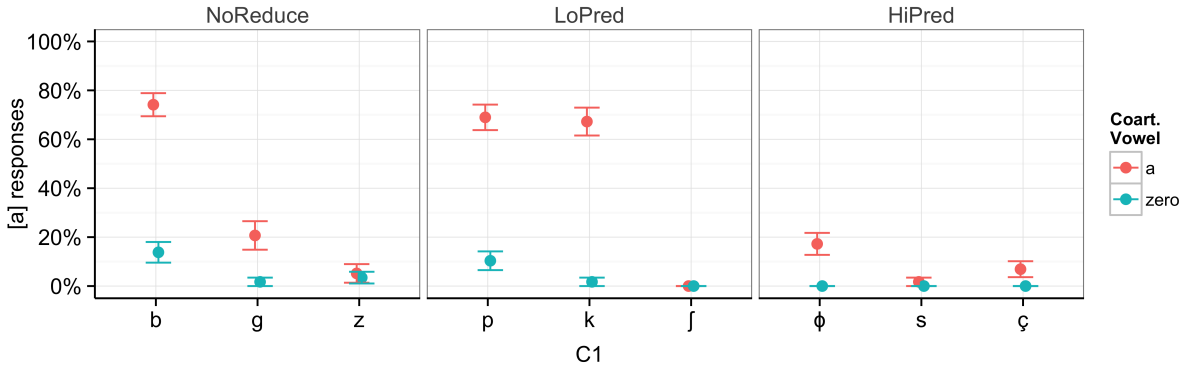

Figure 4.9: <a> responses for naturally vowel-less vs. spliced [a] tokens.

If the responses to naturally vowel-less tokens are taken as the default for phonotactically illegal sequences, the responses to spliced tokens show that there is an additional effect of sensitivity to phonetic cues. Phonotactically illegal consonant clusters are indeed repaired, but the epenthetic vowel is chosen due to a combination of phonotactic predictability and sensitivity to phonetic cues. Japanese listeners seem more sensitive to coarticulatory information of high vowels across all environments but also non-high vowels like /a/ in contexts where coarticulatory information is easier to detect.

4.2.3 Tokens with no vowel and no burst/short frication noise

The results discussed in §4.2.2 for spliced vowel-less but burst-ful tokens (splice-2) show that Japanese listeners are biased towards perceiving a vowel between heterorganic consonant clusters,

and that the choice of vowel is sensitive to the coarticulatory cues present in the C_1 burst/frication noise. Numerous studies have shown that the presence of a stop burst or frication noise in phonotactically illegal sequences are often interpreted as signaling the presence of a vowel (see Davidson and Shaw 2012, Hsieh 2013 for English; Furukawa 2009, Whang 2016 for Japanese; Kang 2003 for Korean). This section therefore discusses the results of splice-4 tokens, where the target vowel has been spliced out completely and C_1 also has been spliced out leaving just the closure for stop C_1 and <15 ms of frication noise for fricative C_1.

The responses to all splice-4 tokens are summarized in Table 4.13 below. A mixed logit model was fit to test whether the rates of <∅, i, u, a> responses were significantly affected by the identify of the vowel that was spliced out. Stop C_1 and fricative C_1 were analyzed separately. The results revealed that the responses were not significantly different regardless of the target vowel, with the exception of spliced [u̥] tokens where C_1 was /b/, which drove up <u> responses ($p = 0.002333$). Because the effect was limited to a single consonant, this section collapses the responses across all target vowels and focuses more on vowel detection.

	NoReduce			LoPred			HiPred		
	eb˺ko	eg˺to	ez˺po	ep˺ko	ek˺to	eʃ˺po	eɸ˺ko	es˺po	eç˺to
a	0.05	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.01
i	0.08	0.13	0.16	0.07	0.02	0.36	0.09	0.10	0.45
u	0.32	0.17	0.34	0.07	0.09	0.11	0.11	0.14	0.10
∅	0.55	0.68	0.47	0.85	0.87	0.52	0.78	0.76	0.45

Table 4.13: Responses for VC_1˺C_2V tokens with medial vowel and C_1 burst/frication noise spliced out.

The results show first and foremost that the rate of <∅> responses never reaches 100%. This is perhaps expected for fricative C_1, since there was ∼15 ms of frication remaining in the tokens. Factors contributing to the results for stop C_1, on the other hand, are less obvious. A mixed logit model was fit separately for the stops and fricatives since the the fricative tokens had a short frication noise remaining whereas the stop tokens had no burst at all. The full model for both data

subsets had the following structures. The fixed effects included context, V_1, and their interaction. All stimuli used in the experiment had the form $V_1C_1(V)C_2V_2$, where the order of V_1-V_2 was always either [e-o] or [o-e]. V_1 was included as a predictor to test whether the ordering of the initial and final vowels had a significant effect on vowel detection, which would suggest that there might be V-to-V coarticulatory cues that the participants are picking up on. The random effects included by-participant and by-stimulus random intercepts as well as by-participant random slopes for context, V_1, and their interaction.

Shown first below in Table 4.14 is the result of the final model for the stop-only subset. Since the HiPred context had no stops, the subset only includes NoReduce and LoPred contexts with the latter as the baseline. The interaction term was shown to be a non-significant contributor to the fit of the model (p = 0.5463) and thus was removed.

	Estimate	Std. Error	z	Pr(>|z|)	
(Intercept)	-1.8118	0.3162	-5.730	1.00e-08	***
V_1 = [o]	-0.4413	0.3369	-1.310	0.19	
NoReduce	1.5019	0.3493	4.299	1.71e-05	***

Table 4.14: Mixed logit model result for vowel detection in spliced vowel-less and burst-less stop tokens.

The results show that V_1 did not have a significant effect, but the rate of vowel detection was significantly higher for NoReduce tokens than LoPred tokens. A possible explanation for this effect is that the C_1 in NoReduce tokens had consistent phonation during closure, as shown in Figure 4.10 below.

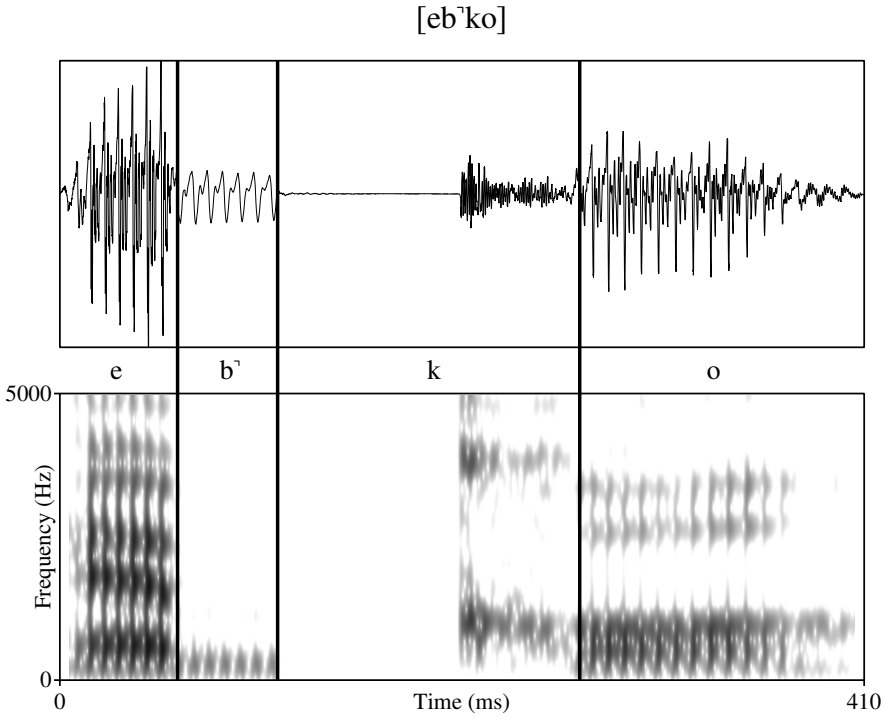

Figure 4.10: Spliced vowel-less, burst-less token created from [ebako].

The mixed logit model for the fricative-only subset also shows that the vowel detection rate for the NoReduce fricative [z] is significantly higher than for HiPred fricatives although not higher than the LoPred fricative [ʃ] ($p = 0.658$). For the fricatives, only context was a significant contributor to the fit of the model, and thus V_1 ($p = 0.81919$) and V_1:Context ($p = 0.82666$) were excluded from the fixed effects structure of the final model. The results are shown below in Table 4.15.

	Estimate	Std. Error	z	Pr(>|z|)	
(Intercept)	-0.9063	0.3023	-2.998	0.00272	**
LoPred	0.7666	0.5488	1.397	0.16243	
NoReduce	1.0569	0.5065	2.087	0.03691	*

Table 4.15: Mixed logit model result for vowel detection in splice-4 fricative tokens.

Although the fact that vowel detection rates never fall to 0% can be easily explained by the presence of prevoicing for NoReduce tokens and the 15 ms frication noise for the fricatives, the

10+% of vowel detection for the LoPred stops [p̚, k̚] is still somewhat puzzling. Without a vowel and without a burst between C_1 and C_2, a token such as [ep̚ko] contains a doubly long stop closure, much like a geminate medial consonant as in [ekko]. Geminate consonants are phonotactically legal in Japanese and require no repair. Nevertheless, participants report perceiving a vowel some of the time. It is possible that some participants are picking up on the mismatch between the transitional cues out of V_1 and into V_2. This seems unlikely, however, in that transitional cues into a vowel often outweighs transitional cues out of a vowel for Japanese listeners (Fujimura et al., 1978) and that Japanese listeners rely more on centroid spectral cues than on formant transitions (Hirai et al., 2005). Perhaps a more likely explanation is one of task effect. Although the stimuli sounded as though they contain a geminate obstruent, there was no geminate option given as a possible answer. This might have kept the participants from fully eliminating the vowel-ful answer choices, and having been exposed to numerous vowel-ful tokens (both acoustically and perceptually) during the task, the participants might have assumed that a vowel should be present at least some of the time.

4.2.4 Main findings

There were five main findings in the perceptual experiment. First, Japanese listeners seem to sometimes confuse the high vowel that is phonotactically the most likely after a given C_1 with ∅ even when the high vowel is 40 ms long and fully phonated. This sort of confusion was not observed with the low vowel /a/, which typically does not reduce in Japanese. Second, results from naturally vowel-less tokens revealed that the vowel most often perceptually epenthesized between illicit clusters is /u/, largely due to the fact that it is phonotactically the most probable vowel after most obstruents in Japanese. This is further supported by the finding that after /ʃ, ç/, which is most often followed by /i/ rather than /u/, the choice of epenthetic vowel is in fact /i/. Third, participants successfully identified spliced high vowels in splice-2 tokens (full C_1 with target vowel completely spliced out) at rates significantly higher than the baseline rates observed in naturally vowel-less

tokens. Identification rates of spliced /a/ were significantly lower and limited to after stops. Fourth, related to the third finding, identification rates of high vowels were lowest in HiPred contexts, suggesting that listeners are less sensitive to low-level coarticulatory cues in contexts where the phonotactics typically is sufficient for identifying the target vowel. Lastly, <∅> responses never quite reach 100% even for splice-4 tokens where both C_1 and target vowel were fully spliced out.

4.3 Discussion and conclusion

The production experiment in the previous chapter showed that reduced high vowels are more likely devoice and retain vowel coarticulatory cues in low predictability contexts where both high vowels are possible (e.g., /ʃi, ʃu/ → [ʃi̥, ʃu̥]) and that reduced high vowels are more likely to delete in high predictability contexts where only one high vowel is possible (e.g., /*si, su/ → [s]). Given the varying degrees of coarticulatory cues depending on phonotactic predictability, the aim of the current chapter was to test whether Japanese listeners are more sensitive to coarticulatory cues in low predictability contexts during perception. Broadly speaking, that overt consonant clusters are mapped to a phonotactically legal CVC sequence, neutralizing the contrast between CC and CVC sequences as Dupoux and colleagues have shown. However, the specific vowel recovered is modulated by CV co-occurrence probabilities in Japanese, as well as by detailed phonetic information.

To discuss these points in more detail, first, the perception of full-vowel tokens showed that there is confusion between /u/ and ∅, even when there is a 40 ms-long, phonated [u]. It is possible that this confusion arises because /u/ is indeed the default epenthetic vowel in Japanese, making it equivalent to ∅. However, a survey of biphone co-occurrence probabilities in the Corpus of Spontaneous Japanese revealed that /u/ also happens to be the most common vowel after most consonants, making it difficult to attribute the seemingly default status of /u/ as stemming simply from its shortness (Dupoux et al., 1999, 2011). Furthermore, similar confusion with ∅ is observed for

/i/ after /ʃ, ç/, suggesting that the choice of epenthetic vowel must be conditioned by the phonotactics of the language at least in part.

Second, the perception of vowel-less tokens further suggests that Japanese listeners confuse vowel-ful and vowel-less tokens with a tendency towards vowel-fulness. The results for splice-4 (vowel-less and burst-less) tokens in particular showed that Japanese listeners interpret even the most minute acoustic cues such as prevoicing of stops as signaling the presence of a vowel (§4.2.3). However, participants do not seem to simply perceive a default vowel. A comparison between naturally vowel-less and spliced vowel-less tokens showed that spliced tokens drive up the rate of target vowel responses significantly. This suggests that while heterorganic C_1C_2 sequences are perceived as being equivalent to C_1VC_2 as Dupoux and colleagues claim, the particular vowel is again not simply the "default" but is also determined by acoustic information in the signal. The participants, therefore, are recovering the vowel that is the most likely based on the phonetic cues contained in the burst/frication noise of C_1.

Third, the rate of high vowel identification was above chance at 40% across all contexts in spliced vowel-less tokens. Specifically, recovery rates were the highest in LoPred contexts as predicted, and the recovery rates were significantly lower for HiPred contexts, also as predicted. Recovery rates in NoReduce contexts fell somewhere between the two reducing contexts. The high rates of recovery suggest that Japanese listeners are hypersensitive to vowel coarticulatory cues, and the lower rate of recovery in HiPred contexts additionally suggests that sensitivity to coarticulatory cues are conditioned by phonotactic predictability.

Lastly, the sensitivity coarticulatory cues in Japanese listeners is limited to high vowels. The participants were worst at identifying /a/. Non-high vowels are typically not reduced in Japanese, and thus Japanese listeners have relatively little experience recovering them.

There are two major points to be made about these results. First, perhaps the terms perceptual epenthesis and "illusory" vowel epenthesis should be not used interchangeably. The results reported in this chapter show that phonotactic repair of heterorganic clusters in Japanese is not illusory. Due

in part to the highly productive process of high vowel reduction, phonetic cues perceived from the signal are used by Japanese listeners to identify a specific segment (i.e., perceptual epenthesis). Conversely, it may very well be the case that the phonotactic repair by Brazilian Portuguese listeners truly is an "illusory" one triggered primarily by phonotactic violations.

Second, the fact that vowel identification rates are lower in contexts where phonotactic predictability is high further bolsters the idea that predictability can enhance or obscure phonetic cues that facilitate recoverability both during production and perception. Stated differently, while Japanese listeners are sensitive to coarticulatory cues of high vowels in general, the same high vowel cues are nonetheless less utilized in contexts where their language experience tells them that low-level cues can be ignored.

A more general point that the results make is that language-specific coarticulation and sensitivity to these cues are also context-specific. It has been observed for some time that listeners attend to the types of cues that are the most informative in their language. For example, Korean listeners are more sensitive to V-to-C formant transitions than English speakers (Hume et al., 1999) because coda obstruents are obligatorily unreleased in Korean but optionally so in English (Kang, 2003). Another example is that English listeners tend to interpret the presence of nasality during a vowel as a coarticulatory cue for an upcoming nasal consonant, whereas Bengali speakers interpret the same cue as signaling a nasal vowel (Lahiri and Marslen-Wilson, 1991). What the experimental results of this chapter additionally show is that listeners take into account the context when considering how informative the cues they attend to are. Japanese listeners are sensitive to CV coarticulation, but specifically for high vowels and not /a/. Attending to /a/ coarticulatory cues is a waste of cognitive resources because the low vowel never reduces in Japanese and thus the listener can simply wait for a more robust, higher amplitude cue—the vowel itself. Similarly, attending to low level coarticulatory cues that typically help recover the same vowel as one's phonotactic knowledge is also a waste of cognitive resources, and thus phonetic cues are less utilized in high

predictability contexts. In other words, the implementation of and the sensitivity to certain phonetic cues are context-specific in addition to being language-specific.

CHAPTER 5

Learning Japanese high vowel reduction

5.0 Introduction

The results of the production experiment in Chapter 3 showed that the process of high vowel reduction can result in complete deletion of the target vowel in high-predictability contexts. What is puzzling about this is that Japanese is well-known for its strong phonotactic preference for a CVCV structure (Shibatani, 1990; Kubozono, 2015) which biases Japanese listeners towards vowel recovery during perception. This was shown to be the case in the perception experiment in Chapter 4 as well. What we have then is a dislike of consonant clusters during perception despite their frequent occurrence during production as a result of reduction.

In this chapter, I present a preliminary computational model that explores how the two seemingly contradictory aspects of Japanese phonology (i.e., process of vowel deletion and phonotactic restriction against consonant clusters) might be learned. As will be discussed in more detail in the subsequent sections, Japanese infants generally seem to ignore contrasts between CCV and CVCV

sequences during perception quite early, by around one year of age. Japanese children take longer, however, to learn high vowel reduction, favoring CVCV sequences until around five years of age. To capture the observed acquisition process, the model combines a phonotactic learning mechanism that relies solely on overt forms without access to higher levels of representation and a lexicon-based mechanism that learns alternations by matching one or more overt forms that correspond to the same lexical item.

5.0.1 The acquisition of Japanese high vowel reduction

Before presenting the model, this section first discusses what is known about the acquisition of high vowel reduction in Japanese children, beginning with the first issue a learner faces in the acquisition of any language: the input. Studies on infant-directed speech (IDS) often report a number of significant differences between infant-directed speech and adult-directed speech—e.g., expanded vowel space and F0 range. So how does Japanese IDS compare when it comes to high vowel reduction? Japanese has a strong preference for a CVCV structure. This was shown to be true of nonce words in numerous psycholinguistic studies (Dupoux et al., 1999, *et seq.* and Chapter 4 of this dissertation), where Japanese listeners confuse CCV and CVCV sequences, preferring to map them both to CVCV. The preference for CVCV structure has also been shown in lexical words in the analyses of Yamato, Sino-Japanese, and loanword lexical strata (Ito, 1986, *et seq.* and Chapter 2). Japanese IDS can use canonical, unreduced forms to facilitate structure learning on the one hand, but this would obscure the high vowel reduction process that is an integral part of adult speech. On the other hand, providing adult-like speech with vowel reduction would obscure the CVCV preference.

Given this seemingly irreconcilable conflict, Japanese caretakers seem to prefer to provide adult-like speech to infants. Fais et al. (2010) investigated how the speech of Japanese mothers of one-year-old infants differ when speaking to their children (IDS) and to adults (ADS). Reduction

rates of high vowels were calculated by identifying all instances of contexts in which high vowel reduction is expected and checking both auditorily and visually (waveform and spectrogram) for vowel presence in the acoustic signal. The results revealed that while there are some differences in prosodic cues, the rates of high vowel reduction in Japanese IDS and ADS are virtually identical. For both IDS and ADS, the reduction rates were around 85% for lexical words and around 20% for nonce words. Furthermore, Fais et al. also reported that nonce words tended to get reduced more with more use.

A more recent paper by Martin et al. (2014) also investigated Japanese high vowel reduction in IDS. High vowel reduction rates were calculated similarly to Fais et al. (2010), where all instances of high vowels between two voiceless obstruents were identified and then coded for reduction status by a trained phonetician. Non-high vowels between two voiceless obstruents were also identified and coded for their reduction status. The results reported by Martin et al. (2014) are somewhat different from that of Fais et al. (2010). In their study, the rate of reduction for high vowels in ADS was 90% overall, whereas in IDS it was 77%. A statistical analysis revealed that the difference was significant ($p < 0.0001$). Also, although reduction is typically described as being limited to high vowels, Japanese speakers have been shown to reduce non-high vowels as well, albeit much less frequently. Martin et al. (2014) report in their study that in ADS, the rate of non-high vowel reduction was extremely low at 2%, while in IDS the rate was significantly higher at 11% ($p < 0.001$). In other words, Japanese mothers tend to reduce high vowels less but reduce non-high vowels more when speaking to their children.

The fact that the elicitation methods used for IDS in the two studies were different could have contributed to the difference between the results from Fais et al. (2010) and Martin et al. (2014). In the Fais et al. study, the IDS sample consisted of spontaneous speech intermixed with read speech, which were later differentiated during analysis. In the Martin et al. study, there was only spontaneous speech and no read speech. Another possible explanation is that the ages of the children in the studies were different. As the authors of both papers note, a number of studies on IDS

have shown that the characteristics of IDS change according to the age of the infant being addressed. For example, results from Bernstein Ratner (1984, as cited in Fais et al. 2010) show that vowels were more exaggerated in speech directed towards infants who have begun producing 2-4 word utterances than in speech directed towards pre-speech or holophrastic (younger) infants. Likewise, Malsheen (1980) report that the difference in VOT for voiced versus voiceless stops was greater in speech to infants between 1;3 (1 year; 3 months) and 1;4 than in speech to either younger (0;6 - 0;8) or older (2;0 - 5;0) infants. Results from such studies suggest that as infants begin speaking, IDS changes accordingly to emphasize certain linguistic structures such as segments and words. The children in Fais et al. (2010) were 1;0, typically an age at which children produce very limited number of words if any. The children in Martin et al. (2014), on the other hand, were between 1;5 and 2;1, an age at which infants typically go through a rapid development in production, and also the age range during which exaggerated, less adult-like production of vowels and consonants is reported (Malsheen, 1980; Bernstein Ratner, 1984).

If it is the case that Japanese IDS changes with regards to rates of high vowel reduction depending on the age of the infant, one must ask whether infants learn high vowel reduction before or after the change. Behavioral studies by Kajikawa et al. (2006) and Mugitani et al. (2007) show that infants are sensitive to the difference between reduced and unreduced sequences at the age of 0;6, but this sensitivity is noticeably diminished by the age of 1;0, and even more so by the age of 1;6. In other words, Japanese infants have already learned the process of high vowel reduction and have learned to ignore the difference between C_1C_2 and C_1VC_2 sequences by the age of 1;0. This is an age when IDS is reportedly virtually identical to ADS in terms of high vowel reduction rates. If it is the case that infant-directed speech changes according to development, it seems safe to assume that the rates of high vowel reduction in Japanese IDS is similar to that of ADS, as reported in Fais et al. (2010).

There are very few studies that have looked at the *production* of high vowels in Japanese infants. However, a study by Imaizumi et al. (1999) looked at the developmental differences between

children learning different dialects of Japanese, and found that reduction rates are generally low for Japanese children before reaching adult-like levels around the age of five.

To summarize, it seems Japanese infants learn the process of high vowel reduction quite early in development at least in perception, and that mastery of producing reduced high vowels is acquired much later. Importantly for the current study, it seems safe to assume that the input to the child learner contains a high number of consonant clusters or consonant cluster-like sequences as a result of reduced forms produced by adult speakers.

5.0.2 Models of phonological learning

Most computational models of phonological acquisition fall broadly into two categories, which the current model attempts to combine: phonotactic learners and alternation learners. Models of phonotactic learning typically define the learning problem as finding the appropriate ranking or weighting for a universal set of constraints (e.g., Tesar and Smolensky, 2000; Prince and Tesar, 2004; Coetzee and Pater, 2008) or as finding rankings as well as the constraints themselves ('constraint induction'; e.g., Hayes, 1999; Hayes and Wilson, 2008a). Although phonotactic knowledge plays a role in both perception (Dupoux et al., 1999) and production (Davidson and Stone, 2003), thus far computational models of phonotactic learning have generally focused on perception. For example, the Maximum Entropy Model (MaxEnt; Hayes and Wilson, 2008a) and the Minimal Generalization Learner (MGL; Albright and Hayes, 2003) learns the phonotactic grammar of a given language, and the resulting grammar is used to predict well-formedness judgments of nonce words.

Other phonotactic learning models focus on the contribution of phonotactics to infants' discovery of word boundaries in continuous speech (Adriaans and Kager, 2010; Blanchard et al., 2010; Daland and Pierrehumbert, 2011). Such models take into account the fact that infants are sensitive to various aspects of their native language, such as phonetic categories (Werker and Tees, 1984; Werker and Lalonde, 1988; Maye et al., 2002) and phonotactics (Jusczyk et al., 1994; Mattys

and Jusczyk, 2001) before they are 1;0 and as early as 0;6. Infants around this age have also been shown to be able to extract words from a continuous stream of speech (Jusczyk and Aslin, 1995; Saffran et al., 1996). In other words, infants already have quite sophisticated knowledge of their native phonology presumably before they have acquired a sufficiently detailed lexicon.

Most work on alternation learning has been experimental (Pater and Tessier, 2003; White, 2014) or theoretical (Tesar and Prince, 2007), and unlike phonotactic models, alternation models are typically rule-based learners that have access to a morphologically detailed lexicon. A common assumption in the alternation learning literature is that phonotactic learning is more or less complete by the time alternation learning begins (Tesar and Prince, 2007) and that the goal of alternations is to form phonotactically preferred surface structures (Pater and Tessier, 2003). Recent work by Peperkamp et al. (2006) presents a computational model for allophonic alternation learning, which is extended to account for a broader scope of alternations by Calamaro and Jarosz (2015). These models focus on the viability of correctly identifying alternation rules in a given language statistically, but it is not immediately obvious as to how these rules, once learned, can be implemented into a computational model that must evaluate speech.

5.0.3 The proposed model

The model being proposed in the current chapter explores how a phonotactic grammar learned from overt forms and knowledge of alternations in the lexicon can both be used to account for high vowel deletion and CV preference in Japanese. Phonotactic learning is argued to begin early, based on acoustic input (i.e., overt forms) in the absence of a lexicon (Hayes, 2004; Prince and Tesar, 1999), although this reliance on overt forms from phonotactic learning lessens once a sufficiently morphologically detailed lexicon is acquired that enables phonological alternation and underlying (or phonemic) form learning (Tesar and Prince, 2007). Because phonotactic learning happens at an early age, it likely interacts with the acquisition of high vowel reduction and a CVCV preference

that interprets reduced forms as containing a vowel. In addition to a phonotactic component, the model includes an alternation learning component. This is because high vowel reduction results in an alternation between reduced and unreduced high vowels. Numerous studies have consistently shown that high vowel reduction rates are generally above 90% between two voiceless obstruents, while it is well below 10% elsewhere (Tsuchida, 2001; Fujimoto, 2015). In other words, reduced and non-reduced high vowels are in near complementary distribution of each other. By combining the two learning mechanisms, the aim is to find that the seemingly contradictory preferences for CVCV structure and reduction of high vowels can be learned from the same input data. The additional benefit of implementing the two approaches is that it provides a chance to test the claim that phonotactic knowledge might aid alternation learning even when the two processes should lead to contradictory outputs.

The phonotactic and alternation learning mechanisms are applied to a large corpus of real Japanese speech data taken from the Corpus of Spontaneous Japanese (Maekawa, 2003) and the combined model's performance is evaluated against the production data from Chapter 2 and the perception data from Chapter 3. Computational models of phonological acquisition have typically focused on English and Dutch, primarily because these are the two languages that have speech corpora readily available for analysis. Modeling work of other languages has had to rely on artificial data tailored specifically for the phenomenon being modeled, or on corpora of insufficient quantity or quality for rigorous work. The model presented here is trained on a subset of the Corpus of Spontaneous Japanese, which consists of 7.5 million words, or about 660 hours of speech in total. The subset used is the *CSJ-Core*, which contains approximately 500,000 words (roughly 45 hours of speech) that have been meticulously segmented and labeled phonemically and sub-phonemically.

5.0.4 The corpus

The annotations provided by the CSJ-Core are geared towards varying levels of linguistic analysis, ranging from sentence-level annotations to annotations regarding phonetic variations in the recorded data. The most relevant annotations for the model are what the CSJ-Core refers to as the word level, the phoneme level, and the phone level transcriptions provided in the corpus. For each word, which the CSJ-Core represents in Unicode characters of Japanese orthography, there are "phoneme" and "phone" level transcriptions. The phoneme level transcriptions are largely based on the kana syllabary and consequently employ a mix of /underlying/ and ⟨surface⟩ level transcriptions. The phone level transcriptions correspond most closely to the [overt] level, but there are syllabary-based aspects to the transcriptions as well. Take the Sino-Japanese compound 'full enjoyment', for example. It is represented as <満喫>, <maNkicu>, and <maNkjIcU> at the word, phoneme, and phone levels, respectively. In the transcriptions, <N> is the moraic nasal coda, <kj> is a "phonetically palatalized voiceless velar stop", <c> is an alveolar affricate, and uppercase vowels are devoiced vowels. This means that the phoneme level transcription would correspond to *maNkitsu* and the phone level to *maN̥k̟i̥tsu̥*. Compare these with /maNkit/, ⟨maŋk̟itsu⟩, and [maŋk̟i̥tsu̥]. The "phoneme" level transcription contains elements of both underlying and surface level transcriptions – underlying-like with an unassimilated nasal coda and a velar stop that is not fronted but surface-like with an epenthetic final vowel after a CVC root and the resulting allophony of /t/. The phone level transcription is most like the overt form, but it too contains an unassimilated nasal coda. The two transcription levels, therefore, required some modifications before they could be used as input for the model.

In general, the "phone" level was modified to simulate the [overt] level as closely as possible, showing all assimilations and reductions. Hence, phone level transcriptions will be notated with square brackets. The "phoneme" level was left largely untouched, employing a mix of surface and underlying forms that is closely tied to the syllabary of Japanese for two reasons. First, lexical strata

is not marked in the corpus, making it difficult to identify and convert epenthetic and allophonic segments. Second, underlying forms are not accessible during acquisition, and knowledge of the kana syllabary would affect how lexical items are represented underlyingly. For example, having received the word 'to like' only as [ski], a child would assume that the underlying form is the same as the overt form (Tesar and Smolensky, 1998). However, when the child learns to read and sees that the [s] is written with the same syllabary as [su] in a word like [sugoi], that knowledge would undoubtedly affect how the word [ski] is represented underlyingly. Although the phoneme level is a hybrid of two levels, these transcriptions will be treated as underlying forms and notated with forward slashes hereafter accordingly.

There are also some specific modifications that were made. First, although the CSJ-Core makes a distinction between "phonetically palatalized" <Cj> and "phonologically palatalized" <Cy> consonants in its transcription convention (Maekawa, 2003; Maekawa and Kikuchi, 2005), this distinction does not seem to be motivated by the acoustics. The "phonetically palatalized" segments occur exclusively before /i/ in the corpus while the "phonologically palatalized" transcriptions are used before all other vowels, suggesting that <Cj> is really a notation for coarticulation. For most consonants, the contrast between <Cj> and <Cy> would be neutralized (e.g., <cji, hji> → [ʧi, çi] vs. <cyu, hyu> → [ʧu, çu]), although there perhaps is a contrast for velar stops (e.g., <kji, kyu> → [k̟i, kʲu]). It is unclear how the phonetically and phonologically palatalized sounds would differ acoustically at this point. Based on the results from chapter 4, where Japanese listeners showed sensitivity to the coarticulatory differences between between [ʃi, ʃu] but not other fricatives, differences between <sj> and <sy> were maintained but collapsed for all other consonants.

Second, the production experiment in Chapter 3 showed that high vowels in high-predictability environments are more likely to delete than in low-predictability environments, but this is not reflected in the corpus. The corpus was transcribed by Japanese linguists, who assumed that reduced vowels are always devoiced rather than deleted. To reflect the experimental results, vowels tran-

scribed as devoiced in the phone level transcriptions of the corpus were deleted using the following probabilities:

- Baseline by vowel height (V):
    - Short high vowels = 0.15
    - Other vowels = 0.05
- Changes by environments:
    - Non-reducing = V $-$ 0.05
    - Low-predictability = V $+$ 0.15
    - High-predictability = V $+$ 0.65

Articulatory data on the actual deletion rates of high vowels in Japanese are largely lacking in the literature. The probabilities above, therefore, were set to ensure that there is a reasonable number of consonants clusters in the input based on the indirect acoustic data from Chapter 3. The deletion probabilities only applied to vowels already transcribed as reduced in the corpus, so for example, if the word for /kita/ 'north' was transcribed as [kʲita] at the phone level with no reduced vowel, it was left unchanged. Conversely, if the same word was transcribed as [kʲi̥ta], the reduced [i̥] would have a 40% probability of being deleted (15% (high vowel baseline) + 15% (low-predictability context addition)). Similarly, if the word /suki/ 'to like' was transcribed as [su̥kʲi], the reduced [u̥] would have a 80% probability of being deleted (15% + 65%) because /s_k/ is a high-predictability reducing environment. Non-reducing vowels in non-reducing environments such as the first vowel in [çḁku] 'hundred' would never delete.

5.1 The model

The model proposed in this chapter simultaneously learns phonotactics and allophonic alternations, which can be represented as the flow chart in Figure 5.1 below. The phonotactic learning mechanism has access only to overt forms and induces constraints to infer possible surface forms. The mechanism calculates observed/expected ratios (e.g., (Pierrehumbert, 1993; Frisch et al., 2004) of all biphones in the input, then induces markedness constraints for underrepresented biphones and faithfulness constraints for overrepresented biphones.

The alternation learning mechanism has access to the lexicon and all its representational levels. The Corpus of Spontaneous Japanese provides orthographic transcriptions as well as phonemic and phonetic transcriptions, which are being used as underlying and overt forms, respectively. The orthographic form is being used as a stand-in for the semantic properties of a given word, and the model builds a lexicon by keeping track of each underlying form and one or more overt forms that correspond to the same meaning (i.e., orthographic form). The model induces rules that pair up underlying sequences to one or more corresponding overt sequences. These rules function like constraints and penalize underlying and overt sequences that do not match the rule. For example, if the model learns the rule /suk/ $\leftrightharpoons$ [s̥uk], an instance of /suk/ in the underlying form that does not correspond to a [s̥uk] in the overt form would incur a violation. There are no limits to how many overt forms can be paired with one underlying form and vice versa, as long as there is lexical support. What this means is that the model does not learn alternation rules explicitly but rather indirectly via multiple, simple rules that convert an input sequence to an output sequence. The phonotactic constraints and the conversion rules together form a constraint base that functions as the phonological grammar.

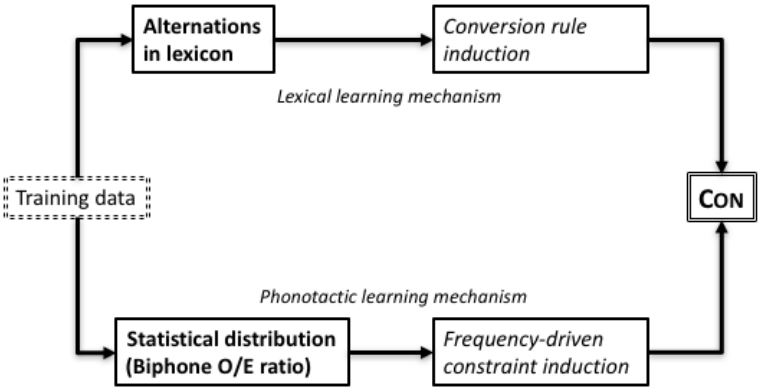

Figure 5.1: Simultaneous phonotactic and lexical learning.

Both learning mechanisms are employed in the model so that the resulting grammar can make contradicting predictions during perception and production. For example, the model could learn a phonotactic grammar that prohibits all consonant clusters and repair the overt input [skiː] 'ski' to /sukiː/ during perception, seemingly capturing a CV preference. However, given the underlying input /sukiː/, the phonotactic grammar would return [sukiː], failing to show high vowel deletion. The lexically driven alternation learning mechanism could counteract the phonotactic grammar's indiscriminate rejection of consonant clusters during production by learning that /suk/ sequences most often map to [sk], hence returning [skiː] as the output for /sukiː/.

5.1.1 Phonotactic learning

5.1.1.1 Phonotactic constraint induction

The phonotactic learning component of the model is based on Frequency-Driven Constraint Induction (Adriaans and Kager, 2010). The phonotactic learner calculates observed/expected ratios (O/E; Pierrehumbert, 1993; Frisch et al., 2004) of all biphones that occur in the input data, and induces constraints by setting thresholds on the O/E values. Markedness constraints are induced for underrepresented biphones with O/E ratios lower than 0.5 (e.g., **kt* assigns a violation for

every instance of *kt* in the output). The model by Adriaans and Kager (2010) also induces so-called 'contiguity' constraints for overrepresented biphones with O/E ratios higher than 2.0.

To illustrate the phonotactic learning mechanism of the model presented in (Adriaans and Kager, 2010), suppose that the model receives the following words as input: [kụ̥toː, kta]. Focusing on the word-initial biphones for the moment, the model learns by calculating observed/expected (O/E) ratios that [kụ̥, kt] are both likely to occur in the language. However, when the model receives [kubi, kumo, kugi, kud͡ʒi] as additional input, the O/E of [ku] goes up significantly while the O/E for [kụ̥, kt] drops. In this way, the O/E ratios of biphones rise and fall based on the data, and when the O/E ratio of a particular biphone sequence falls below 0.5, the model induces a markedness constraint (e.g., **kụ̥*: assign a violation for every instance of *kụ̥* in the output'). When the O/E ratio is 2.0 or higher, the model induces a CONTIGUITY constraint (e.g., CONTIG-*ku*: Assign a violation for every contiguous instance of *ku* in the input that is not also contiguous in the output).

There are a number of difference from the model presented in Adriaans and Kager (2010; STAGE) and the current model. First, STAGE is a lexiconless model built for the purpose of word segmentation in continuous, unsegmented speech. The phonotactic learning component of the current model is also lexiconless, learning strictly from overt forms. However, because the task of the current model is not word segmentation but the recovery of reduced vowels, the input string was shortened to consist of "intonational phrases" as segmented in the corpus, which could range from a single word to an entire sentence.

Second, instead of CONTIGUITY constraints that penalize elements adjacent in the input that are not also adjacent in the output, the phonotactic learner for the current model induces instead a constraint that functions as a general correspondence constraint, which will be called CORRESPONDENCE (hereafter COR) for its combined function as a MAX, DEP, and IDENT constraint. The COR constraints penalize length differences in the input and output as well as changes in segments. For example, given the input /p_1t_2/, CONTIGUITY-*pt* would disallow *[p_1at_2] but allow [$a_jp_1t_2$, $p_1t_2a_j$] since the adjacency of the input cluster is preserved in the output forms. Outputs that change

the identify of the input segments as in $p_1t_2/ \rightarrow [b_1d_2]$ is also allowed, since the contiguity of the corresponding input-output segments are preserved. COR-*pt* induced by the phonotactic learner of the current model would disallow epenthesis *[$a_j p_1 t_2$, $p_1 a_j t_2$, $p_1 t_2 a_j$] because [a_j] does not have a corresponding segment in the input. Deletion would also be disallowed as in /p_1t_2/ $\rightarrow$ *[p_1_, _t_2, $\varnothing$] because not all input segments have a corresponding segment in the output. Lastly, changes in the identity are also disallowed.

Third, instead of the strict domination of constraints used in STAGE, the current model uses weighted constraints (Legendre et al., 1990; Smolensky and Legendre, 2006) to allow cumulative effects of lower-ranked constraints in overcoming a higher-ranked constraint. For example, given the input /ɸuki/ and the hypothetical constraints and weights COR-*ɸu* (weight = 0.9), **uk* (weight = 0.5), **ki* (weight = 0.5), the faithful candidate [ɸuki] (weight = -1.0) would lose to the reduced candidate [ɸu̥ki̥] (weight = -0.9) despite obeying the highest weighted COR-*ɸu* constraint. The weights of the constraints induced by the model are determined by the expected value of the constraint's biphone (Adriaans and Kager, 2010).

Lastly, while STAGE uses features to capture phonological generalizations, the current model uses symbolic representations that reflect coarticulatory effects seen in both Chapters 3 and 4. For example, /ʃ/ is represented differently before /i, u/ (e.g., /ʃi/ $\rightarrow$ [ɕi] /ʃu/ $\rightarrow$ [ʃu]). The feature-based generalization mechanism was removed from the current model, not because features were deemed irrelevant to the problem at hand but rather to simplify the architecture of the model while representing the relevant phonetic details. Simulation results later in the chapter shows that the model fails to converge on an output when it encounters a sequence that was absent in the input. While including a feature-based generalization mechanism would help mitigate this issue, such a change would also raise the issue of how cross-linguistic, low level variations in phonetic implementation should be represented using abstract, and perhaps universal, features. This is discussed in the conclusion of this chapter.

5.1.1.2 Phonotactic constraints in action

Table 5.1 below shows how the input /sugi/ is evaluated by the phonotactic constraints learned by the model. The model learned the following COR constraints: COR-*su*, COR-*s̥u*, COR-*ug*^{j}, and COR-*g*^{j}*i* with O/E ratios of 3.30, 9.17, 17.02, and 6.61, respectively. The model also learned the following markedness constraints: **sg*^{j} and **g*^{j}*i̥* with O/E ratios of 0.06 and 0.19, respectively. The model did not induce any constraints for *u̥g*^{j} because the biphone had an O/E ratio of 1.76, falling between the thresholds for COR (O/E $\geq$ 2.0) and markedness (O/E $\leq$ 0.5) constraints. COR-*ug*^{j}, COR-*g*^{j}*i* and **g*^{j}*i̥* are excluded from Table 5.1 below for the sake of space and also to focus on the first vowel.

Candidates (b-d) each incur a violation of COR-*su* for not remaining faithful to the input sequence /su/. Candidate (c) incurs an additional violation of **sg*^{j}. The faithful candidate (a) is selected as the winner with no violations, resulting in a total weight of zero.

		/sugi/	COR-*su* (9.04e-04)	COR-*s̥u* (1.16e-04)	COR-*ug*^{j} (3.05e-05)	**sg*^{j} (2.52e-05)	*total weight*
✓	a.	[sug^{j}i]					0.0
	b.	[su̥g^{j}i]	−1				-9.04e-04
	c.	[sg^{j}i]	−1			−1	-9.29e-04
	d.	[zug^{j}i]	−1				-9.04e-04

Table 5.1: No reduction in voicing environment.

The following in Table 5.3 is an evaluation of /ski/, focusing on the initial cluster. In addition to the constraints shown above, the model also learned **sk*^{j} and **uk*^{j}, both with O/E ratios of 0.45. No constraint was induced for *u̥k*^{j} because the biphone had an O/E ratio of 1.40, falling below the induction thresholds. The faithful candidate (a) violates **sk*^{j} and is eliminated. Candidate (b) violates **uk*^{j}, leaving candidate (c) as the winner, which incurs no violations.

	/ski/	COR-*su* (9.04e-04)	COR-*su̥* (1.16e-04)	**uk*^{j} (2.84e-05)	**si̥* (2.65e-04)	**sk*^{j} (2.35e-04)	*total weight*
	a. [sk^{j}i]					−1	-2.35e-04
	b. [suk^{j}i]			−1			-2.84e-05
✓	c. [su̥k^{j}i]						0.00
	d. [si̥k^{j}i]				−1		-2.65e-04

Table 5.2: Repair through epenthesis.

The two tables above show that at least two clusters are prohibited by the phonotactic grammar, despite their occurrence in the input. The phonotactic learner induced 1,097 constraints in total, and a survey of those constraints revealed that only the following four clusters had O/E ratios high enough to induce COR constraints: *ɸk*, *ɸts*, and *tsk*, all of which are high-predictability environments. By comparison, there were ∼160 markedness constraints prohibiting all other heterorganic clusters that occurred in the training data (e.g., [kt]). It should be noted that the model does not induce any constraints for heterorganic clusters not encountered in the training data (e.g., [gt]), which means that the model learns a stronger bias against voiceless clusters (i.e, reducing environments) than voiced or mixed clusters.

The reason for this induced bias against clusters is, in fact, simple. Although consonant clusters do occur in Japanese, it is usually the result of a rather restricted process. Only high vowels typically reduce, and almost exclusively between two voiceless obstruents. This means that when all possible biphones of Japanese are considered, the relative number of clusters is rather small. Below is a modified table from (Maekawa and Kikuchi, 2005: p. 211) summarizing how common reducing environments are in a representative subset of the CSJ-Core.

Vowel	C_1	C_2	*N*	*% of total*
u *(N=39,481)*	[-voice]	[-voice]	10,999	27.86%
	[-voice]	[+voice]	14,984	37.95%
	[+voice]	[-voice]	5,689	14.41%
	[+voice]	[+voice]	7,809	19.78%
i *(N=47,905)*	[-voice]	[-voice]	13,599	28.38%
	[-voice]	[+voice]	12,775	26.67%
	[+voice]	[-voice]	9,326	19.47%
	[+voice]	[+voice]	12,205	25.48%

Table 5.3: Distribution of reducing environment, reproduced from Maekawa and Kikuchi (2005).

To summarize, despite high vowel reduction's status as a near-obligatory process, reducing environments are relatively uncommon when the entire language of Japanese is concerned, leading to a low O/E ratio and hence markedness constraints for clusters.

5.1.2 Alternation learning

5.1.2.1 The lexicon

A lexicon allows the model to keep track of what input forms correspond to what meaning (Apoussidou, 2007), and eventually acquire a paradigm over the lexicon. Since there are no semantic representations in CSJ-Core, orthographic forms provided in Unicode characters are used as stand-ins for the semantic properties of a given word. For each orthographic form, CSJ-Core also provides the phonemic (underlying) and phonetic (overt) forms associated with the word. For example, the words for 'to like' and 'an opening' are both /suki/ phonemically. The two words are orthographically different, however—<好き> 'to like' and <隙> 'an opening'—allowing the model to acquire them as separate words. Furthermore, the lexical learner, unlike the phonotactic learner is given access to both the phonemic and phonetic transcriptions. This allows the model to keep track of one or more phonetic forms that correspond to one phonemic form. This also means that the model is not learning phonemic forms on its own, but rather building connections among corresponding semantic (orthographic transcription), phonemic, and phonetic forms that the model is assumed to have

learned already. What this means for the model is that homophonous words can be distinguished from each other based on the orthographic representation of the words. Admittedly, this lexicon building process is only indirectly related to lexical acquisition. However, since the focus here is the role of an established lexicon rather than the acquisition of it, the current method suffices. It should be noted, however, that the size of the lexicon and the types of input the model receives during training might change the grammar that the model learns.

Furthermore, because the phonetic transcriptions were modified to delete a large percentage of the vowels that were originally transcribed as devoiced in the corpus (30%∼80% for high vowels), the model can keep track of phonetic variations between unreduced, devoiced, and deleted vowels. For example, shown below in Table 5.4 is a toy lexicon built from the corpus. The word 'to like' occurred 140 times in the CSJ-Core, 12 times with the the first vowel deleted, five times with both vowels devoiced, once with the first vowel deleted and the second vowel devoiced, and once with both vowels deleted. The word 'a opening' occurred once with the first vowel deleted. The word 'after/over' occurred 44 times, 42 times without reduction and twice with the first vowel deleted. With these words, the model would create a lexical dictionary that looks like the following:

Word	*Gloss*	*Underlying*	*Surface*
好き	'to like'	/suki/ (x140)	[sk^{j}i] (x110), [sk^{j}i̥] (x2), [sk^{j}] (x2), [su̥k^{j}i] (x24), [su̥k^{j}i̥] (x1), [su̥k^{j}] (x1)
隙	'an opening'	/suki/ (x1)	[sk^{j}i] (x1)
過ぎ	'after/over'	/sugi/ (x44)	[sug^{j}i] (x42), [sg^{j}i] (x2)

Table 5.4: Toy lexicon.

5.1.2.2 Conversion rule induction

With the lexicon in place, the model can now learn alternations between voiced and devoiced vowels via conversion rules. The alternation learning mechanism simply keeps track of the environments in which an underlying vowel surfaces as either unreduced or reduced and induces underlying-overt conversion rules. The observed probability of the overt form given an underlying form is assigned

as the weight of the conversion rule. This means that the model can learn multiple conversion rules involving the same underlying sequence, each with different weights. The conversion rules function as weighted constraints, where a violation is assigned for every instance of an input to output conversion that does not match the conversion rule. It should be noted that conversions rules indiscriminately map an input sequence to an output sequence. This means that alternations are learned indirectly through multiple conversion rules that target the same underlying high vowels, with the rule that has the highest weight in a given context dominating the evaluation.

The exact sequence length the learner keeps track of may vary depending on the language being acquired. For example, the model can induce biphone conversion rules or triphone conversion rules from the toy lexicon in Table 5.4. Focusing on the initial /CVC/, if the model were inducing biphone conversion rules, the model would learn that /su/ may phonetically be [su] with a probability of 0.227 (42 out of 185), [s̥u] with a probability of 0.141 (26 out of 185), or [s_] with a probability of 0.632 (117 out of 185). It also would learn that /uk/ may have the phonetic forms [uk^{j}], [u̥k^{j}], or [_k^{j}] with probabilities of 0.000, 0.184, and 0.886, respectively. The model would also learn that /ug/ can phonetically be [ug^{j}], [u̥g^{j}], or [_g^{j}] with probabilities of 0.955, 0.000, and 0.045, respectively. The resulting conversion rule with their corresponding weights are presented in Table 5.5 below.

conversion rule	*weight*
/su/ ≒ [su]	0.227
/su/ ≒ [s̥u]	0.141
/su/ ≒ [s]	0.632
/uk/ ≒ [u̥k^{j}]	0.184
/uk/ ≒ [k^{j}]	0.886
/ug/ ≒ [ug^{j}]	0.955
/ug/ ≒ [g^{j}]	0.045

Table 5.5: Example of biphone conversion rules and weights.

The process is the same for triphones, but there are fewer conversion rules because triphone rule induction involves keeping track of larger chunks. Again from the toy lexicon in Table 5.4, the model would learn that of the 35 times the phonemic sequence /suk/ occurred, it never had the

phonetic form [suk^{j}], but [s̥uk^{j}] had a probability of 0.184 (26 out of 141), and [sk^{j}] a probability of 0.886 (125 out of 141). For the phonemic sequence /sug/, the phonetic form [sug^{j}] had a probability of 0.955 (42 out of 44), [s̥ug^{j}] had a probability of 0.000, and [sg^{j}] a probability of 0.045 (2 out of 44). The model would therefore induce the following triphone conversion rules and weights:

conversion rule	*weight*
/suk/ ≒ [s̥uk^{j}]	0.171
/suk/ ≒ [sk^{j}]	0.829
/sug/ ≒ [sug^{j}]	0.943
/sug/ ≒ [s̥ug^{j}]	0.434
/sug/ ≒ [sg^{j}]	0.013

Table 5.6: Example of triphone conversion rules and weights.

The rest of this chapter uses triphone rules. A series of pilot simulations were run to test the effectivity of biphone conversion rules and triphone conversion rules, both separately and combined. Overall, models using triphone rules outperformed the others. The advantage of triphone rules over biphone rules is perhaps expected for production simulations, since Japanese high vowel reduction requires access to both consonants flanking the target vowel. On the other hand, triphone rules sometimes failed to converge on a single output when a test item contained a novel triphone sequence. An example of this is shown in Table 5.8. However, biphone rules also produced more unexpected outputs during perception simulations due to highly weighted [C] ≒ /VC/, which often overrode similar [C] ≒ /CV/. Also, combining the rules into a single set for evaluation led to middling performance, revealing no advantage over the triphone rules.

5.1.2.3 Conversion rules in action

Using the conversion rules from Table 5.6 but with the actual weights that the model learned from the entire corpus, let us consider the example in Table 5.7 below, where the input is the underlying form /suki/. The faithful candidate (a) violates /suk/ ≒ [sk^{j}] and /suk/ ≒ [s̥uk^{j}], since the candidate contains neither [sk^{j}] nor [s̥uk^{j}] that corresponds to the /suk/ in the input. Candidate (a) also violates

uki/ ≒ [kʲi] for retaining the /u/ vowel. The devoiced candidate (b) violates /suk/ ≒ [skʲ], /uki/ ≒ [ukʲi], and uki/ ≒ [kʲi]. Candidate (d) violates all of the rules. The deleted candidate (c) is chosen as the winner despite violating /suk/ ≒ [sṵkʲ] and /uki/ ≒ [ukʲi], with the highest total weight of -0.469.

	/suki/	/suk/ ≒ [skʲ] (0.242)	/suk/ ≒ [sṵkʲ] (0.047)	/uki/ ≒ [ukʲi] (0.442)	/uki/ ≒ [kʲi] (0.355)	*total weight*
	a. [sukʲi]	−1	−1		−1	-0.644.
	b. [sṵkʲi]	−1		−1	−1	-1.039
✓	c. [skʲi]		−1	−1		-0.469
	d. [sugʲi]	−1	−1	−1	−1	-1.089

Table 5.7: Correct deleted form selected.

Compare the evaluation above with an evaluation of the same input using just the phonotactic constraints, shown below in Table 5.8. The faithful candidate is selected as the optimal candidate because of the highly weighted COR-*su* constraint that keeps the vowel from reducing.

	/suki/	COR-*su* (9.04e-04)	COR-*sṵ* (1.16e-04)	**ukʲ* (2.84e-04)	**skʲ* (2.35e-04)	*total weight*
✓	a. [sukʲi]			−1		-2.84e-04
	b. [sṵkʲi]	−1				-2.17e-03
	c. [skʲi]	−1			−1	-1.14e-03

Table 5.8: Repair through epenthesis.

If the input is /sugi/, where the high vowel is in a voicing environment, the faithful candidate containing a voiced vowel is selected, as shown in Table 5.9, which is also the candidate chosen by the phonotactic constraints.

	/sugi/	/sug/ ≒ [sgʲ] (0.003)	/sug/ ≒ [sṵgʲ] (0.001)	/ugi/ ≒ [ugʲi] (0.947)	/ugi/ ≒ [gʲi] (0.037)	*total weight*
✓	a. [sugʲi]	−1	−1		−1	-0.041
	b. [sṵgʲi]	−1		−1	−1	-0.987
	c. [sgʲi]		−1	−1	−1	-0.985
	d. [sukʲi]	−1	−1	−1	−1	-0.988

Table 5.9: Correct unreduced form selected.

Where the alternation grammar would have trouble, however, is when input contains a sequence that the model has never encountered during the training phase. One such case in our simulations was the word /ʃugʲoː/ 'training', shown in Table 5.10. The model had never encountered the triphone /ʃugʲ/ during training, and thus cannot narrow down the candidate set. Because the palatalized voiced velar stop /gʲ/ is phonemic in Japanese, the conversion rules /ʃug/ ≒ [ʃug] and /ʃug/ ≒ [ʃugʲ] do not apply.

	/ʃugʲoː/	/ʃug/ ≒ [ʃug] (0.929)	/ʃug/ ≒ [ʃu̥gʲ] (0.071)	/ʃukʲ/ ≒ [ʃukʲ] (1.000)	*total weight*
!	a. [ʃugʲoː]				0.000
	b. [ʃu̥gʲoː]				0.000
	c. [ʃegʲoː]				0.000
	d. [ʃugoː]				0.000

Table 5.10: No optimal output.

Table 5.10 also shows that implementing a feature-based generalization mechanism would help the model's performance. Although the model never encountered /ʃugʲ/ during training, it did learn the conversion rule /ʃukʲ/ ≒ [ʃukʲ], which is different from the target sequence by a single feature, namely voicing.

5.1.3 Combining phonotactic constraints and conversion rules

Given the two components of the model, the EVAL mechanism of the model is tiered such that the conversion rules evaluate the candidates generated by GEN first, with the phonotactic constraints becoming active only when the conversion rules fail to narrow down the candidates to a single output. The two components were tiered in this way primarily to maximize the model's performance, but there is empirical support for the implementation, where speech processing seems to rely primarily on lexical processes with the phonotactic grammar becoming active when lexical activation fails (Shademan, 2006; Vitevitch and Luce, 1999). In lexical decision tasks, participants often take longer to process lexical items with high neighborhood densities (i.e., a high number of other, minimally

different lexical items, such ⟨bæt⟩ → ⟨pæt⟩, ⟨kæt⟩, ⟨but⟩, ⟨bit⟩, etc; McClelland and Elman, 1986; Norris, 1994; Vitevitch and Luce, 1998). On the other hand, participants process nonce words containing sequences with high phonotactic probabilities faster than nonce words containing low probability sequences (Vitevitch et al., 1997; Vitevitch and Luce, 1999). An apparent contradiction arises between these two lines of work in that sequences with high phonotactic probability are found in words with high neighborhood densities, and low probability sequences are found in words with low neighborhood densities. This apparent conflict has been attributed to there being lexical and postlexical tiers of speech processing, with lexical tier processes dominating lexical processing and postlexical processes dominating nonce-word processing (Shademan, 2006; Vitevitch and Luce, 1999). The speed of processing is relative within their lexical and postlexical tiers, however, meaning that slowly processed high density lexical items are still processed faster than quickly processed high probability nonce words. This was shown in Vitevitch and Luce (1999) in particular, and the overall slowness of nonce word processing is presumably due to high probability nonce words activating lexical neighbors. This lexical activation ultimately fails to resonate (match a lexical item), at which point the phonotactic grammar comes into effect.

The tiered implementation of the lexical conversion rules and the phonotactic grammar can be illustrated as below in Figure 5.2. Given an input, the GEN mechanism generates a set of candidates, which is then evaluated first by the conversion rules. One or more candidates that have been assigned the highest weight by the conversion rules are then passed on to the phonotactic grammar for further evaluation. If there still are more than one optimal candidate after phonotactic evaluation, the EVAL mechanism simply chooses one at random.

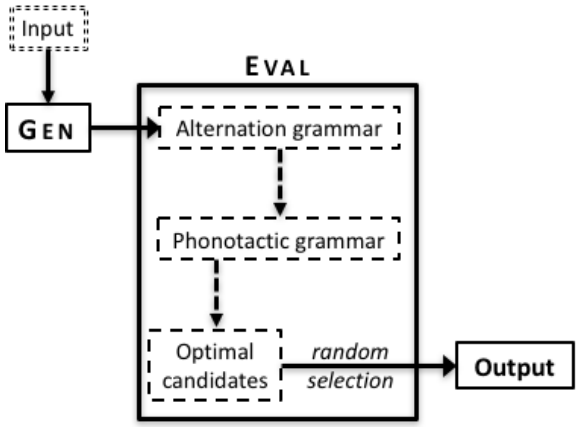

Figure 5.2: Tiered EVAL mechanism.

The main benefit of the tiered approach is that for inputs that require reduction, the conversion rules can eliminate non-reduced candidates before the phonotactic grammar can impose a CV preference to the output. This is illustrated in Table 5.11 below with the input /suki/ 'to like'. Because the conversion rules apply first, the only candidate that gets passed on to the phonotactic grammar is candidate (c) with high vowel deletion.

A. Lexical level

	/suki/	/suk/ ≒ [sk^{j}] (0.242)	/suk/ ≒ [s̥uk^{j}] (0.047)	/uki/ ≒ [uk^{j}i] (0.442)	/uki/ ≒ [k^{j}i] (0.355)	*total weight*
	a. [suk^{j}i]	−1	−1		−1	-0.644.
	b. [s̥uk^{j}i]	−1		−1	−1	-1.039
✓	c. [sk^{j}i]		−1	−1		-0.469
	d. [sug^{j}i]	−1	−1	−1	−1	-1.089

B. Postlexical level

	/suki/	COR-*su* (9.04e-04)	COR-*s̥u* (1.16e-04)	**uk^{j}* (2.84e-04)	**sk^{j}* (2.35e-04)	*total weight*
✓	c. [sk^{j}i]	−1			−1	-1.14e-03

Table 5.11: Two-tier grammar selects correct reduced output.

Additionally, if the conversion rules fail to eliminate candidates, as was the case with /ʃug^{j}oː/ 'training', it is up to the phonotactic grammar to select the optimal candidate. This is shown in

Table 5.12 below. Because the conversion rules fail to eliminate any candidate, all candidates get passed on for phonotactic evaluation. At the postlexical level, COR-*ʃu̥* does not apply to any of the candidates because there is no devoiced vowel in the input. With the exception of candidate (a), all candidates violate COR-*ug*^{j} for not keeping faithful to the underlying /ug^{j}/ sequence. Candidate (c) incurs an additional violation of the MARKEDNESS constraint **ʃe*. Because candidate (a) has the highest total weight of 0.0, it is selected correctly as the optimal candidate.

A. Lexical level

	/ʃugʲoː/	/ʃug/ ⇋ [ʃug] (0.929)	/ʃug/ ⇋ [ʃu̥gʲ] (0.071)	/ʃukʲ/ ⇋ [ʃukʲ] (1.000)	*total weight*
a.	[ʃugʲoː]				0.000
b.	[ʃu̥gʲoː]				0.000
c.	[ʃegʲoː]				0.000
d.	[ʃugoː]				0.000

B. Sublexical level

	/ʃugʲoː/	COR-*ʃu̥* (1.67e-05)	COR-*ugʲ* (3.05e-05)	COR-*gʲoː* (1.52e-05)	*total weight*
✓ a.	[ʃugʲoː]				0.000
b.	[ʃu̥gʲoː]		−1		-3.05e-05
c.	[ʃegʲoː]		−1		-3.05e-05
d.	[ʃugoː]		−1	−1	-4.57e-05

Table 5.12: Two-tier grammar selects correct non-reduced output.

Note that the underlying forms used for lexicon building assumed that there were no heterorganic consonant clusters underlyingly. This means that when confronted with a non-native phonotactic sequence such as /kra/ as an underlying input, the lexical conversion rules would not apply, leaving the work to the phonotactic grammar. On the other hand, if the same sequence is given but as the phonetic input [kra] for perception, the highest ranked conversion rule /kur/ ⇋ [kr] with a weight of 0.013 would apply, giving /kura/ as the optimal candidate.

5.2 Simulations

5.2.1 Overview

There are three main goals of the simulations in this chapter, the first of which is to verify whether a preference for CV structure can be learned with rules and constraints statistically induced from real spontaneous Japanese data. With numerous empirical evidence for a CV preference in Japanese listeners (see Chapters 2 and 4 for more discussion), it would be surprising to find that this preference is not learnable from the data. If, however, such a result is found, it would suggest that the illusory vowel epenthesis phenomenon in Japanese listeners reported in works by Dupoux and colleagues (Dupoux et al., 1999, 2011) and other related works cannot simply be attributed to phonotactic repair.

The second goal is to explore whether alternation learning can be captured statistically by inducing simple conversion rules based on phonemic-to-phonetic mappings observed in the lexicon. If the learning mechanism is successful, the results should show a context-based preference for reduced and unreduced vowels with high vowels deleting in high predictability environments and devoicing in low predictability environments. This contrasts with the phonotactic grammar of Japanese which predominantly prefers unreduced vowels and prohibits almost all consonant clusters.

Lastly, by combining phonotactic and alternation learning mechanisms, the model seeks to find additional evidence that phonotactics is indeed helpful in alternation learning. If the results show that the combined, tiered model outperforms each of its componential models, this would provide some support for the idea that phonotactic knowledge is helpful in alternation learning (Pater and Tessier, 2003; Jarosz, 2006; Tesar and Prince, 2007). On the other hand, it does not mean that phonotactics is not helpful in alternation learning if the tiered model does not perform the best. The model is admittedly in its primary stages and a failure to capture the interaction of the two

processes by the current model would require an investigation into the kinds of errors the model makes before a conclusion can be drawn.

5.2.2 Background assumptions

Like most grammars based on an Optimality Theoretic framework, the model is assumed to consist of GEN, CON, and EVAL mechanisms (Prince and Smolensky, 1993/2004). The simulations used the same tokens used in the production and perception experiments. To simplify the task and evaluation of the model, GEN was limited to manipulating the first vowel in production simulations (e.g., /suki/ → [ski, s̥ki, saki, suki]), and the second vowel in perception simulations (e.g., [epuko] → /epuko, epu̥ko, epako, epko/).

CON consists of phonotactic constraints and constraint-like conversion rules that were induced by the model as described above. The rules and constraints and rules are weighted as in Harmonic Grammar (Legendre et al., 1990; Smolensky and Legendre, 2006; Pater, 2012) rather than in strict domination as in classic OT. It is admittedly possible to use strict domination, which was in fact how constraint rankings were implemented in STAGE. However, the decision to use weighted constraints was to allow whether cumulative effects of lower-ranked constraints to overcome a higher-ranked constraint if possible.

The EVAL mechanism of the model uses the rules and constraints in CON to evaluate the set of candidates for each input, first using the conversion rules to filter out non-optimal candidates, then using the phonotactic constraints to further narrow down the choice of candidate. If more than one candidate remains after the phonotactic evaluation, one is chosen at random as the final output.

5.2.3 Methodology

The training data comes from a subset the Corpus of Spontaneous Japanese (CSJ). The entire CSJ contains about 7 million words spoken by 1,418 native speakers of standard modern Japanese, which

corresponds to roughly 650 hours of recorded speech. The CSJ consists mainly of monologues in the variety of academic presentation speech and simulated public speech. The speech data was recorded live using a head-mounted microphone at a sampling frequency of 48 kHz and 16-bit precision, which was then down-sampled to 16 kHz and stored. The entire corpus is phonemically transcribed and morphologically analyzed in terms of word-boundaries and parts of speech. The subset used as training data for the current model is called the CSJ-Core, which includes 500,000 words or 45 hours of speech from approximately 200 speakers. The CSJ-Core includes additional narrow phonetic transcriptions including vowel reduction status, as well as consonantal allophony (Maekawa, 2003; Maekawa and Kikuchi, 2005). As noted above, CSJ-Core transcriptions assume strict devoicing, so before the corpus was used as training data, vowels transcribed as devoiced in the corpus were deleted with probabilities ranging from 0%∼80% depending on context.

5.3 Production results

To compare each of the models' performance to that of the production experiment in Chapter 3, d′ values were calculated for both the experimental results and each of the models using the *dprime.mAFC(Pc, m)* function of the *psyphy* package in R (Knoblauch, 2014), where *Pc* is the proportion of correct responses and *m* is the number of answer choices. *Pc* was defined as the sum of hit and correct rejection rates divided by the total number of responses (Knoblauch and Maloney, 2012). Since the number of responses for reducing and non-reducing tokens were equal, the calculation can be simplified as shown in the equation below:

Proportion correct (*Pc*):

$$Pc = \frac{P(\text{reduced}|\text{reducing}) + P(\text{unreduced}|\text{non-reducing})}{2} \tag{5.1}$$

Devoiced and deleted outputs were collapsed as reduced because there is no contrast between devoiced and deleted tokens in Japanese. *m* was set at three with the following answer choices: reduced, unreduced, and wrong vowel. Although the actual number of possible candidates generated by the model's GEN is 11 (5 voiced vowels, 5 devoiced vowels, 1 deleted vowel), the output candidates were collapsed into the three categories because the primary task of the model is to reduce in reducing environments and do nothing elsewhere. In the case of the deleted vowel, the quality of the deleted vowel is not crucial here because its coarticulatory effects are reflected in the preceding consonant in the data. If the vowel in the output is neither the corresponding reduced nor unreduced of the input vowel, it's identity and reduction status does not matter. Also, using a low *m* value also results in a more conservative d-prime.

5.3.1 A1: Summary of production experiment

The model's production performance was compared to the experimental results in section 3.2.1.1 of Chapter 3, summarized again below in Table 5.13. Reduction rates were 99.4% in reducing environments and 10% non-reducing environments, the latter being driven mostly by /ʧ/-initial and /s/-initial tokens. This means that the hit and rejection probabilities were 0.994 and 0.900 (1 - false alarm), which gives a *Pc* of 0.947. A *Pc* of 0.947 with an *m* of 3 yields a d-prime of 2.672.

C_1	reducing	non-reducing
/k/	0.979	0.055
/ʃ/	0.986	0.080
/ʧ/	1.000	0.191
/ɸ/	1.000	0.042
/s/	1.000	0.214
/ç/	1.000	0.015
overall	0.994	0.100

Table 5.13: Reduction rate by token type from 22 Japanese participants.

5.3.2 A2: Simulation using phonotactics only

For all simulation results, the probabilities shown are the means from 22 test simulations, the same number of times as the number of participants in the production experiment. In the *reduced* columns are the rates in which the target high vowels in the input surfaced as the corresponding reduced vowel (e.g., /i/ → [i̥]) while in the *unreduced* columns are rates in which the underlying vowel surfaced faithfully as a full vowel (e.g., /i/ → [i]). The numbers in the *wrong vowel* columns refer to cases that fall into neither of these categories, where the output contained a different vowel altogether (e.g., /i/ → [ḁ, u]).

The results for the phonotactics only model are as shown in Table 5.14 below, which serves as the baseline model. With a hit rate of 8.9% and rejection rate of 98.5%, the phonotactics-only model has a *Pc* of 0.537 or a d-prime value of 0.677.

C1	reducing			non-reducing		
	reduced	unreduced	wrong vowel	reduced	unreduced	wrong vowel
/k/	0.093	0.465	0.442	0.000	1.000	0.000
/ʃ/	0.143	0.095	0.762	0.000	1.000	0.000
/ʧ/	0.164	0.055	0.782	0.000	1.000	0.000
/ɸ/	0.000	1.000	0.000	0.000	1.000	0.000
/s/	0.011	0.950	0.039	0.014	0.909	0.077
/ç/	0.123	0.114	0.764	0.000	1.000	0.000
overall	0.089	0.447	0.464	0.002	0.985	0.013

Table 5.14: Phonotactics only: Mean probabilities from 22 test simulations.

The results show first that there is a preference for unreduced CVCV structure in both reducing and non-reducing environments. In non-reducing environments, the model's performance essentially at ceiling with a non-reducing rate of 98.5%. Although a similar CVCV bias is evident in reducing contexts, the model's output is almost evenly split between unreduced vowels (44.7%) and wrong vowels (46.4%), suggesting that the output is chosen at random. A survey of the model's output, in fact, revealed that phonotactic evaluation often failed to eliminate any candidate. This failure seems to stem from two factors. First, the bias against reduction seems to stem mainly from

VC constraints. *Vk* and *Vt* constraints in particular are highly overrepresented in the input due to CVC Sino-Japanese roots, keeping the phonotactic model from reducing vowels in these contexts. Second, of the ∼1,400 biphones in the input, only ∼900 constraints were induced due to O/E ratios that fell between the two thresholds of 0.5 and 2.0. In the absence of markedness constraints against certain CV and VC sequences, the model simply had to choose candidates at random. This suggests that changing the constraint induction thresholds could help the model's overall performance in reducing tokens.

5.3.3 A3: Simulation using alternation only

The results for the alternation-only model are presented below in Table 5.15. An overall hit rate of 93.1% and rejection rate of 48.8% gives a *Pc* of .710 the alternation model has a d-prime value of 1.272.

C1	reducing			non-reducing		
	reduced	unreduced	wrong vowel	reduced	unreduced	wrong vowel
/k/	0.955	0.000	0.045	0.287	0.526	0.187
/ʃ/	1.000	0.000	0.000	0.459	0.412	0.129
/ʧ/	0.632	0.023	0.345	0.596	0.061	0.343
/ɸ/	1.000	0.000	0.000	0.005	0.904	0.091
/s/	1.000	0.000	0.000	0.000	1.000	0.000
/ç/	1.000	0.000	0.000	0.798	0.030	0.172
overall	0.931	0.004	0.065	0.358	0.488	0.154

Table 5.15: Alternation only: Mean probabilities from 22 test simulations.

The results show that alternation learning from the CSJ was somewhat successful but with some notable problems. Reduction rates in reducing environments are qualitatively more similar to the experimental results than the phonotactics-only model. For reducing tokens, the alternation-only model has an overall reduction rate of 93.1%. Wrong vowel errors, which are being treated as a miss here, are also low at 6.5% and are limited to /k/ and /ʧ/ tokens specifically. However, the reduction rate for /ʧ/ reducing tokens are much lower than expected at 63.2%, with wrong vowel errors making

of 34.5% of the model's output. Closer examination of the model's output revealed that the error in vowel choice was limited to /ʧit/ contexts. The reason for the error in this context was that there were no word-initial /ʧit/ sequences in the training data. This meant that no conversion rule applied to any input with word-initial /ʧit/, and thus the model was selecting a random candidate as optimal.

Although the reduction rates for non-reducing tokens are much lower than for the reducing tokens at 35.8%, this rate is also higher than the experimental results, which was 10%. Also, unlike in the case of reducing tokens where the alternation model made comparatively fewer vowel errors than the phonotactic model, the reverse was true for non-reducing tokens. All C_1 contexts had some wrong vowel error with the exception of /s/ tokens, resulting in an overall wrong vowel error rate of 15.4%. A close examination of the model's output revealed that in addition to the issue of novel sequences, the model had also learned a number of conversion rules from speech errors, such as /hid/ $\leftrightharpoons$ [çi̥d], leading the model to favor reduction in some non-reducing environments.

5.3.4 A4: Simulation using phonotactics and alternations

The results of the tiered model are presented in Table 5.16 below. While the overall numbers seem similar to the alternation-only model at first glance, closer examination of the results reveals one significant improvement. With the exception of /ʧ/-initial reducing tokens, the tiered model makes zero wrong vowel errors, driving up the overall hit rate to 93.6% from 93.1% and also driving up the rejection rate to 66.1%. This gives a *Pc* of 0.799 and a d-prime value of 1.648.

C1	reducing			non-reducing		
	reduced	unreduced	wrong vowel	reduced	unreduced	wrong vowel
/k/	0.950	0.050	0.000	0.250	0.750	0.000
/ʃ/	1.000	0.000	0.000	0.450	0.550	0.000
/ʧ/	0.664	0.045	0.291	0.556	0.444	0.000
/ɸ/	1.000	0.000	0.000	0.000	1.000	0.000
/s/	1.000	0.000	0.000	0.000	1.000	0.000
/ç/	1.000	0.000	0.000	0.778	0.222	0.000
overall	0.936	0.016	0.048	0.339	0.661	0.000

Table 5.16: Proposed model: Mean probabilities from 22 test simulations.

Even in the case of /ʧ/-initial reducing tokens where wrong vowel errors were not completely corrected, the error rates are still lowered slightly, improving the reduction rate accordingly. The reduction rates for non-reducing tokens are still high compared to the experimental results, but with zero wrong vowel errors, the model's 33.9% reduction rate bring it much closer to the experimental results.

5.3.5 Interim discussion: production

Summarized below in Table 5.27 are the hit rate, rejection rate, and d-prime value that correspond to the production experiment and each simulation. Confidence intervals also shown for the model simulations. Output candidates were chosen at random when the model failed to narrow the choice to a single, and some variance resulted from the 22 simulations that were run per model. Failure to converge on a single output candidate was mostly a problem for the phonotactic model as discussed earlier, and accordingly the confidence intervals are also wider.

simulation	hit	rejection	*d*-prime
Production experiment	0.994	0.900	2.672
Tiered model	0.936 ± 0.002	0.661 ± 0.003	1.648 ± 0.012
Alternation model	0.931 ± 0.003	0.488 ± 0.004	1.272 ± 0.013
Phonotactic model	0.089 ± 0.010	0.985 ± 0.009	0.677 ± 0.031

Table 5.17: Hit rate, correct rejection rate, and d-prime with 95% CI of all models.

Of the three models that were tested, the phonotactics-only model in Simulation A2 showed the strongest CVCV preference across all contexts, confirming that statistically induced phonotactic constraints can indeed capture the strong CVCV preference in Japanese. Additionally the alternation model fared better than the phonotactic model in predicting higher reduction rates in reducing environments. Although the alternation model was more successful than the phonotactic model in overall hit rate, it still suffered from wrong vowel errors in non-reducing environments. The alternation model's wrong vowel errors reveal first that real speech data is noisy and makes alternation learning difficult as noted by Peperkamp et al. (2006). Furthermore, the use of a largely segmental representation seems to make the model ineffective when it comes to novel sequences, a problem that arose in the /ʧ/ tokens. The tiered model showed that both of these issues can be partially resolved with the addition of phonotactic evaluation, supporting the notion that phonotactics can help alternation learning (Pater and Tessier, 2003; Tesar and Prince, 2007).

A possible solution to the remaining issues could be to have the model learn a CV bias. The rules and phonotactic constraints—the latter in particular—revealed a preference for certain VC sequences that overrode CV sequences. Given that Japanese high vowel reduction first and foremost relies on C_1, a CV bias could prove beneficial for the model's performance. In addition, Endress and Bonatti (2007) argue that when processing speech on-line there are two mechanisms at play. The first mechanism rapidly extracts structural information (syllables) from speech, which is then fed to a slower mechanism that detects statistical regularities within the extracted structures. In other words, there is an innate linguistic restriction on what sequences statistical computations can be applied to. If this is indeed the case, it does not seem unreasonable to think that there is a linguistic bias for calculating distributional probabilities of CV rather than VC sequences, especially given the crosslinguistic preference for CV sequences. The phonotactic learner currently induces constraints for all biphones indiscriminately, but one way a CV bias can be introduced is to modify the model so that CV constraints evaluate candidates before VC constraints, resulting in a three-tier model.

5.4 Perception results

As was the case in the production simulations, the model's performance was also compared to the perception experimental results presented in Section 4.2 of Chapter 4, evaluated for the model's ability to select the correct vowel in the output. Twenty-nine monolingual Japanese listeners (16 women, 13 men) were recruited for the perception experiment in Tokyo, Japan. The participants' ages ranged from 18 to 24 years. The stimuli were in the form $VC_1(V_T)C_2V$, where V_T is the target vowel and C_1 and C_2 are determined based on the stimulus group the token belongs to. The stimuli were divided into three groups: no reduction (No-Reduce) where vowel reduction is not expected, low predictability (Lo-Predict) where vowels should devoice, and high predictability (Hi-Predict) where vowels should delete. Below in Table 5.18 are the stimuli.

No-Reduce	eb_ko	ez_po	eg_to	ob_ke	oz_pe	og_te
Lo-Predict	ep_ko	eʃ_po	ek_to	op_ke	oʃ_pe	ok_te
Hi-Predict	eɸ_ko	es_po	eç_to	oɸ_ke	os_pe	oç_te

Table 5.18: Stimuli for Experiment 2.

There were 252 stimulus items in total. The stimulus forms shown in Table 5.18 were first recorded in a sound-attenuated booth with /i, a, u/ as target vowels (V_T). From each of the recordings, four additional tokens were created by removing from the right half of V_T (splice 1), the remaining half of V_T (splice 2), then half of the C_1 burst/frication noise (splice 3), then the remaining half of the C_1 burst/frication noise leaving only the closure for stops and $\sim$15 ms for fricatives (splice 4). The result of the splicing process is a gradual decrease of vowel coarticulatory information available in the burst/frication noise of C_1. In addition, a naturally produced vowel-less token was also recorded for each stimulus form, where C_1 was released if a stop. Since the focus of this chapter is to model the process of how reduced high vowels are recovered in Japanese, only the naturally produced vowel-less stimuli and the splice 2 tokens (i.e., spliced tokens with no V_T but full C_1 burst/frication) are discussed.

To evaluate the performance of the model relative to the empirical data, *d*-prime values were calculated again using the *dprime.mAFC(Pc, m)* function of the *psyphy* package in R (Knoblauch, 2014). For the perception simulations, *Pc* was defined as the sum of all correct identifications divided by the number of total responses. Because all contexts had the same number of tokens, this means that *Pc* simply the mean of correct identification rates. *m* was set at 4, since there were four answer choices in the experiment (i.e., <a, i, u, ∅>). GEN was also modified accordingly so that the target vowel could only be changed into one of the four choices.

Proportion correct (*Pc*):

$$Pc = \frac{P(\mathrm{CV_1C} \mid \mathrm{C\underset{\circ}{V}_1C}) + ... + P(\mathrm{CV}_n\mathrm{C} \mid \mathrm{C\underset{\circ}{V}}_n\mathrm{C})}{n} \text{ where } \underset{\circ}{\mathrm{V}} = [\underset{\circ}{\mathrm{a}}, \underset{\circ}{\mathrm{i}}, \underset{\circ}{\mathrm{u}}, \varnothing] \tag{5.2}$$

This definition of *Pc* can be further broken down in terms hit rates and correct rejection rates. A hit is when there is a coarticulated vowel in the stimulus, and the vowel is correctly identified in the response (e.g., [ek$\underset{\circ}{\mathrm{u}}$to] → /ekuto/). Similarly, a correct rejection is when there is no coarticulated vowel, as in the naturally produced vowel-less stimuli, and the response is ∅ (e.g., [espo] → /espo/). All results tables in the current section will therefore present both hit rates and correct rejection rates together as correct identifications.

5.4.1 B1: Summary of perception experiment

Summarized below in Table 5.19 is the rate of matches between the stimulus and participant response from the perception experiment presented in section 4.2.2 of Chapter 4. For example, given the stimulus [eb$\underset{\circ}{\mathrm{a}}$ko], the rate of /ebako/ responses were 74%. A full breakdown of responses are provided in Appendix A.1. The *mean* column shows that identification rates are high for /i, u/ but low for /a/. The results also show that participants reported hearing a vowel where there is none 66.9% of the time (= 1 − 0.331). The high rates for /i, u/ are unsurprising given that these vowels

are specifically targeted for reduction, and thus Japanese listeners are more inclined to recover them over other vowels.

	NoReduce			LoPred			HiPred			
	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to	*mean*
a	0.74	0.21	0.05	0.69	0.67	0.00	0.17	0.02	0.07	*0.291*
i	0.59	0.71	0.48	0.93	0.90	0.76	0.66	0.55	0.86	*0.716*
u	0.67	0.57	0.55	0.67	0.74	0.66	0.55	0.47	0.43	*0.590*
∅	0.41	0.50	0.38	0.36	0.38	0.19	0.33	0.33	0.10	*0.331*
overall	0.603	0.498	0.365	0.663	0.673	0.403	0.428	0.343	0.365	0.482

Table 5.19: Match rate of stimulus & response in perception experiment.

The proportion correct and *d*-prime are summarized below in Table 5.20. Given the *d*-prime metrics above, the proportion correct for the experiment is the mean of the identification rates for /a, i, u, ∅/, since these rates show the successful identification of the coarticulated vowels in the stimuli. *Pc*, therefore is 0.482 (= (0.291 + 0.716 + 0.590 + 0.331) ÷ 4), and the *d*-prime value for the experiment is 0.781. The overall Hit rate is lowered by the inclusion of /a/, a vowel that is typically not a target for recovery in Japanese. When /a/ is excluded, the *d*-prime increases to 0.981, since the new *Pc* is higher at 0.546 (0.716 + 0.590 + 0.331) ÷ 3). When calculating d-prime, *m* is always 4, since there were four answer choice regardless of which vowels are being included for calculating *Pc*.

simulation	hit	rejection	*d*-prime
All answers	0.532	0.331	0.781
/a/ removed	0.653	0.331	0.981

Table 5.20: Proportion correct and *d*-prime of perception experiment.

5.4.2 B2: Simulation using phonotactics only

For all simulation results, the probabilities shown are the means from 29 test simulations, the same number of times as the number of participants in the perception experiment. As in the experiment

results above, the probabilities shown below in Table 5.21 are the match rates between the input (stimulus) and the output of the phonotactics-only model (response). A full breakdown of the model's output is provided in Appendix A.2. The overall performance of the model is quite poor, with an overall identification rate of ∼25%.

	NoReduce			LoPred			HiPred			
	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to	*mean*
a	0.241	0.362	0.000	0.190	0.241	0.000	0.000	0.000	0.707	*0.193*
i	0.276	0.379	0.000	0.224	0.379	0.172	0.000	0.000	0.138	*0.174*
u	0.241	0.000	0.569	0.224	0.000	0.155	0.534	0.448	0.000	*0.241*
∅	0.207	0.379	0.500	0.310	0.310	0.655	0.534	0.500	0.190	*0.398*
overall	0.241	0.280	0.267	0.237	0.233	0.246	0.267	0.237	0.259	0.252

Table 5.21: Match rate of stimulus & response of phonotactics-only model.

The *d*-prime value of the phonotactics-only model can be calculated in the same way as with the experiment results. The same d-prime calculations as the perception experiment are shown below in Table 5.22. D-prime values are essentially at zero, revealing that the model is performing at chance.

simulation	hit	rejection	*d*-prime
All answers	0.203 ± 0.017	0.398 ± 0.036	0.006 ± 0.110
/a/ removed	0.208 ± 0.027	0.398 ± 0.036	0.080 ± 0.080

Table 5.22: Proportion correct and *d*-prime of phonotactic model with 95% CI.

5.4.3 B3: Simulation using alternation only

Unlike in the case of the production simulations where the alternation grammar utilized triphones, the alternation grammar for perception uses biphones instead, since the identity of the vowel to be recovered depends solely on the preceding consonant. Shown below in Table 5.23 are the match rates between the input (stimulus) and the output of the alternation-only model (response). A full breakdown of the model's output is provided in Appendix A.3. The overall performance of the

model is closer to the experimental results than the phonotactics-only model, but with notable differences. First, the /a/ identification rate is at 100% compared to the experiment's 29.1%. The reason for this high identification rate of /a/ was due to the fact that there were a small number of instances of [a, a̰] alternations in the lexicon, which triggered the induction of[C_1a̰C_2] $\leftrightharpoons$ /C_1aC_2/ type alternation rules. Unlike phonotactic constraints where a small number of instances would lead to a markedness constraint prohibiting the rare sequence, the alternation learner assigns overly high weights. Second, the identification rates for /i/ are lower than for /u/, which is the reverse of what was seen in the experiment.

	NoReduce			LoPred			HiPred			
	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to	*mean*
a	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*1.000*
i	0.000	1.000	0.293	0.000	1.000	1.000	0.534	0.172	1.000	*0.555*
u	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.276	*0.920*
∅	0.000	0.190	0.259	0.000	0.000	1.000	0.000	1.000	0.224	*0.297*
overall	0.500	0.798	0.638	0.500	0.750	1.000	0.634	0.793	0.625	0.693

Table 5.23: Match rate of stimulus & response of alternation-only model.

The *d*-prime value of the alternation-only model with and without /a/ are shown below in Table 5.24. The exclusion of the extremely high /a/ identification rate brings the model's *d*-prime closer to the target performance of the human participants, although the Hit rates are still higher.

simulation	hit	rejection	*d*-prime
All vowels	0.819 ± 0.008	0.297 ± 0.024	1.468 ± 0.072
High vowels only	0.739 ± 0.014	0.297 ± 0.024	1.127 ± 0.071

Table 5.24: Proportion correct and *d*-prime of alternation model with 95% CI.

5.4.4 B4: Simulation using phonotactics and alternations

Shown below in Table 5.25 are the match rates between the input (stimulus) and the output of the tiered model (response). A full breakdown of the model's output is provided in Appendix A.4. The

overall performance of the model is essentially identical to the alternation-only model as shown in both Tables 5.25 and 5.26. An examination of the model's outputs revealed that the alternation module allowed only one candidate to be passed on to the phonotactics grammar in most cases, hence the addition of the phonotactics module did little to the final output of the model.

	NoReduce			LoPred			HiPred			
	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to	*mean*
a	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*1.000*
i	0.000	1.000	0.259	0.000	1.000	1.000	0.500	0.259	1.000	*0.558*
u	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.259	*0.918*
∅	0.000	0.276	0.207	0.000	0.000	1.000	0.000	1.000	0.293	*0.308*
overall	0.500	0.819	0.617	0.500	0.750	1.000	0.625	0.815	0.638	0.696

Table 5.25: Match rate of stimulus & response of tiered model.

simulation	hit	rejection	*d*-prime
All vowels	0.826 ± 0.011	0.308 ± 0.023	1.479 ± 0.071
High vowels only	0.744 ± 0.015	0.308 ± 0.023	1.150 ± 0.087

Table 5.26: Proportion correct and *d*-prime of tiered model with 95% CI.

5.4.5 Interim discussion: perception

Shown below in Table 5.27 is a summary of all simulations, with and without /a/. The tiered and alternation-only models have essentially the same performance, again because the alternation module removes all but one candidate in most cases, leaving no other candidates for the phonotactics module to evaluate. The tiered and alternation-only models have higher Hit rates overall than what was observed in Japanese listeners. However, all models including the phonotactics-only model have similar rejection rates as the participants.

simulation	hit	rejection	*d*-prime
Perception experiment	0.532	0.331	0.781
Tiered model	0.826 ± 0.011	0.308 ± 0.023	1.479 ± 0.071
Alternation model	0.819 ± 0.008	0.297 ± 0.024	1.468 ± 0.072
Phonotactic model	0.203 ± 0.017	0.398 ± 0.036	0.006 ± 0.110

Simulation with /a/ removed

Perception experiment	0.653	0.331	0.981
Tiered model	0.744 ± 0.015	0.308 ± 0.023	1.150 ± 0.087
Alternation model	0.739 ± 0.014	0.297 ± 0.024	1.127 ± 0.071
Phonotactic model	0.208 ± 0.027	0.398 ± 0.036	0.080 ± 0.080

Table 5.27: Proportion correct and *d*-prime with 95% CI of all models.

The perception simulation results suggest that perceptual vowel epenthesis in Japanese listeners might not be driven primarily by the phonotactics of the language as has been argued in previous studies (Dupoux et al., 1999, 2011). First, the correct rejection rates are rather low, even for the alternation model. Second, the hit rate of the phonotactics-only model is below 25%, significantly worse than the other two models as well as the experiment results. Adding an alternation grammar module to the model, in fact, brings the model's performance more in line with the behavioral data, supporting a recent study by Durvasula and Kahng (2015), who found that the identity of perceptually epenthesized vowels in Korean listeners is better explained by an account that takes into consideration allophonic alternations of a language than a purely phonotactic account. The alternation grammar acquired by the model uses phonetic details encoded in the symbolic representations of segments and essentially maps any [C_1(V)C_2] sequence to an underlying /C_1VC_2/ sequence, further enforcing the phonotactic preference for a CVCV sequence. Furthermore, although not significantly so, the tiered model performed better than the alternation model, providing support to the notion that phonotactic knowledge can enhance alternation learning (Tesar and Prince, 2007; Pater and Tessier, 2003). This was also shown in the production simulations.

5.5 Discussion

The purpose of this chapter was to investigate how a learner of Japanese might acquire both the phonotactic preference for a CVCV structure during perception and the highly productive process of high vowel reduction that frequently violates this preference during production. The simulations show that a model that learns both phonotactic probabilities as well as alternation rules can successfully capture these seemingly contradictory processes in the phonological grammar of Japanese, but not without some issues. First, the results showed that the use of a largely segmental representation makes the model ineffective when it comes to novel sequences. One possible solution to this issue is to implement the feature-based generalization mechanism of StaGe that was not implemented in the current iteration of the model. The generalization mechanism of StaGe utilizes a single feature abstraction mechanism that combines two or more similar constraints. For example, since Cor-|gu| and Cor-|bu| differ only by place, a more general constraint Cor-|x∈{g,b};y∈{u}|, which says, "When the sequence /g/ or /b/ followed by /u/ is in the input, have a /g/ or /b/ followed by /u/ in the output," can be induced. The weight of such generalized constraints is calculated as the average of the componential specific constraints. This same generalization mechanism can be applied to the conversion rules as well, which would allow the model to deal with novel sequences more flexibly.

Furthermore, the current iteration of the model assumes that the learner has access to all levels of representation in the lexicon, including the underlying form, much of which a naive learner presumably never actually encounters during acquisition. The question that must be answered then is, how does the learner arrive at an abstract underlying representation from varying surface forms, if at all. Ogasawara (2013) found that Japanese listeners access reducible lexical items more quickly when presented with a reduced overt form. Ogasawara (2013) interprets this finding to mean that Japanese learners might be storing frequently reduced words as reduced. However, another compatible interpretation of the results is that while the vowel in a reducible word is

represented underlyingly (especially if the listener is literate), the underlying form has a stronger link to a reduced overt form (either deleted or devoiced, depending on the lexical item and listener), facilitating lexical access. Essentially, overt [CC] sequences are regarded as the preferred overt form of underlying /CuC/ and /CiC/ sequences, similar to how the conversion rules in the model are formulated. Regardless of the exact representation, it would be interesting to model how a learner might arrive at the lexical form of choice based solely on the surface forms.

5.6 Conclusion

The model presented in this chapter attempted some novel approaches to the issue of acquisition. Previous studies on general language acquisition have focused on unifying perception and production grammars through error-driven mechanisms, where learners change their current grammar only when they encounter a mismatch between what the grammar produces versus what the grammar perceives given the output form of adults. Generally, a lexicon is required to make error-detection possible (Escudero, 2005; van Leussen and Escudero, 2015). In an Optimality Theoretic approach to acquisition, a grammar is simply constraint rankings, and thus a number of constraint reranking strategies have been proposed, such as the Constraint Demotion Algorithm (CDA; Tesar, 1995) and the Gradual Learning Algorithm (GLA; Boersma and Hayes, 2001). However, empirical evidence shows that infants have a rather sophisticated knowledge of their native language before they begin to speak or even understand words. Since their knowledge does not seem dependent on the ability to produce speech, this suggests that perception and production grammars are separate.

Rather than making an explicit connection between production and perception grammars, the model instead bridges the two with conversion rules between varying surface forms and an established underlying form. Because multiple surface forms can correspond to a single underlying form, the learner does not need to make 'corrections' per se, but rather the learner can overcome errors as more and more surface forms are linked to the right underlying form, which eventually

will outweigh initial errors. Again, the model does require more work in addressing how a learner might arrive at an underlying form entirely from the surface form. This seems like a promising extension to the current model, and explicitly encoding the different lexical strata of Japanese might help improve the model's overall performance.

Furthermore, both CDA and GLA have been shown to be insufficient in learning phonotactics. The Maximum Entropy model (MaxEnt;Hayes and Wilson, 2008b) focuses specifically on phonotactic learning and constraint induction. It moves away from the assumption of universal constraints and argues that a universal feature set and a constraint format are sufficient enough to induce phonotactic constraints. MaxEnt is a batch learning model, where a set of all logically possible constraints are induced from the given a feature set, which are then ranked to maximize the probability of the entire input data set given. The constraints therefore still exist independently of the data much like in the cases of CDA and GLA. The current model, takes a more data-driven approach to constraint induction like STAGE, where constraints are not represented a priori but induced as the model processes the input data. It is encouraging that the two-tier grammar of the current model seems to make the correct prediction regarding the trajectory of how a Japanese infant might learn high vowel reduction, and the current data-driven approach to phonotactic learning may more accurately simulate language acquisition.

Lastly, the alternation learning mechanism of the model was applied to a non-typical alternation of specific segments within a root. Alternation typically refers to morpho-phonological alternations, such as in Korean, where the final coda in /pat^{h}/ 'field' surfaces as lax and unreleased utterance-finally (i.e., [pat$^{\neg}$]) but faithfully as aspirated when a vowel-initial particle follows (i.e., [pat^{h} + e] 'field + DAT'). The model's lexicon was morphological simple, consisting only of root words, which is perhaps part of the reason for the model's shortcomings in its current iteration. However, the conversion rule induction mechanism is flexible enough to be applied to a more morphologically complex lexicon, which could be a promising direction for the model, making it applicable to a broader range of alternations observed across different languages.

CHAPTER 6

Conclusion

6.0 Summary

The primary goal of this dissertation was to investigate the interaction of two sublexical processes in speech processing. A number of studies have shown that speakers control the amount of information in their speech signals to facilitate recoverability, at the expense of efficiency if necessary. The types of knowledge that have been shown to affect recoverability-driven gestural coordination are lexical predictability (Hall et al., forthcoming), phonetic contrasts in a given language (Silverman, 1997), ordering effects of features (Bladon, 1986), and also segments within a word or cluster (Chitoran et al., 2002). This dissertation investigated in detail the effects of phonotactic predictability on fine-grained coarticulatory cues using Japanese high vowel reduction as a test case. A secondary goal of the dissertation was to also better understand how reduced high vowels are manifested acoustically in Japanese, and how this process, which seemingly results in sequences that violate the strong CVCV phonotactics of the language, might be learned.

6.1 Experimental results

The main finding of this dissertation was that implementation of coarticulatory cues is not only language-specific, is also context-specific, where phonetic cues that are regarded as more informative, and thus aiding recoverability, are enhanced during production and more attended to during perception. The production experiment presented in Chapter 3 showed that Japanese speakers provide more coarticulatory information when a given context allows for both high vowels, making the target reduced vowel less predictable. The reverse was also found to be true, where the same phonetic cues that are enhanced in some contexts are weakened and not attended to in contexts where the cues carry less informative value. When the target vowel is highly predictable from context, speakers provided fewer perceptible cues, essentially deleting the vowel. Although the experiment was an acoustic one, providing indirect insight into how gestures are coordinated, the results nonetheless support previous findings that knowledge of one's native language affects whether the speaker chooses to enhance certain phonetic cues or not by reconfiguring gestural timing. Furthermore, lengthening of C_1 burst/frication in reduced environments was shown to only occur in stops and affricates. This is perhaps also due to recoverability-related issues, where stops and affricates are considered too short to carry sufficient perceptible cues compared to fricatives, and thus are lengthened.

At the core of the notion of recoverability is that certain cues are made more perceptible by the speaker for the sake of the listener. Chapter 4 presented a perception experiment that tested this idea in more detail. Specifically, if speakers are providing more coarticulatory cues in low-predictability contexts and less coarticulatory cues in high-predictability contexts with the listener in mind, the prediction then is that listeners are similarly inclined to attend to the same coarticulatory cues in low-predictability contexts where such information is more helpful. The results confirmed the prediction, showing first that Japanese listeners are more sensitive to high vowel cues than low vowels cues in general due to their experience with high vowel reduction. Second, participants

were even more sensitive to high vowel cues in low-predictability contexts, where the cues are more informative than in high-predictability contexts. The results also showed that while Japanese listeners have a tendency to recover a vowel even in the complete absence of vocalic information, it is not solely the need to repair phonotactically illegal sequences that drives Japanese listeners to perceive a vowel as argued by Dupoux et al. (1999, 2011). Instead, it seems Japanese listeners are in fact interpreting coarticulatory cues in the signal that previous studies have failed to control for.

The two experiments show a strong influence of phonotactic knowledge on how reduced vowels are coarticulated with preceding obstruents, further strengthening the idea that top-down linguistic knowledge informs the speaker and listener of the contexts in which recoverability might be in jeopardy. This knowledge leads the speaker to adjust coarticulatory cues accordingly to enhance perceptibility of the target segment. Likewise, the same knowledge informs the listener to pay closer attention to phonetic cues in contexts where he or she may require additional help for successful recovery, but ignore the same cues in contexts where such help is not necessary for recovery.

6.2 Representing high vowel reduction

In light of the results presented in Chapters 3 and 4, I return to the issue of representing high vowel reduction. Despite the popularity of high vowel reduction as a topic of study, the level in which the process occurs is seldom discussed explicitly. The prevalent assumption in the literature is that reduction in Japanese is a postlexical process (see Hirayama 2009 for a detailed discussion), but the question of how late in the phonological grammar reduction applies remains unanswered.

As discussed previously in Chapter 2, reduction applies to both underlying and epenthesized high vowels, suggesting that high vowel reduction applies after phonotactic repairs are made, late in the grammar. Shown again below in (25) are examples of Sino-Japanese roots that illustrate this point. The word ‘definitive answer’ in (25a) begins with a CVC root, which does not have an

underlying high vowel. The word ‘hard fight’ in (25b), on the other hand, begins with a CV root that does have an underlying high vowel. Regardless of their underlying representation, however, the high back vowel is reduced in both words, potentially resulting in heterorganic clusters that should be prohibited by the phonotactics of Japanese.

(25) a. ‘definitive answer (certain+reply)’

/kak+toː/ → [kak̥toː, kak̲ toː]

b. ‘hard fight (difficult+battle)’

/ku+toː/ → [k̥toː, k̲ toː]

This conflict between overt clusters and the phonotactics of Japanese, in fact, has led a number of scholars to propose that reduction must only result in a devoiced vowel (Hirayama, 2009; Kondo, 2005; Tsuchida, 1997). Reduced high vowels, however, often are phonetically deleted (Pinto, 2015; Kawahara et al., 2016; also Chapter 3 of this dissertation).

I proposed in Chapter 2 that phonotactic restrictions and high vowel reduction must apply at separate levels (Boersma, 2009; Hayes, 1999; Zsiga, 2000) to account for the reduction of both underlying and epenthetic vowels. In Figure 6.1 below are five levels of representation that are argued to be minimally necessary for a phonological grammar that can account for both production and perception (Boersma, 2007, 2009). Noted next to the levels of representation are types of constraints that are active in each level.

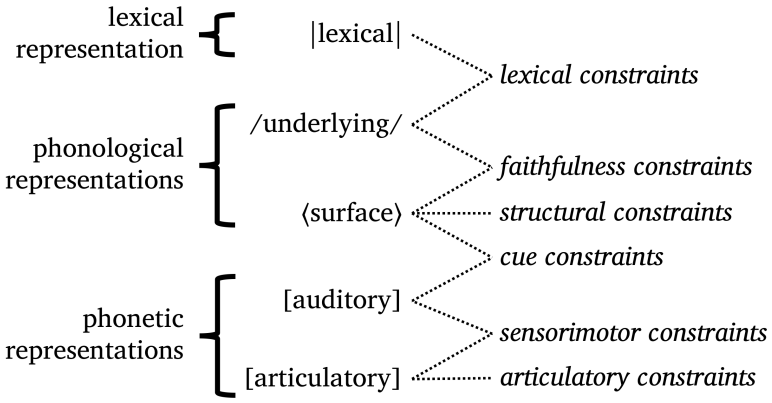

Figure 6.1: Minimally necessary levels of representation for a bidirectional model of phonology.

I provide here a brief summary of how perception from auditory forms and production of articulatory forms are handled within this framework (see Boersma 1998 *et seq.* for more details). Building on observations that phonetic implementation is largely language-specific, phonetic cues that are relevant to a given language are included as part of the phonological grammar in the form of *cue constraints*, which are active in the evaluation of both surface and auditory forms. Below in Table 6.1 is a tableau adapted from Boersma (2009) to show how a non-native, illegal coda is perceived by a Japanese listener. Superscript [^{k}] indicates a [k] release burst. Because the output is a surface form, the structural constraint CODACONDITION is active as well as a number of cue constraints that help interpret acoustic cues. The constraint */ / [^{k}] penalizes an auditory [k] burst that is not represented in the surface form. The cue constraints */o/ [ ] and */u/ [ ] penalize the epenthesis of /o/ and /u/ in the surface form, respectively.

[tak^{k}]	CODACONDITION	*/ / [^{k}]	*/o/ [ ]	*/u/ [ ]
/.tak./	*!			
/.ta./		*!		
☞ /.ta.ku./				*
/.ta.ko./			*!	

Table 6.1: Perception of obstruent coda in Japanese

During production, the input is a surface form and the output is a pair of auditory and articulatory forms derived from a given surface form. The necessity of an articulatory form is obvious, but a crucial assumption is that an auditory form is also necessary because the speaker must have an auditory target. Since both auditory and articulatory forms are being produced, cue constraints, sensorimotor constraints, and articulatory constraints evaluate the output. I follow Boersma (2009) and assume that sensorimotor knowledge is perfect for the sake of simplicity (i.e., sensorimotor constraints are ranked very high or very low).[1] An example of an articulatory constraint from Boersma (2009) is *[*periodic, final plosive*], which prohibits phonation in final obstruent codas in languages like English.

Since underlying and epenthesized high vowels both undergo reduction in Japanese, high vowel reduction must happen at the articulatory level, where structural constraints are inactive. There are a number of ways to formalize reduction using articulatory constraints. I propose here two possible variants. First, the articulatory constraint can be formulated as *[*vowel, short c.g.*], which prohibits a closed glottis gesture that is too short associated with a vowel. Another way to formulate high vowel reduction in terms of articulatory constraints would be to explicitly encode the context, such as *[*s.g.*][*vowel, short c.g.*][*s.g.*], which prohibits a short closed glottis gesture associated with a vowel between two spread glottis gestures.

The choice between the two articulatory constraints might depend on the speaker. Although high vowel reduction is nearly obligatory between voiceless obstruents, word-final reduction is optional and speaker-dependent. Therefore, for speakers who also reduce word finally, the constraint would be *[*vowel, short c.g.*], while for speakers who only reduce between two voiceless obstruents, the constraint would be *[*s.g.*][*vowel, short c.g.*][*s.g.*]. The constraints can include more detail, such as the duration of [c.g.], but the threshold for what constitutes a short glottal gesture may vary

[1] In practice, a perfect sensorimotor knowledge means that there is essentially no distinction between [auditory] and [articulatory] forms because the articulatory form would piggy-back on the auditory form during production and would be inactive during perception, making it unnecessary for the articulatory form to be discussed separately from the auditory form.

by speaker and speech style. For the current discussion, it suffices to simply point out that it is high vowels that are the most likely to reduce because they are inherently short regardless of the threshold.

Shown in Table 6.2 below is a simple tableau showing how the articulatory constraint must rank above a relevant cue constraint to produce a reduced output.

/.ku.toː./	*[*vowel, short c.g.*]	*/u/ [ ]
[k^{k}utoː]	*!	
☞ [k^{k}u̥toː]		
[k^{k}toː]		*!

Table 6.2: Production of reduced high vowel in Japanese

Although the example in Table 6.2 seemingly prefers devoicing to deletion, it would not be difficult to formulate articulatory constraints to prohibit not just the phonation but an entire vowel in certain contexts when the vowel is too short.

Shown below in Figure 6.2 is a summary of where rendaku, phonotactic repairs, high vowel epenthesis, and high vowel reduction apply in Japanese. Structural constraints like NoCoda, CodaCondition, Onset, and *Complex are active at the ⟨surface⟩ level. Violations in the underlying form during production are repaired differently depending on the lexical stratum a word belongs to. Violations in the auditory form during perception, on the other hand, are generally repaired through high vowel epenthesis. Reduction occurs at the articulatory level.

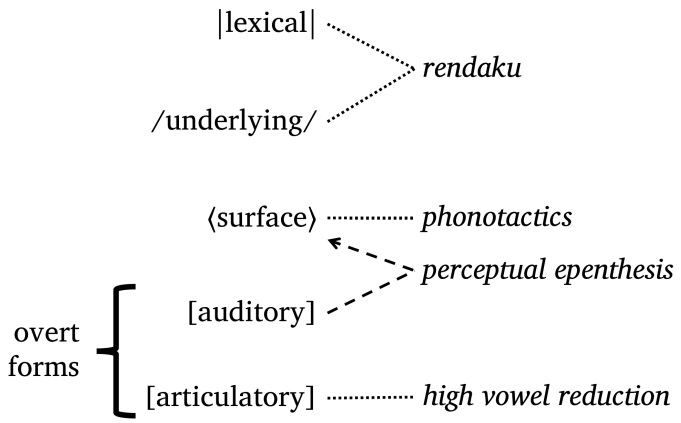

Figure 6.2: Phonological processes at each level of representation.

Given Figure 6.2 above, the process of perceptual repair would be represented as in (26) below. Using the Sino-Japanese compound 'definitive answer' as an example, the heterorganic cluster in the reduced auditory form [kakto:] is repaired via perceptual epenthesis at the ⟨surface⟩ level via *u*-epenthesis to satisfy CODACONDITION and *COMPLEX. The repaired form is then mapped to a corresponding underlying form, which allows access of the lexical form that involves two Sino-Japanese roots. Assuming a perfect sensorimotor knowledge, activation of an articulatory form is possible but not necessary.

(26) 'definitive answer' (perception)

[kakᵾto:, kakto:] → ⟨ka.ku.to:⟩ → /kakto:/ → |kak+to:|

The reduction process during production is essentially the reverse of above as shown in (27) below. The two morphemes involved in forming the word 'definitive answer' are accessed at the |lexical| level, which together result in a heterorganic cluster at the /underlying/ level. The cluster is repaired through *u*-epenthesis at the ⟨surface⟩ level to satisfy CODACONDITION and *COMPLEX. Finally, the epenthesized high vowel is reduced at the [articulatory] level. Even if the vowel were to

delete as in [kakto:], the consonant cluster in the output does not constitute a phonotactic violation since phonotactic constraints do not evaluate phonetic/overt forms.

(27) 'definitive answer' (production)

|kak+to:| → /kakto:/ → ⟨ka-ku.to:⟩ → [kakʉ̥to:,kakto:]

Because phonotactic constraints do not evaluate reduced forms during production in the proposed framework, the prediction then is that reduced outputs are not resyllabified. Using the Sino-Japanese compound 'definitive answer' again as an example, once the word is syllabified as ⟨ka.ku.to:⟩, the overt form after reduction applies would retain the structure even in the case of vowel deletion, leading to the overt form [ka.k.to:]. The experimental results in Chapter 3, in fact, showed that most onset obstruents in reduced syllables are longer than or comparable in length with the same obstruents in unreduced syllables, and the lack of a shortening effect suggests that reduced syllables did not undergo resyllabification.

The separation of ⟨surface⟩ and [overt] levels also helps explain how underlying obstruents can be produced as their allophones even in cases of vowel deletion. Consider the following examples in (28), where (28a) is a verb from the Yamato stratum and (28b) is a Sino-Japanese compound. Even though the underlying high vowels are ultimately deleted in the overt form, the initial obstruents still undergo allophony because it is triggered at the surface level first.

(28) a. 'to make (make+infinitive)'

|tukur+u| → /tukuru/ → ⟨tsukuru⟩ → [tskuru, *tkuru]

b. 'rectangle (four+angle)'

|si+kak| → /sikak/ → ⟨ʃikaku⟩ → [ʃkaku, *skaku]

In summary, although phonotactic constraints and high vowel reduction are apparently at odds with each other in Japanese phonology, the levels in which they apply differ. During production,

high vowel reduction can target both underlying and non-underlying high vowels because they are both represented at the ⟨surface⟩ level. Apparent phonotactic violations in the [overt] form that result from cases of high vowel deletion is also not problematic because structural constraints repair phonotactic violations at the ⟨surface⟩ level, not the [overt] level. This process is reversed in the case of perception, where reduced high vowels in the [overt] form (either devoiced or deleted) are recovered through phonotactic repair at the ⟨surface⟩ level due to cue constraints that allow the interpretation of low-level phonetic cues and structural constraints that prefer CV sequences.

6.3 Modeling simulations

Based on the results of Chapter 3, which showed that high vowel reduction sometimes results in vowel deletion, and on the results of Chapter 4, which showed that Japanese listeners are subject to a CVCV bias, Chapter 5 presented a computational model that investigated how the two contradictory aspects of Japanese phonology might be acquired from the same input data. The model induced and combined phonotactic constraints and alternation rules, trained on data from the Corpus of Spontaneous Japanese. The output of the model was compared to the results in Chapters 3 and 4. The production simulations showed that while a phonotactic grammar learned from the surface forms of Japanese learns a strong CVCV bias, this bias can be overcome by allowing lexicon-based alternation rules to evaluate the potential output candidates first. This confirms the idea that knowledge of phonotactics and alternations can fine tune each other Pater and Tessier (2003); Tesar and Prince (2007). As was the case with production simulations, the perception simulations also showed that allowing the alternation rules to evaluate the input first improves the model's performance.

Additionally, a phonotactics-only model which operated strictly on surface forms had the worst performance overall, and in fact the addition of a phonotactic grammar to the alternation grammar did little to improve the model's perception performance. This supports the findings

of Durvasula and Kahng (2015), where alternation rules better predicted the perceptual repairs performed by Korean speakers than phonotactics. This is not to say that phonotactics is not playing a role. On the contrary, the alternation rules are in fact functioning as a more fine-tuned set of phonotactic knowledge, since it has access to underlying forms and their corresponding surface forms. In other words, the interaction of the model's two tiers can be interpreted as lexical, top-down phonological rules processing the input first, which is then passed on to be analyzed by the sublexical phonotactic constraints that operate strictly over phonetically detailed surface forms. If the higher-level process successfully narrows down the output candidate, the lower-level process essentially does nothing, similar to what was shown in the experimental results of Chapters 3 and 4.

6.4 Future work

Although this dissertation focused on the effects of phonotactic predictability on reduction, predictability can come from multiple other sources, such as morphophonological processes, sentential contexts, and even differences between the grammars of the speaker and listener. If the insight from the literature on recoverability and this dissertation is correct, linguistic units that are deemed to be in less recoverable environments should be made more perceptible by the speaker for the sake of the listener. For example, schwa in English can sometimes delete in words such as "potato" /pə̲teɪto/ → [pʰ_tʰeɪɾo, pʰə̲tʰeɪɾo]. If predictability effects come from higher levels such as sentential context, then the schwa deletion should happen more often in (29) than in (30) below, which in fact have been shown to be the case in numerous studies (Lieberman, 1963; Fowler and Housum, 1987; Bybee and Scheibman, 1999; Aylett and Turk, 2004; Pluymaekers et al., 2005).

(29) French fries are made from potatoes.

(30) These puppies look like potatoes.

Regardless of where predictability comes from or whether the unit to be recovered is a feature, segment, or word, a question that arises is how does the speaker decide which cues would be helpful for the listener? It is not possible for the speaker to fully know whether or how the grammar of the listener differs, even if familiar with the listener. Recoverability-driven processes of the speaker then must be based on a mental model of the listener (Arnold, 2008), which undoubtedly will have inaccuracies. This perhaps explains why speech directed at listeners who are assumed to have drastically different grammars such as non-native speakers or infants differ from speech directed at listeners from similar linguistic backgrounds. While studies have shown that segments that carry more disambiguating information for words reduce less (van Son and Pols, 2003a,b), there have also been studies that show that speakers reduce the second mention of a word even if the listener is hearing the word for the first time (Bard et al., 2000). These contradictory findings show that the contribution of the listener's concerns during speech is still unclear. It would be helpful then to investigate how speakers formulate a model of the listener, if at all, and how this model can be manipulated through explicit instructions and/or interactions.

The experiments presented in this dissertation, as with all experiments, can be improved and applied to other languages to investigate more deeply into the contribution of predictability and recoverability in speech. The primary concern when constructing the carrier sentences used for the production experiment in Chapter 3 was to keep participants from figuring out which words were the stimuli, and although frequency was controlled for, the resulting sentences failed to control for semantic predictability. The stimuli used in the experiment were contexts in which reduction is considered to have been phonologized, which is apparent from the near-ceiling reduction rates. Fricative-fricative and affricate-fricative contexts are known to have significantly lower reduction rates and are considered more phonetic. Using these tokens would perhaps help better test for predictability from sentential contexts. It is admittedly difficult to control all of these factors in a single experiment, but it may be worthwhile endeavor. Furthermore, a follow-up to the perception

experiment, which was an identification task, such as an ABX or a 4IAX (4-interval forced choice) task would be helpful in probing further into how listeners interpret low-level phonetic cues.

The issue of language-specific, context-specific coarticulation can be further investigated through other phenomena, such as Korean nasal-lateralization (e.g., /ʃinla/ → [ʃilla] 'name of an ancient Korean kingdom') and the related lateral-nasalization (/tamlʲək/ → [tamnʲəkˀ] 'courage, guts'). Previous works have focused primarily on the role of sonority (Davis and Shin, 1999) or word-structure (Kang, 2002), but the question remains as to how sensitive Korean speakers are to nasality or laterality cues in these contexts, and whether predictability (phonotactic or semantic) plays a role in the degree of neutralization.

In regards to the computational model, there are a number of improvements that can be made, some of which were discussed in Chapter 5, such as underlying form learning and allowing for a more morphologically complex lexicon. The latter improvement in particular would allow the model to tackle not only the issue of more traditional, morphophonological alternations as mentioned in the previous chapter, but also the problem of L2 phonotactic learning. The crucial difference between L1 and L2 acquisition is that in the latter case, the learner already has a phonological grammar before any learning begins. Previous works on L2 learning, therefore, assume that the initial state of the L2 grammar is the acquired L1 grammar (Escudero, 2009; Trapman and Kager, 2009). This means that the model presented in Chapter 5 can serve as the initial state for L2 training.

6.5 Concluding remarks

This study focused on Japanese high vowel reduction as a test case for the interaction of phonotactics and coarticulation. A production experiment showed that high vowels are more likely to be devoiced, retaining more phonetic cues of the target vowel in contexts where the target vowel is less predictable phonotactically. Conversely, high vowels are more likely to delete if the context provides high predictability. Similarly, a perception experiment showed that listeners are more

sensitive to coarticulatory cues in contexts of low high vowel predictability. Additionally, listeners were sensitive to high vowel cues across all contexts, but were sensitive to low vowel cues only after stops, revealing a selective sensitivity to CV coarticulation. Listeners across different languages show selective sensitivity to phonetic cues, such as when native Spanish listeners fail to show sensitivity to F1 difference between [i, ɪ] (Kondaurovaa and Francis, 2010). What the perception experiment adds is that listeners can also have selective sensitivity to phonetic cues of the same vowels, depending on the phonotactic context.

Lastly, a computational model that combines lexicon-less phonotactic learning and lexicon-based alternation learning showed that the two processes can capture both the production and perception results better than either process alone, confirming the idea that phonotactic and alternation learning can fine tune each other. The results together suggest that top-down knowledge modulates bottom-up processes (i.e., enhancement of phonetic cues during production and attending to phonetic cues during perception).

Appendix: Perception simulation results

A Computational modeling results: Perception

A.1 B1: Perception experiment

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	**0.74**	0.21	0.05	**0.69**	**0.67**	0.00	0.17	0.02	0.07
i	0.00	0.03	0.09	0.00	0.00	**0.57**	0.10	0.03	**0.52**
u	0.12	**0.38**	**0.55**	0.09	0.19	0.26	0.21	0.47	0.28
∅	0.14	**0.38**	0.31	0.22	0.14	0.17	**0.52**	**0.48**	0.14

Table A.1.3: Responses for [VCa̰CV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	0.02	0.00	0.02	0.00	0.00	0.00	0.02	0.00	0.02
i	0.02	0.00	0.12	0.00	0.00	0.09	0.03	0.05	**0.47**
u	**0.67**	**0.57**	**0.55**	**0.67**	**0.74**	**0.66**	**0.55**	0.47	0.43
∅	0.29	0.43	0.31	0.33	0.26	0.26	0.40	**0.48**	0.09

Table A.1.4: Responses for [VCṵCV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	0.09	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
i	**0.59**	**0.71**	**0.48**	**0.93**	**0.90**	**0.76**	**0.66**	**0.55**	**0.86**
u	0.10	0.10	0.41	0.03	0.03	0.10	0.19	0.24	0.03
∅	0.22	0.16	0.10	0.03	0.07	0.14	0.16	0.21	0.10

Table A.1.5: Responses for [VC$_{\circ}^{i}$CV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	0.14	0.02	0.03	0.10	0.02	0.00	0.00	0.00	0.00
i	0.10	0.05	0.09	0.24	0.02	**0.55**	0.07	0.07	**0.76**
u	0.34	0.43	**0.50**	0.29	**0.59**	0.26	**0.60**	**0.60**	0.14
∅	**0.41**	**0.50**	0.38	**0.36**	0.38	0.19	0.33	0.33	0.10

Table A.1.6: Responses for [VCCV] environment.

A.2 B2: Phonotactics only

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	0.241	0.362	0.000	0.190	0.241	0.000	0.000	0.000	0.707
i	0.379	0.362	0.000	0.293	0.362	0.224	0.000	0.000	0.190
u	0.190	0.000	0.483	0.259	0.000	0.121	0.517	0.534	0.000
∅	0.190	0.276	0.517	0.259	0.397	0.655	0.483	0.466	0.103

Table A.2.1: Phonotactics only output for [VC$_{\circ}^{a}$CV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	0.276	0.276	0.000	0.345	0.379	0.000	0.000	0.000	0.586
i	0.328	0.379	0.000	0.155	0.276	0.103	0.000	0.000	0.224
u	0.241	0.000	0.569	0.224	0.000	0.155	0.534	0.448	0.000
∅	0.155	0.345	0.431	0.276	0.345	0.741	0.466	0.552	0.190

Table A.2.2: Phonotactics only output for [VC$_{\circ}^{u}$CV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	0.241	0.276	0.000	0.259	0.345	0.000	0.000	0.000	0.672
i	0.276	0.379	0.000	0.224	0.379	0.172	0.000	0.000	0.138
u	0.172	0.000	0.552	0.328	0.000	0.224	0.569	0.552	0.000
∅	0.310	0.345	0.448	0.190	0.276	0.603	0.431	0.448	0.190

Table A.2.3: Phonotactics only output for [VC̬iCV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	0.328	0.207	0.000	0.293	0.259	0.000	0.000	0.000	0.655
i	0.241	0.414	0.000	0.155	0.431	0.155	0.000	0.000	0.155
u	0.224	0.000	0.500	0.241	0.000	0.190	0.466	0.500	0.000
∅	0.207	0.379	0.500	0.310	0.310	0.655	0.534	0.500	0.190

Table A.2.4: Phonotactics only output for [VCCV] environment.

A.3 B3: Alternation only

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
i	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
u	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
∅	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.3.1: Alternation only output for [VC̬aCV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.241
i	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.293
u	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.276
∅	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.190

Table A.3.2: Alternation only output for [VC̬uCV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	1.000	0.000	0.224	1.000	0.000	0.000	0.466	0.310	0.000
i	0.000	1.000	0.293	0.000	1.000	1.000	0.534	0.172	1.000
u	0.000	0.000	0.207	0.000	0.000	0.000	0.000	0.276	0.000
∅	0.000	0.000	0.276	0.000	0.000	0.000	0.000	0.241	0.000

Table A.3.3: Alternation only output for [VC̥iCV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	1.000	0.310	0.362	1.000	0.000	0.000	1.000	0.000	0.328
i	0.000	0.224	0.190	0.000	1.000	0.000	0.000	0.000	0.207
u	0.000	0.276	0.190	0.000	0.000	0.000	0.000	0.000	0.241
∅	0.000	0.190	0.259	0.000	0.000	1.000	0.000	1.000	0.224

Table A.3.4: Alternation only output for [VCCV] environment.

A.4 B4: Phonotactics and alternation

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
i	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
u	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
∅	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.4.1: Tiered model output for [VC̥aCV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.224
i	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.276
u	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.259
∅	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.241

Table A.4.2: Tiered model output for [VC̥uCV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	1.000	0.000	0.190	1.000	0.000	0.000	0.500	0.379	0.000
i	0.000	1.000	0.259	0.000	1.000	1.000	0.500	0.259	1.000
u	0.000	0.000	0.328	0.000	0.000	0.000	0.000	0.172	0.000
∅	0.000	0.000	0.224	0.000	0.000	0.000	0.000	0.190	0.000

Table A.4.3: Tiered model output for [VC̥iCV] environment.

	NoReduce			LoPred			HiPred		
response	eb_ko	eg_to	ez_po	ep_ko	ek_to	eʃ_po	eɸ_ko	es_po	eç_to
a	1.000	0.224	0.276	1.000	0.000	0.000	1.000	0.000	0.172
i	0.000	0.241	0.207	0.000	1.000	0.000	0.000	0.000	0.259
u	0.000	0.259	0.310	0.000	0.000	0.000	0.000	0.000	0.276
∅	0.000	0.276	0.207	0.000	0.000	1.000	0.000	1.000	0.293

Table A.4.4: Tiered model output for [VCCV] environment.

Bibliography

Adriaans, Frans, and René Kager. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62:311–331.

Akamatsu, Tsutomu. 1997. *Japanese phonetics: Theory and practice*. Munchen: Lincom Europa.

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90:119–161.

Apoussidou, Diana. 2007. The Learnability of Metrical Phonology. Doctoral Dissertation, University of Amsterdam.

Archangeli, Diana, and Douglas Pulleyblank. 1994. *Grounded Phonology*. Cambridge, MA: MIT Press.

Arnold, Jennifer E. 2008. Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes* 23:495–527.

Aylett, Matthew, and Alice Turk. 2004. . the smooth redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47:31–56.

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem Lang* 59:390–412.

Bard, E.G., A.H. Anderson, C. Sotillo, Sotillo, M. Aylett, G. Doherty-Sneddon, and A. Newlands. 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language* 42:1–22.

Barlow, Jessica A. 2001. The structure of /s/-sequences: evidence from a disordered system. *Journal of Child Language* 28:291–324.

Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.

Becker, Michael, and Maria Gouskova. 2016. Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry* 47:391–425.

Beckman, Mary. 1982. Segmental duration and the 'mora' in Japanese. *Phonetica* 39:113–135.

Beckman, Mary, and A. Shoji. 1984. Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica* 41:61–71.

Beddor, P. S., K. B. McGowan, J. E. Boland, A. W. Coetzee, and A. Brasher. 2013. The time course of perception of coarticulation. *Journal of the Acoustical Society of America* 133:2350–66.

Bernstein Ratner, N. 1984. Patterns of vowel modification in mother–child speech. *Journal of Child Language* 11:557–78.

Bladon, A. 1986. Phonetics for hearers. In *Language for hearers*, ed. G. McGregor, 1–24. Oxford, Pergamon Press.

Blanchard, D., J. Heinz, and R. Golinkoff. 2010. Modeling the contribution of phonotactic cues to the problem of word segmentation. *Journal of Child Language* 37:487–511.

Boersma, Paul. 1998. *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.

Boersma, Paul. 2007. Some listerner-oriented accounts of h-aspiré in French. *Lingua* 117:1989–2054.

Boersma, Paul. 2009. Cue constraints and their interactions in phonological perception and production. In *Phonology in perception*, ed. Paul Boersma and Silke Hamann, 55–110. Berlin: Mouton de Gruyter.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32:45–86. Available as ROA-348 on Rutgers Optimality Archive, http://roa.rutgers.edu.

Browman, Catherine P., and Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6:201–251.

Browman, Catherine P., and Louis Goldstein. 1992a. Articulatory phonology: An overview. *Phonetica* 49:155–180.

Browman, Catherine P., and Louis Goldstein. 1992b. 'Targetless' schwa: An articulatory analysis. In *Papers in laboratory phonology ii: Gesture, segment, prosody*, ed. G. Docherty and R. Ladd, 26–56. Cambridge: Cambridge University Press.

Browman, Catherine P., and Louis Goldstein. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la Communication Parlée* 5:25–34.

Brown, R. W., and D. C. Hildum. 1956. Expectancy and the perception of syllables. *Language* 32:411–419.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82:711–733.

Bybee, Joan, and Joanne Scheibman. 1999. The effect of usage on degrees of constituency: the reduction of don't in english. *Linguistics* 37:575–596.

Byrne, Brian, and Ruth Fielding-Barnsley. 1995. Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology* 87:488–503.

Calamaro, Shira, and Gaja Jarosz. 2015. Learning general phonological rules from distributional information: A computational model. *Cognitive Science* 39:647–666.

Chaney, Carolyn. 1992. Language development, metalinguistic skills, and print awareness in 3-year-old children. *Applied Psycholinguistics* 13:485–514.

Chaney, Carolyn. 1994. Language development, metalinguistic awareness, and emergent literacy skills of 3-year-old children in relation to social class. *Applied Psycholinguistics* 15:371–94.

Chang, Charles B. 2010. First language phonetic drift during second language acquisition. Doctoral Dissertation, UC Berkeley.

Chitoran, Ioana, Louis Goldstein, and Dani Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In *Papers in Laboratory Phonology VII*, ed. Natasha Warner and Carlos Gusshoven. Berlin: Mouton de Gruyter.

Cho, Taehong, and Peter Ladefoged. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27:207–229.

Coetzee, Andries W., and Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *NLLT* 26:289–337.

Cohn, Abigail C. 1993. Nasalization in English: Phonology or phonetics? *Phonology* 10:43–81.

Cutler, Anne, Takashi Otake, and James M. McQueen. 2009. Vowel devoicing and the perception of spoken Japanese words. *Acoustical Society of America* 125:1693–1703.

Daland, R., and J. Pierrehumbert. 2011. Learning diphone-based segmentation. *Cognitive Science* 35:119–155.

Davidson, Lisa, and Jason Shaw. 2012. Sources of illusion in consonant cluster perception. *Journal of Phonetics* 40:234–248.

Davidson, Lisa, and Maureen Stone. 2003. Epenthesis versus gestural mistiming in consonant cluster production: an ultrasound study. In *Proceedings of WCCFL 22*.

Davis, Stuart, and Seung-Hoon Shin. 1999. The Syllable Contact Constraint in Korean: An Optimality-Theoretic Analysis. *Journal of East Asian Linguistics* 8:285–312.

Dehaene-Lambertz, G., E. Dupoux, and A. Gout. 2000. Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience* 12:635–647.

Dupoux, Emmanuel, Kazuhiko Kakeshi, Yuki Hirose, Christophe Pallier, and Jacques Mehler. 1999. Epenthetic vowels in japanese: A perceptual illusion? *Journal of Experimental Psychology* 25:1568–1578.

Dupoux, Emmanuel, Erika Parlato, Sónia Frota, Yuki Hirose, and Sharon Peperkamp. 2011. Where do illusory vowels come from? *Journal of Memory and Language* 64:199–210.

Durvasula, Karthik, and Jimin Kahng. 2015. Illusory vowels in perceptual epenthesis: The role of phonological alternations. *Phonology* 32:385–416.

Endress, Ansgar D., and Luca L. Bonatti. 2007. Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition* 105:247–299.

Ernestus, Mirjam. 2011. Gradience and categoricality in phonological theory. In *The Blackwell Companion to Phonology*, ed. Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 2115–36. Wiley-Blackwell.

Escudero, Paola. 2005. Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization. Doctoral Dissertation, University of Amsterdam.

Escudero, Paola. 2009. Linguistic perception of "similar" L2 sounds. In *Phonology in perception*, ed. Paul Boersma and Silke Hamann, 151–190. Mouton de Gruyter.

Faber, Alice, and Timothy J. Vance. 2000. More acoustic traces of "deleted" vowels in Japanese. In *Japanese/Korean Linguistics*, ed. Mineharu Nakayama and Charles J. Jr. Quinn, volume 9, 100–113.

Fais, Laurel, Sachiyo Kajikawa, Shigeaki Amano, and Janet F. Werker. 2010. Now you hear it, now you don't: Vowel devoicing in Japanese infant-directed speech. *Journal of Child Language* 37:319–340.

Farnetani, Edda, and Daniel Recasens. 2010. Coarticulation and connected speech processes. In *Handbook of Phonetic Sciences*, ed. William Hardcastle, John Laver, and Fiona Gibbon, volume 2, chapter 9, 316–352. Blackwell.

Flege, J. E., N. Takagi, and V. Mann. 1996. Lexical familiarity and English-language experience affect Japanese adults' perception of / r/ and / l/. *Journal of the Acoustical Society of America* 99:1161–1173.

Forrest, Karen, Gary Weismer, Paul Milenkovic, and Ronald N. Dougall. 1988. Statistical analysis of word-initial voiceless obstruents: preliminary results. *Journal of Acoustical Society of America* 84:115–123.

Fowler, Carol A., and Jonathan Housum. 1987. Talkers' signaling of new and old words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26:489–504.

Fowler, Carol A., and Elliot Saltzman. 1993. Coordination and coarticulation in speech production. *Language and Speech* 36:171–195.

Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language & Linguistic Theory* 22:179–228.

Fujimoto, Masako. 2015. Vowel devoicing. In *Handbook of Japanese Phonetics and Phonology*, ed. Haruo Kubozono, chapter 4. Mouton de Gruyter.

Fujimoto, Masako, Emi Murano, Seiji Niimi, and Shigeru Kiritani. 2002. Differences in glottal opening patterns between Tokyo and Osaka dialect speakers: Factors contributing to vowel devoicingevoicing. *Folia Phoniatrica et Logopedia* 54:133–143.

Fujimura, O., M.J. Macchi, and L.A. Streeter. 1978. Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech* 21:337–346.

Furukawa, Kaori. 2009. Perceptual similarity in loanword adaptation between Japanese and Korean. Master's thesis, University of Toronto.

Gibbon, Fiona, William Hardcastle, and Katerina Nicoladis. 1993. Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: a cross-linguistic investigation. *Language and Speech* 36:261–277.

Gick, Bryan, Ian Wilson, Karsten Koch, and Clare Cook. 2004. Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica* 61:220–233.

Gierut, Judith A. 1999. Syllable onsets: clusters and adjuncts in acquisition. *Journal of Speech, Language, and Hearing Research* 42:708–726.

Goswami, Usha. 2000. Phonological representations, reading development and dyslexia: Towards a cross-linguistic theoretical framework. *Dyslexia* 6:133–151.

Hall, Kathleen Curie, Elizabeth Hume, Florian Jaeger, and Andrew Wedel. forthcoming. The message shapes phonology. Forthcoming.

Hallé, André Pierre, Alberto Domínguez, Fernando Cuetos, and Juan Segui. 2008. Phonological mediation in visual masked priming: Evidence from phonotactic repair. *Journal of Experimental Psychology: Human Perception & Performance* 34:177–92.

Hamann, Silke, and Anke Sennema. 2005. Acoustic differences between German and Dutch labiodentals. *ZAS Papers in Linguistics* 42:33–41.

Han, Mieko S. 1994. Acoustic manifestations of mora timing in Japanese. *Acoustical Society of America* 96:73–82.

Hayes, Bruce. 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In *Functionalism and Formalism in Linguistics*, ed. Michael Darnell, Edith Moravscik, Michael Noonan, Frederick J. Newmeyer, and Kathleen M. Wheatley, 243–285. Amsterdam: John Benjamins.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: The early stages. In *Constraints in Phonological Acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld. Cambridge: Cambridge University Press.

Hayes, Bruce, Robert Kirchner, and Donca Steriade, ed. 2004. *Phonetically-based Phonology*. Cambridge: Cambridge University Press.

Hayes, Bruce, and Colin Wilson. 2008a. A maximum entropy model of phonotactics and phonotactic learning. *LI* 39:379–440.

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

Hirai, Sawako, Keichi Yasu, Takayuki Arai, and Kyoko Iitaka. 2005. Acoustic cues in fricative perception for Japanese native speakers. *Technical Report of Institute of Electronics, Information and Communication Engineers* 104:25–30.

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

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

Hsieh, Chih-Hsiang. 2013. The perception of epenthetic vowels in voiced and voiceless contexts in Japanese. Master's thesis, University of Kansas.

Hume, E., K. Johnson, M. Seo, G. Tserdanelis, and S. Winters. 1999. A cross-linguistic study of stop place perception. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 2069–2072.

Imai, Terumi. 2010. An emerging gender difference in Japanese vowel devoicing. In *A Reader in Sociolinguistics*, ed. Dennis Richard Preston and Nancy A. Niedzielski, volume 219, chapter 6, 177–187. Walter de Gruyter.

Imaizumi, Satoshi, Kiyoko Fuwa, and Hiroshi Hosoi. 1999. Development of adaptive phonetic gestures in children: evidence from vowel devoicing in two different dialects of Japanese. *JASA* 106:1033–1044.

Ito, Junko. 1986. Syllable Theory in Prosodic Phonology. Doctoral Dissertation, University of Massachusetts, Amherst. Published 1988. Outstanding Dissertations in Linguistics series. New York: Garland.

Ito, Junko, Yoshihisa Kitagawa, and Armin Mester. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese Argot. *Journal of East Asian Linguistics* 5:217–294.

Ito, Junko, and Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17:49–73.

Ito, Junko, and Armin Mester. 1995. Japanese phonology. In *Handbook of phonological theory*, ed. John Goldsmith, 817–838. Cambridge, MA: Blackwell.

Ito, Junko, and Armin Mester. 1996. Stem and word in Sino-Japanese. In *Phonological structure and language processing: Cross-linguistic studies*, ed. T. Otake and A. Cutler, 13–44. Berlin: Mouton de Gruyter.

Ito, Junko, and Armin Mester. 1999. The phonological lexicon. In *The handbook of Japanese linguistics*, ed. Natsuko Tsujimura, 62–100. Oxford: Blackwell.

Ito, Junko, and Armin Mester. 2003. Lexical and postlexical phonology in Optimality Theory: evidence from Japanese. *Linguistische Berichte* 11:183–207.

Ito, Junko, and Armin Mester. 2015. Sino-japanese phonology. In *Handbook of Japanese Phonetics and Phonology*, ed. Haruo Kubozono, chapter 7. Mouton de Gruyter.

Jacquemot, Charlotte, Christophe Pallier, Denis LeBihan, Stanislas Dehaene, and Emmanuel Dupoux. 2003. Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *The Journal of Neuroscience* 23:9541–9546.

Jarosz, Gaja. 2006. Rich lexicons and restrictive grammars – Maximum likelihood learning in Optimality Theory. Doctoral Dissertation, Johns Hopkins University.

Jun, Sun-Ah, and Mary Beckman. 1993. A gestural-overlap analysis of vowel devoicing in Japanese and Korean. *Paper presented at the 67th Annual Meeting of the Linguistic Society of America* .

Jusczyk, P. W., P. A. Luce, and J. Charles-Luce. 1994. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33:630–645.

Jusczyk, P.W., and R.N. Aslin. 1995. Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology* 29:1–23.

Kajikawa, Sachiyo, Laurel Fais, Ryoko Mugitani, Janet F. Werker, and Shigeaki Amano. 2006. Cross-language sensitivity to phonotactic patterns in infants. *JASA* 120:2278–2284.

Kang, Hyunsook. 2002. On the optimality-theoretic analysis of Korean nasal-liquid alternations. *Journal of East Asian Linguistics* 11:43–66.

Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 83:536–574.

Kawahara, Shigeto, and Shin-ichiro Sano. 2016. Rendaku and identity avoidance: Consonantal identity and moraic identity. In *Sequential voicing in Japanese: Papers from the NINJAL Rendaku Project*, ed. Timothy J. Vance and Mark Irwin, Studies in Language Companion Series, 176, 47–56. John Benjamins Publishing Company.

Kawahara, Shigeto, Jason Shaw, and James Whang. 2016. Targetless /u/ in Tokyo Japanese. In *Poster at the 16th Conference on Laboratory Phonology*. Ithaca, NY.

Kaye, Jonathan D. 1992. Do you believe in magic? The story of s+C sequences. *SOAS: Working Papers in Linguistics* 2:293–313.

Kindaichi, Haruhiko. 1995. *Shin Meikai Nihongo Akusento Jiten [Japanese Accent Dictionary]*. Sanseido.

Kiss, Zoltán, and Zsuzsanna Bárkányi. 2006. A phonetically-based approach to the phonology of /v/ in Hungarian. *Acta Linguistica Hungarica* 53:175–226.

Knoblauch, Kenneth. 2014. *psyphy: Functions for analyzing psychophysical data in R*. URL https://CRAN.R-project.org/package=psyphy, r package version 0.1-9.

Knoblauch, Kenneth, and Laurence T. Maloney. 2012. *Modeling Psychophysical Data in R*. Use R! Springer New York. URL https://books.google.com/books?id=AGMEsjX8LSMC.

Koda, Keiko. 1988. Effects of l1 orthographic representation on l2 phonological coding strategies. *Journal of Psycholinguistic Research* 18:201–222.

Kondaurovaa, Maria V., and Alexander L. Francis. 2010. The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: comparison of three training methods. *Journal of Phonetics* 38:569–87.

Kondo, Mariko. 2005. Syllable structure and its acoustic effects on vowels in devoicing. In *Voicing in Japanese*, ed. Harry van der Hulst, Jan Koster, and Henk van Riemsdijk, 229–246. Mouton de Gruyter.

Kong, Eun Jong, Mary Beckman, and Jan Edwards. 2012. Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics* 40:725–744.

Kubozono, Haruo. 2015. Loanword phonology. In *Handbook of Japanese Phonetics and Phonology*, ed. Haruo Kubozono, chapter 8, 313–362. Mouton de Gruyter.

Kumagai, Shuri. 1999. Patterns of linguopalatal contact during Japanese vowel devoicing. *The 14th International Congresses of Phonetics Sciences* 375–378.

Kurisu, Kazutaka. 2001. The Phonology of Morpheme Realization. Doctoral Dissertation, University of Santa Cruz.

Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2016. *lmertest: Tests in linear mixed effects models*. URL https://CRAN.R-project.org/package=lmerTest, r package version 2.0-30.

Lahiri, Aditi, and W. D. Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38:245–294.

Legendre, Geraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the twelfth annual conference of the cognitive science society*, 388–395. Mahwah, NJ: Lawrence Erlbaum Associates.

van Leussen, Jan-Willen, and Paola Escudero. 2015. Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology* 6:1000.

Lieberman, Philip. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6:172–187.

Maddieson, Ian, and Karen Emmorey. 1985. Relationship between semivowels and vowels: Cross-linguistic investigations of acoustic difference and coarticulation. *Phonetica* 42:163–174.

Maekawa, Kikuo. 2003. Corpus of Spontaneous Japanese: Its design and evaluation. *Proceedings of the ISCA & IEEE workshop on spontaneous speech processing and recognition (SSPR)* .

Maekawa, Kikuo, and Hideaki Kikuchi. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report. In *Voicing in Japanese*, ed. Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara. Mouton de Gruyter.

Malsheen, B.J. 1980. Two hypotheses for phonetic clarification in the speech of mothers to children. In *Child phonology: Perception*, ed. G.H. Yeni-Komshianm, J.F. Kavanaugh, and C.A. Ferguson, volume 2, 173–184. New York: Academic Press.

Mann, V. A. 1986. Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". *Cognition* 24:169–196.

Marslen-Wilson, W. D. 1987. Functional parallelism in spoken word-recognition. *Cognition* 25:71–102.

Martin, Andrew, Akira Utsugi, and Reiko Mazuka. 2014. The multidimensional nature of hyper-speech: Evidence from Japanese vowel devoicing. *Cognition* 132:216–228.

Martin, Samuel E. 1952. *Morphophonemics of Standard Colloquial Japanese*. Baltimore: Linguistic Society of America.

Mattingly, Ignatius G. 1981. Phonetic representation and speech synthesis by rule. In *The cognitive representation of speech*, ed. J. Myers, J. Laver, and Anderson J., 415–420. North-Holland Publishing Company.

Mattys, S. L., and P. W. Jusczyk. 2001. Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78:91–121.

Mattys, S. L., L. White, and J. F. Melhorn. 2005. Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General* 134:477–500.

Maye, J., J. F. Werker, and L. Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82:B101–B111.

McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16:331–399.

McCawley, James D. 1968. *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.

McClelland, J. L., and J. L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18:1–86.

Morelli, Frida. 1999. The phonotactics and phonology of obstruent clusters in Optimality Theory. Doctoral Dissertation, University of Maryland.

Moreton, Elliott, and Shigeaki Amano. 1999. Phonotactics in the perception of japanese vowel length: evidence for long-distance dependencies. *EUROSPEECH* 99:82–86.

Mugitani, Ryoko, Laurel Fais, Sachiyo Kajikawa, Janet F. Werker, and Shigeaki Amano. 2007. Age-related changes in sensitivity to native phonotactics in Japanese infants. *JASA* 122:1332–1335.

Näätänen, Risto, Anne Lehtokoski, Mietta Lennes, Marie Cheour, Minna Huotilainen, Antti Iivonen, Martti Vainio, Paavo Alku, Risto J Ilmoniemi, Aavo Luuk, Jüri Allik, Janne Sinkkonen, and Kimmo Alho. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385:432–434.

NHK. 1985. *Nihongo Hatsuon Akusento Jiten [Japanese Pronunciation Accent Dictionary]*. Tokyo: Nihon Hoso Kyokai [Japan Broadcasing Corporation].

Nittrouer, Studdert-Kennedy M., S., and R. S. McGowan. 1989. The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *J. Speech Hear. Res.* 32:120–132.

Nogita, Akitsugu, Noriko Yamane, and Sonya Bird. 2013. The Japanese unrounded back vowel /ɯ/ is in fact rounded central/front [ʉ - ʏ]. *Ultrafest VI* .

Norris, D. 1994. Shortlist: A connectionist model of con- tinuous speech recognition. *Cognition* 52:189–234.

Ogasawara, Naomi. 2013. Lexical representation of Japanese high vowel devoicing. *Language and Speech* 56:5–22.

Ogasawara, Naomi, and Natasha Warner. 2009. Processing missing vowels: Allophonic processing in Japanese. *Language and Cognitive Processes* 24:376–411.

Okada, Hideo. 1991. Japanese. *Journal of the International Phonetic Association* 21:94–96.

Okamoto, Shigeko. 1995. "Tasteless" Japanese: less "feminine" speech among young Japanese women. In *Gender articulated: Language and the socially constructed self*, ed. Kira Hall and Mary Bucholtz, 297–325. New York: Routledge.

Otake, T., G. Hatano, A. Cutler, and J. Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* .

Pater, Joe. 2012. Serial harmonic grammar and berber syllabification. In *Prosody Matters: Essays in Honor of Lisa Selkirk*, ed. Toni Borowsky, Shigeto Kawahara, Takahito Shinya, and Mariko Sugahara, 43–72. London: Equinox Press.

Pater, Joe, and Anne-Michelle Tessier. 2003. Phonotactic knowledge and the acquisition of alternations. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 1177–1180.

Peperkamp, Sharon, Rozenn Le Calvez, Jean-Pierre Nadal, and Emmanuel Dupoux. 2006. The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition* 101:B31–B41.

Pierrehumbert, Janet. 2001. Exemplar dynamics: word frequency, lenition, and contrast. In *Frequency effects and the emergence of linguistic structure*, ed. Joan Bybee and Paul Hopper, 137–157. Amsterdam: John Benjamins.

Pierrehumbert, Janet B. 1993. Dissimilarity in the Arabic verbal roots. In *Proceedings of the North East Linguistics Society*, ed. A. Schafer, volume 23, 367–381. Amherst, MA: GLSA.

Pinto, Francesca. 2015. High vowels devoicing and elision in japanese: a diachronic approach. In *International Congress of Phonetic Sciences 18*.

Pitt, Mark, and James McQueen. 1998. Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language* 39:347–370.

Pluymaekers, Mark, Mirjam Ernestus, and Harald R. Baayen. 2005. Lexical frequency and acoustic reduction in spoken dutch. *Journal of the Acoustical Society of America* 118:2561–2569.

Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA, and Oxford, UK: Blackwell. Available as ROA-537 on the Rutgers Optimality Archive, http://roa.rutgers.edu.

Prince, Alan, and Bruce Tesar. 1999. Learning phonotactic distributions. October, 1999.

Prince, Alan, and Bruce Tesar. 2004. Learning phonotactic distributions. In *Constraints in phonological acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, 245–291. Cambridge, UK: Cambridge University Press.

R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Saffran, J. R., R. N. Aslin, and E. L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274:1926–1928.

Sawashima, M. 1971. Devoicing of vowels. *Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo)* 6:7–13.

Selkirk, Elisabeth. 1982. The syllable. In *The structure of phonological representations*, ed. Harry van der Hulst and Norval Smith, volume Part II. Dordrecht: Foris Publications.

Shademan, Shabham. 2006. Is phonotactic knowledge grammatical knowledge? In *Proceedings of the 25th West Coast Conference on Formal Linguistics*, ed. Donald Baumer, David Montero, and Michael Scanlon, 371–379.

Sharoff, Serge. 2008. Lemmas from the internet corpus. URL http://corpus.leeds.ac.uk/frqc/internet-jp.num.

Shibatani, Masayoshi. 1990. *The Languages of Japan*. Cambridge: Cambridge University Press.

Shogakukan. 2013. Daijisen Zoubo/Shinsouban (Digital Version). URL http://dictionary.goo.ne.jp/.

Sibata, Takeshi. 1994. Gairaigo ni okeru akusento kaku no ichi [The location of accent in loanwords]. In *Kokugo ronkyū 4: Gendaigo hōgen no kenkyū [Studies in Japanese 4: Studies in modern Japanese and dialects]*, ed. Kiyoji Sato, 338–418. Tokyo: Meiji Shoin.

Silverman, Daniel. 1997. Phasing and Recoverability. Doctoral Dissertation, UCLA, Los Angeles.

Smith, Jennifer. 2006. Loan phonology is not all perception: Evidence from Japanese loan doublets. In *Japanese/Korean Linguistics 14*, ed. Timothy J. Vance.

Smolensky, Paul, and Géraldine Legendre. 2006. *The harmonic mind: From neural computation to Optimality-Theoretic grammar*. Cambridge, MA: MIT Press.

So, Connie K., and Catherine T. Best. 2010. Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech* 53:273–293.

Solé, Maria-Josep. 1992. Phonetic and phonological processes: The case of nasalization. *Language and Speech* 35:29–43.

van Son, R.J.J.H., and L.C.W. Pols. 2003a. An acoustic model of communicative efficiency in consonants and vowels taking into account context distinctiveness. In *Proceedings of ICPhS 15*, 2141–2144. Barcelona.

van Son, R.J.J.H., and L.C.W. Pols. 2003b. Information structure and efficiency in speech production. In *Proceedings of Eurospeech 2003*, 769–772. Geneva.

Stevens, Kenneth. 1998. *Acoustic phonetics*. Cambridge: MIT Press.

Stuart, Morag, and Max Coltheart. 1988. Does reading develop in a sequence of stages? *Cognition* 30:139–81.

Taft, Marcus. 2006. Orthographically influenced abstract phonological representation: Evidence from non-rhotic speakers. *Journal of Psycholinguistic Research* 35:67–78.

Taft, Marcus, and Gail Hambly. 1985. The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language* 24:320–335.

Tateishi, Koichi. 1989. Phonology of Sino-Japanese morphemes. In *University of massachusetts occasional papers in linguistics 13*, 209–235. Amherst: GLSA Publications.

Tesar, Bruce. 1995. Computational Optimality Theory. Doctoral Dissertation, University of Colorado, Boulder, CO.

Tesar, Bruce. 1997. An iterative strategy for learning metrical stress in Optimality Theory. In *Proceedings of the 21st annual boston university conference on language development*, ed. Elizabeth Hughes, Mary Hughes, Annabel Greenhill, Elizabeth Hughes, Mary Hughes, and Annabel Greenhill, 615–626. Somerville, MA: Cascadilla Press.

Tesar, Bruce, and Alan Prince. 2007. Using phonotactics to learn phonological alternations. In *CLS 39*, volume 2, 209–237.

Tesar, Bruce, and Paul Smolensky. 1996. Learnability in Optimality Theory (long version). Technical Report, Department of Cognitive Science.

Tesar, Bruce, and Paul Smolensky. 1998. Learning Optimality-Theoretic grammars. *Lingua* 106:161–196.

Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: The MIT Press.

Trapman, Mirjam, and René Kager. 2009. The acquisition of subset and superset phonotactic knowledge in a second language. *Language Acquisition* 16:178–221.

Tsuchida, Ayako. 1994. Fricative-vowel coarticulation in Japanese devoiced syllables: Acoustic and perceptual evidence. *Working Papers of the Cornell Phonetics Laboratory* 9:183–222.

Tsuchida, Ayako. 1997. Phonetics and phonology of Japanese vowel devoicing. Doctoral Dissertation, Cornell University.

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

Tsuchida, Ayako, Shigeru Kiritani, and Seiji Niimi. 1997. Two types of vowel devoicing in Japanese: Evidence from articulatory data. *Journal of Acoustical Society of America* 101:3177.

Vance, Timothy. 1987. *An Introduction to Japanese Phonology*. New York: SUNY Press.

Vance, Timothy J. 2008. *The sounds of Japanese*. New York: Cambridge University Press.

Vance, Timothy J. 2015. Rendaku. In *Handbook of Japanese Phonetics and Phonology*, ed. Haruo Kubozono, chapter 10, 397–441. Mouton de Gruyter.

Varden, J. Kevin. 1998. On high vowel devoicing in standard modern Japanese. Doctoral Dissertation, University of Washington.

Varden, J. Kevin. 2010. Acoustic correlates of devoiced Japanese vowels: velar context. *The Journal of English and American Literature and Linguistics* 125:35–49.

Varden, J. Kevin, and Tsutomu Sato. 1996. Devoicing of Japanese vowels by Taiwanese learners of Japanese. *Proceedings of International Conference on Spoken Language Processing* 96.2:618–621.

Vitevitch, Michael S., and Paul A. Luce. 1998. When words compete: Levels of processing in spoken word perception. *Psychological Science* 9:325–329.

Vitevitch, Michael S., and Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40:374–408.

Vitevitch, Michael S., Paul A. Luce, Jan Charles-Luce, and David Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40:47–62.

Warner, Natasha, and Takayuki Arai. 2001a. Japanese mora-timing: A review. *Phonetica* 58:1–25.

Warner, Natasha, and Takayuki Arai. 2001b. The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America* 109:1144–1156.

Wedel, Andrew. 2012. Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition* 4:319–355.

Werker, J.F., and C.E. Lalonde. 1988. Cross-language speech perception: initial capabilities and development change. *Developmental Psychology* 24:672–683.

Werker, J.F., and R.C. Tees. 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7:49–63.

Whang, James. 2016. Perception of illegal contrasts: Japanese adaptations of Korean coda obstruents. In *Proceedings of Berkeley Linguistics Society*, volume 36.

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130:96–115.

Wilson, Colin, Lisa Davidson, and Sean Martin. 2014. Effects of acoustic–phonetic detail on cross-language speech production. *Journal of Memory and Language* 77:1–24.

Yoshioka, H. 1981. Laryngeal adjustment in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica* 38:236–351.

Yoshioka, H., A. Löfqvist, and H. Hirose. 1982. Laryngeal adjustments in Japanese voiceless sound production. *Journal of Phonetics* 10:1–10.

Yuen, Chris L-K, and Kathleen Hubbard. 1998. Vowel devoicing and gender in Japanese. Presented at the LSA in New York, NY.

Zhang, Jie. 2004. *The role of contrast-specific and language-specific phonetics in contour tone distribution*, chapter 6, 157–190. Cambridge University Press.

Zimmerer, Frank, Rei Yasuda, and Henning Reetz. 2013. Architekt or Archtekt? Perception of devoiced vowels produced by Japanese speakers of German. In *Proceedings of Interspeech*, 417–420. Lyon.

Zsiga, Elizabeth. 2000. Phonetic alignment constraints: consonant overlap in English and Russian. *Journal of Phonetics* 28:69–102.