

# Learning to reduce and recover: the role of phonotactics and alternations in Japanese high vowel reduction

James Whang and Frans Adriaans

## Abstract

In standard modern Japanese, high vowels /u/ and /i/ alternate between full articulation and gestural reduction, depending on the consonantal context. When an unaccented high vowel occurs between two voiceless obstruents, it is either devoiced or deleted. This highly productive process poses a substantial challenge for phonological learning, as it results in output forms that violate the CVCV phonotactics of Japanese. The current study presents a computational model that induces and combines phonotactic constraints and alternation rules to account for the simultaneous acquisition of CVCV phonotactics and the process of high vowel reduction. The model is trained on data from the Corpus of Spontaneous Japanese, and the output of the model is compared to reduction rates obtained from Japanese speakers in a production experiment. Simulations show that both alternations and phonotactic constraints are necessary for acquiring Japanese high vowel reduction, and that phonotactics can aid in fine-tuning the learned alternations.

## 1 Introduction

### 1.1 Japanese high vowel reduction

In standard modern Japanese, high vowels /u/ and /i/ undergo gestural reduction when the vowels are (i) unaccented and (ii) flanked by two voiceless obstruents. For example, while the /u/ in /kúfi/

‘free use’ and /kufi/ ‘skewer’ both occur between two voiceless consonants, only /kufi/ ‘skewer’ undergoes reduction because the vowel is unaccented, surfacing as [kufi]. Likewise, the /u/ is unaccented in both /kuki/ ‘stem’ and /kugi/ ‘nail’, but only /kuki/ ‘stem’ undergoes reduction to [kuki] because the /u/ is flanked by two voiceless stops, namely /k/. Furthermore, an unaccented high vowel may also be reduced word-finally if it is preceded specifically by a voiceless fricative or affricate and is followed by a pause, as in /ikimasu#/ → [ikimasu#] ‘will go (formal)’ (Han 1962; Vance 1987). In what follows, we will focus on reduction in non-final position.

High vowel reduction is a highly productive process that is considered to be an integral feature of standard modern Japanese (Imai 2010). Even dictionaries with explicit instructions as to the reduction environments exist (Kindaichi 1995; NHK 1985). As such, the phenomenon is quite well-known and has received significant interest in the field of phonetics and phonology. While the phenomenon is more commonly called high vowel *devoicing* rather than reduction, there is an ongoing debate regarding whether the process concerns only loss of voicing (Han 1994, 1962; Tsuchida 1997, 2001; Varden 2010, 1998), or whether it entails complete deletion of the vowel (Hirose 1971; Yoshioka 1981; Vance 1987, 2008; Ogasawara 2013). While the exact acoustic manifestation of the so-called reduced vowel is debatable, there seems to be agreement that Japanese high vowel reduction is a gradient process that can range from partial loss of voicing to complete deletion of the vowel, and some additionally argue that high vowel reduction perhaps results in consonant clusters (Pinto 2015).

What is puzzling about the reduction of high vowels in Japanese is that the language is well-known for its strong phonotactic preference for a CVCV structure (Shibatani 1990). Numerous perception studies have shown that the phonotactic restriction is so strong that Japanese listeners report hearing the high vowel [u] and sometimes [i] between two consonants even in the complete absence of vocalic material (Dupoux et al. 1999, 2011; Dehaene-Lambertz et al. 2000). In other words, high vowel reduction in Japanese is a process that seemingly systematically violates the phonotactics of the language. The main goal of this paper, therefore, is to investigate how the process of Japanese high vowel reduction might be acquired by building a computational model

that combines phonotactic and alternation learning mechanisms and by testing whether the learned phonotactic constraints might be helpful in alternation learning.

## 1.2 The acquisition of Japanese high vowel reduction

Japanese is argued to have a strong preference for a CVCV structure, and a series of perception studies by Dupoux and colleagues, often referred to as the *ebzo* test, investigated the issue of what it is that Japanese listeners perceive when the phonotactic preference is in conflict with the surface form they encounter (Dupoux et al. 1999, 2011; Dehaene-Lambertz et al. 2000). In other words, do Japanese listeners perceive clusters as clusters or something that conforms to the CVCV preference? The results showed that even in the complete absence of a vocalic segment, Japanese listeners seem to perceive a vowel, and the vowel [u] and sometimes [i] in particular. Stated differently, the illusory epenthesis process is the reverse of what happens during high vowel reduction, where there is a strong inclination to recover during perception the same high vowels that are systematically reduced during production.

The fact that the phonological grammar of Japanese requires learning a strong phonotactic preference for a CVCV structure while also learning that this otherwise militant restriction is violable in certain contexts poses a substantial problem for the learner. The question then is, what do we know about the acquisition of high vowel reduction in Japanese children? Before discussing the behavioral evidence from children, we first describe 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? Since it has been argued that Japanese has a strong preference for a CVCV structure (Shibatani 1990), on the one hand Japanese IDS can use canonical, unreduced forms to facilitate structure learning, 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 underlying CVCV structure.

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). High vowel reduction rates 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 reveal 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 rate was around 85% for lexical words and around 20% for nonce words. Furthermore, Fais et al. also report 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 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. Under the assumption 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 in-

fants. 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 consonant clusters as a result of reduced forms produced by adult speakers.

### **1.3 Models of phonological learning**

Most computational models of phonological acquisition fall broadly into two categories: 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; Albright and Hayes 2003; Hayes and Wilson 2008). 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 2008) learns the phonotactic grammar of a given language through constraint induction, 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.

## **1.4 The current study**

The current study combines phonotactic and alternation learning processes into single model. In doing so, the proposed model allows for a more complex view on the learning of Japanese high vowel reduction. Phonotactic learning is included in the model because phonotactics is thought to be learned at an early age, and therefore likely interacts with the learning of high vowel reduction. This interaction is particularly interesting since the production of reduced vowels is in direct conflict with the predominant CVCV phonotactics. In addition to a phonotactic component, the model includes an alternation learning component. This is because high vowel reduction is essentially a process of alternation. 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, we aim 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 gives us 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 is evaluated against new empirical data collected from 22 monolingual Japanese speakers recruited in Tokyo, Japan. 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. The corpus will be described in more detail in §3.3 below.

## 2 The model

The model proposed in this paper assumes that given a linguistic input, a learner simultaneously learns phonotactics and allophonic alternations, which can be represented as the flow chart in Figure 1 below. The phonotactic learning mechanism has access only to the surface forms of words. 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. Because the Corpus of Spontaneous Japanese provides orthographic transcriptions as well as phonemic and phonetic transcriptions, the model first builds a lexicon by keeping track of the



phonemic and one or more phonetic forms that correspond to the same meaning (i.e., orthographic form). The model induces bidirectional rules that pair up an underlying sequence to a surface sequence as observed in the lexicon. These rules function like constraints and penalize underlying sequences that do not surface as the sequence prescribed by the rule. For example, if the model learns the rule  $/bag/ \rightleftharpoons [bak]$ , an instance of  $/bag/$  in the underlying form that does not correspond to a  $[bak]$  in the surface form would incur a violation. There are no limits to how many surface 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 alternations rules explicitly but rather indirectly via multiple, simple input-output conversion rules. The phonotactic constraints and conversion rules together form a constraint base that functions as the phonological grammar.

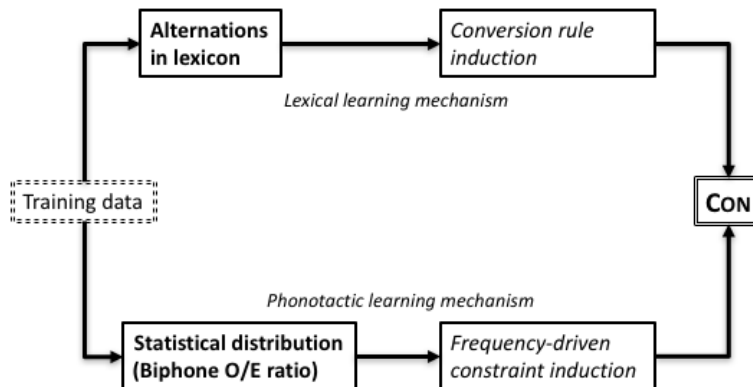


Figure 1: Simultaneous phonotactic and lexical learning.

## 2.1 Phonotactic learning

### 2.1.1 Phonotactic constraint induction

The phonotactic learning process of the model is similar to STAGE (Adriaans and Kager 2010), a model built for the purpose of word-segmentation in continuous, unsegmented speech. Like STAGE, the phonotactic module is an unsupervised, lexicon-less constraint induction model, but rather than using the constraints for word-segmentation, the constraints are applied to a new problem, namely the recovery of reduced vowels. STAGE is a statistical learner, which calculates the

observed/expected ratios (O/E) of all biphones that occur in the input data to induce phone-specific constraints when an assumed threshold is reached. Based on the O/E calculations, a frequency-driven constraint induction mechanism induces OT-style, MARKEDNESS and FAITHFULNESS constraints.

To illustrate the phonotactic learning mechanism of the model, suppose that the model receives the following words as input: [kato:, karu]. Focusing on the word-initial biphones for the moment, the model learns by calculating observed/expected (O/E) ratios that [kə, ka] are both likely to occur in the language. However, when the model receives [kabi, kamo, kagi, kadʒi] as additional input, the O/E of [ka] goes up significantly while the O/E for [kə] 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. When the O/E ratio is 2.0 or higher, the model induces a FAITHFULNESS constraint (e.g., IO-|ka|: Assign a violation for every instance of ka in the input that is not also in the output). It should be noted that these thresholds have been set arbitrarily, and may be changed if needed. Phonotactic constraints are bracketed as |CV| throughout the paper, with MARKEDNESS constraints represented as \*|kə| and interpreted as ‘assign a violation for every instance of [kə] in the output’. FAITHFULNESS constraints are represented as IO-|ka|, and interpreted as ‘assign a violation for every instance of |ka| in the input that is not also in the output’. In contrast to STAGE, which uses strict domination, each constraint in the current model is assigned a weight based on the observed value of the biphone.

### 2.1.2 Phonotactic constraints in action

Table 1 below shows some examples of phonotactic constraints that were learned by the model, along with their corresponding weights. The constraints are used to evaluate the input /kudo:/. The faithful candidate (a) incurs no violation, resulting in a total weight of 0.0000. Candidates (b-d) each incur a violation of the FAITHFULNESS constraint IO-|ku| for not remaining faithful to the underlying /ku/ sequence. Candidates (b) and (c) also incur a violation of IO-|ud| for not remaining faithful to the underlying |ud| sequence.

|                                   | IO- ka <br>(0.0300) | IO- ku <br>(0.0120) | IO- k <sub>u</sub>  <br>(0.0060) | IO- ud <br>(0.0006) | <i>total weight</i> |
|-----------------------------------|---------------------|---------------------|----------------------------------|---------------------|---------------------|
| ✓ a. /kudo:/ ~ [kudo:]            |                     |                     |                                  |                     | 0.0000              |
| b. /kudo:/ ~ [k <sub>u</sub> do:] |                     | −1                  |                                  | −1                  | -0.0126             |
| c. /kudo:/ ~ [kado:]              |                     | −1                  |                                  | −1                  | -0.0126             |
| d. /kudo:/ ~ [gudo:]              |                     | −1                  |                                  |                     | -0.0120             |

Table 1: No reduction in voicing environment.

Because the phonotactics of Japanese systematically prefers CVCV sequences, an output candidate with an epenthised vowel is selected as the winner when a CCV sequence is given as input, as was shown by the *ebzo* tests (Dupoux et al. 1999, 2011). This is illustrated in Table 2 below with a nonce input [kto:].

|                       | IO- ku <br>(0.0300) | IO- ka <br>(0.0120) | * kt <br>(0.0020) | * ku: <br>(0.0004) | <i>total weight</i> |
|-----------------------|---------------------|---------------------|-------------------|--------------------|---------------------|
| a. [kto:] ~ [kto:]    |                     |                     | −1                |                    | -0.0020             |
| ✓ b. [kto:] ~ [kuto:] |                     |                     |                   |                    | 0.0000              |
| c. [kto:] ~ [ku:to:]  |                     |                     |                   | −1                 | -0.0004             |

Table 2: Repair through epenthesis.

## 2.2 Alternation learning

### 2.2.1 The lexicon

A lexicon allows the model to keep track of what input forms correspond to what meaning (Apostolou 2007), and eventually acquire a paradigm over the lexicon. The CSJ-Core provides a word-level representation with Unicode characters of Japanese orthography, along with the phoneme and phone level transcriptions associated with the word. What this means for the model is that homophonous words can be distinguished from each other based on the orthographic representation of the words. For example, the words for ‘hard fight’ and ‘punctuation’ are both /kuto:/ underlyingly. The two words are orthographically different, however—<苦闘> ‘hard fight’ and <句読> ‘punctuation’—allowing the model to acquire them as separate 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 for the moment.

Furthermore, because the CSJ-Core explicitly notes reduction status in the surface level transcriptions, the model can keep track of phonetic variations. For example, suppose that the word ‘hard fight’ occurs five times in the training data, with three times involving a reduced vowel. The word ‘punctuation’, on the other hand occurs once in the input with a reduced vowel. Additionally, the word /kudo:/ ‘driving force’ occurs ten times only with an unreduced vowel. The model would create a lexical dictionary that looks something like the following:

| <i>Word</i> | <i>Gloss</i>    | <i>Underlying</i> | <i>Surface</i>              |
|-------------|-----------------|-------------------|-----------------------------|
| 苦闘          | ‘hard fight’    | /kuto:/ (x5)      | [k̚uto:] (x4), [kuto:] (x1) |
| 句読          | ‘punctuation’   | /kuto:/ (x1)      | [k̚uto:] (x1)               |
| 駆動          | ‘driving force’ | /kudo:/ (x10)     | [kudo:] (x10)               |

Table 3: Toy lexicon.

### 2.2.2 Conversion rule induction

With the lexicon in place, the model can now learn high vowel voicing alternations. The alternation learning mechanism simply keeps track of the environments in which an underlying high vowel surfaces as either unreduced or reduced and induces underlying-surface conversion rules. The observed probability of a surface 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.

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 3. Focusing on the reduction environment (i.e., initial /CVC/), if the model were inducing biphone conversion rules, the model would learn that /ku/ may surface as either [ku] with a probability of 0.687 (11 out of 16) or [k̚u] with a probability of 0.313 (5 out of

16). It also would learn that /ud/ has a probability of 1.000 (10 out of 10) to surface as [ud] while /ut/ may surface as [u̥t] or [ut] with probabilities of 0.667 and 0.333, respectively. The resulting conversion rule with their corresponding weights are presented in Table 4 below.

| <i>conversion rule</i>                       | <i>weight</i> |
|--|---------------|
| /ku/ $\rightleftharpoons$ [ku]               | 0.687         |
| /ku/ $\rightleftharpoons$ [k <sub>u̥</sub> ] | 0.313         |
| /ut/ $\rightleftharpoons$ [u̥t]              | 0.667         |
| /ut/ $\rightleftharpoons$ [ut]               | 0.333         |
| /ud/ $\rightleftharpoons$ [ud]               | 1.000         |

Table 4: Example of biphone conversion rules and weights.

For triphones, the process is the same, but because it is keeping track of larger chunks, there are fewer conversion rules. Again from the toy lexicon in Table 3, the model would learn that of the six times in which an underlying sequence /kut/ occurred, the probability of the surface form [k<sub>u̥</sub>t] is 0.833 (5 out of 6) whereas the probability of [kut] is 0.167 (1 out of 6). For the underlying sequence /kud/, however, there is just one surface form [kud] with the probability of 1.000 (10 out of 10). The model would therefore induce the following triphone conversion rules and weights:

| <i>conversion rule</i>                         | <i>weight</i> |
|--|---------------|
| /kud/ $\rightleftharpoons$ [kud]               | 1.000         |
| /kut/ $\rightleftharpoons$ [k <sub>u̥</sub> t] | 0.833         |
| /kut/ $\rightleftharpoons$ [kut]               | 0.167         |

Table 5: Example of triphone conversion rules and weights.

Because Japanese high vowel reduction requires access to both consonants flanking the target vowel, the model will utilize triphone conversion rules.

### 2.2.3 Conversion rules in action

Using the conversion rules from Table 5 but with the actual weights that the model learned, let us consider the example in Table 6 below. Given an underlying form /kuto:/ as input, none of the candidates are penalized by the highest weighted conversion rule /kud/  $\rightleftharpoons$  [kud] because there is

no /kud/ sequence in the input. Candidates (c-d) have a total weight of -1.000 for incurring one violation for both /kut/  $\rightleftharpoons$  [kut] and /kut/  $\rightleftharpoons$  [kʊt] since the candidates contain neither [kut] nor [kʊt] corresponding to the input sequence /kut/. Candidate (a) incurs one violation for violating /kut/  $\rightleftharpoons$  [kʊt] for a total weight of -0.986. Although candidate (b) incurs a violation of /kut/  $\rightleftharpoons$  [kut], the total weight of the candidate is the highest at -0.014, and thus is selected as the optimal candidate.

|                             | /kud/ $\rightleftharpoons$ [kud]<br>(1.000) | /kut/ $\rightleftharpoons$ [kʊt]<br>(0.986) | /kut/ $\rightleftharpoons$ [kut]<br>(0.014) | <i>total weight</i> |
|-----------------------------|---|---|---|---------------------|
| a. /kuto:/ $\sim$ [kuto:]   |   | -1  |   | -0.986              |
| ✓ b. /kuto:/ $\sim$ [kʊto:] |   |   | -1  | -0.014              |
| c. /kuto:/ $\sim$ [kato:]   |   | -1  | -1  | -1.000              |
| d. /kuto:/ $\sim$ [kudo:]   |   | -1  | -1  | -1.000              |

Table 6: Correct reduced surface form selected given underlying form.

If the input is /kudo:/, where the high vowel is in a voicing environment, the same conversion rules select the correct, non-reduced output candidate because all other candidates would violate the highest weighted /kud/  $\rightleftharpoons$  [kud], as shown in Table 7.

| /kudo:/      | /kud/ $\rightleftharpoons$ [kud]<br>(1.000) | /kut/ $\rightleftharpoons$ [kʊt]<br>(0.986) | /kut/ $\rightleftharpoons$ [kut]<br>(0.014) | <i>total weight</i> |
|--------------|---|---|---|---------------------|
| a. [kuto:]   | -1  |   |   | -1.000              |
| b. [kʊto:]   | -1  |   |   | -1.000              |
| c. [kato:]   | -1  |   |   | -1.000              |
| ✓ d. [kudo:] |   |   |   | 0.000               |

Table 7: Correct unreduced surface form selected given underlying form.

Where the alternation grammar would have trouble, however, is when the underlying form given contains a sequence that the model has never encountered during the training phase. One such case in our simulations was the word /jugʲo:/ ‘training’, shown in Table 8. The model had never encountered the triphone /jugʲ/ during training, and could not narrow the candidate set to an optimal candidate. Because the palatalized voiced velar stop /gʲ/ is phonemic in Japanese, the conversion rule /jug/  $\rightleftharpoons$  [jug] does not apply to any of the input-output pairs given as candidates.

|   | $/\text{ʃug}/ \rightleftharpoons [\text{ʃug}]$<br>(1.000) | $/\text{ʃuk}/ \rightleftharpoons [\text{ʃuk}]$<br>(1.000) | $/\text{ʃum}/ \rightleftharpoons [\text{ʃum}]$<br>(1.000) | <i>total weight</i> |
|---|---|---|---|---------------------|
| ! a. $/\text{ʃug}^{\text{ɪ}}\text{o:}/ \sim [\text{ʃug}^{\text{ɪ}}\text{o:}]$ |   |   |   | 0.000               |
| b. $/\text{ʃug}^{\text{ɪ}}\text{o:}/ \sim [\text{ʃug}^{\text{ɪ}}\text{o:}]$   |   |   |   | 0.000               |
| c. $/\text{ʃug}^{\text{ɪ}}\text{o:}/ \sim [\text{ʃeg}^{\text{ɪ}}\text{o:}]$   |   |   |   | 0.000               |
| d. $/\text{ʃug}^{\text{ɪ}}\text{o:}/ \sim [\text{ʃugo:}]$                     |   |   |   | 0.000               |

Table 8: No optimal output.

### 2.3 Combining phonotactic constraints and conversion rules

As illustrated below in Figure 2, the EVAL mechanism of the model is stratified such that the conversion rules evaluate the candidates generated by GEN first. 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.

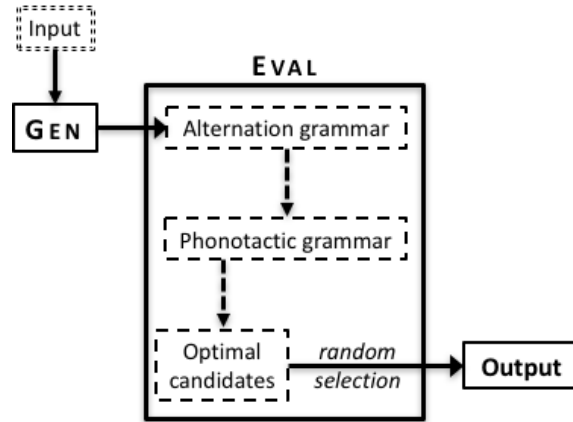



Figure 2: Stratified EVAL mechanism.

This stratified implementation of the alternation and phonotactic grammars is based on insights from Stratal Optimality Theory, where the phonological grammar is assumed to be separated into different-sized domains, and the grammar of a smaller domain applies before the larger domain (Kiparsky 2000). For the purposes of our model, the smaller domain grammar is the alternation grammar, since it is a lexicon-based, word-level grammar. Because phonotactic grammar is as-

sumed to be learned before the learner begins relying on a morphologically detailed lexicon (Tesar and Prince 2007), it is assumed to be part of a larger domain, namely the phrase-level.

The main benefit of stratification is that for inputs that require reduction, the conversion rules can eliminate non-reduced candidates before the phonotactic grammar can impose a CVCV preference to the output. This is illustrated in Table 9 below with the input /kuto:/ ‘hard fight’. Because the conversion rules apply first, the only candidate that gets passed on to the phonotactic grammar is the reduced candidate (b). Despite the fact that candidate (b) violates the FAITHFULNESS constraint IO-|ku|, it is still the winner because there simply are no other candidates.

| A. Word-level  |   |  |   |                     |
|--|---|--|---|---------------------|
|  | /kud/ $\rightleftharpoons$ [kud]<br>(1.000) | /kut/ $\rightleftharpoons$ [k <sub>u</sub> t]<br>(0.986) | /kut/ $\rightleftharpoons$ [kut]<br>(0.014) | <i>total weight</i> |
| a. /kuto:/ $\sim$ [kuto:]  |   | −1   |   | -0.986              |
|  b. /kuto:/ $\sim$ [k <sub>u</sub> to:] |   |  | −1  | -0.014              |
| c. /kuto:/ $\sim$ [kato:]  |   | −1   | −1  | -1.000              |
| d. /kuto:/ $\sim$ [kudo:]  |   | −1   | −1  | -1.000              |


| B. Phrase-level  |                     |                                  |                   |                     |
|--|---------------------|----------------------------------|-------------------|---------------------|
|  | IO- ku <br>(0.0120) | IO- k <sub>u</sub>  <br>(0.0060) | * ut <br>(0.0003) | <i>total weight</i> |
|  b. /kuto:/ $\sim$ [k <sub>u</sub> to:] | −1                  |                                  |                   | -0.0120             |

Table 9: Stratified grammar selects correct reduced output.

Additionally, if the conversion rules fail to eliminate candidates, as was the case with /jug<sup>j</sup>o:/ ‘training’, it is up to the phonotactic grammar to select the optimal candidate. This is shown in Table 10 below. Because the conversion rules fail to eliminate any candidate, all candidates get passed on for phonotactic evaluation. At the phrase level, because vowel length is phonemic in Japanese, the FAITHFULNESS constraint IO-|fu:| does not apply to any of the candidates because the underlying vowel is short. With the exception of candidate (a), all candidates violate IO-|ug<sup>j</sup>| for not keeping faithful to the underlying /ug<sup>j</sup>/ sequence. Candidate (c) incurs an additional violation of the MARKEDNESS constraint \*|fe|. Because candidate (a) has the highest total weight of 0.0, it is selected correctly as the optimal candidate.



| A. Word-level                   |   |   |   |                     |
|---------------------------------|---|---|---|---------------------|
|                                 | $/fug/ \rightleftharpoons [fug]$<br>(1.000) | $/fuk/ \rightleftharpoons [fuk]$<br>(1.000) | $/fum/ \rightleftharpoons [fum]$<br>(1.000) | <i>total weight</i> |
| a. $/fug^j o:/ \sim [fug^j o:]$ |   |   |   | 0.000               |
| b. $/fug^j o:/ \sim [fug^j o:]$ |   |   |   | 0.000               |
| c. $/fug^j o:/ \sim [feg^j o:]$ |   |   |   | 0.000               |
| d. $/fug^j o:/ \sim [fugo:]$    |   |   |   | 0.000               |

| B. Phrase-level                   |                        |                          |                       |                     |
|-----------------------------------|------------------------|--------------------------|-----------------------|---------------------|
|                                   | IO- $ fu $<br>(0.0013) | IO- $ ug^j $<br>(0.0001) | * $ fe $<br>(0.00001) | <i>total weight</i> |
| ✓ a. $/fug^j o:/ \sim [fug^j o:]$ |                        |                          |                       | 0.00000             |
| b. $/fug^j o:/ \sim [fug^j o:]$   |                        | -1                       |                       | -0.00010            |
| c. $/fug^j o:/ \sim [feg^j o:]$   |                        | -1                       | -1                    | -0.00011            |
| d. $/fug^j o:/ \sim [fugo:]$      |                        | -1                       |                       | -0.00010            |

Table 10: Stratified grammar selects correct non-reduced output.

### 3 Simulations

#### 3.1 Overview

There are three main goals of the simulations in this paper, first of which is to verify whether a preference for CVCV structure can be learned with constraints statistically induced from real spontaneous Japanese data. With numerous empirical evidence for a CVCV preference in Japanese listeners, 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.

Second, we explore whether alternation learning can be captured statistically by inducing simple conversion rules based on underlying-to-surface mappings observed in the lexicon. If the learning mechanism is successful, the results should show a context-based preference for reduced and unreduced vowels, unlike the phonotactic grammar of Japanese which singularly prefers unreduced vowels.

Lastly, by combining phonotactic and alternation learning mechanisms, the model allows us to test whether phonotactics is indeed helpful in alternation learning. If the results show that the combined, stratified 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). If the stratified model does not perform the best, however, it could mean that phonotactics is not helpful in alternation learning, but also that a stratified grammar is not the best way to capture the interaction of the two processes.

### **3.2 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). Unlike classic OT, however, where constraints are in strict domination, constraints are assumed to be weighted, as in serial Harmonic Grammar (Pater 2012). The learning mechanism as described in §1 above induces phonotactic constraints and constraint-like conversion rules that make up CON. In classic OT, the GEN process generates potential candidates for output. CON is the set of constraints that the model has acquired through the learning mechanisms described above in §1, and the EVAL mechanism takes the candidate set generated by GEN and arrives at the most optimal candidate or candidates based on the ordered set of constraints in CON.

### **3.3 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 our 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).

The model’s performance was compared to human speakers, evaluated for overall reduction rates and selection of the correct vowel in the output. Twenty-two monolingual Japanese speakers (12 women, 10 men) were recruited for the production experiment in Tokyo, Japan. The participants’ ages ranged from 18 to 24 years. The stimulus items used in the experiment were all lexical items with the reducible high vowels in the first syllable, controlled to be of medium frequency (20 to 100 occurrences, that 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).  $C_1$  types were divided into *low predictability*, after which /i, u/ can occur, and *high predictability* groups, where only one of the high vowels can occur. The low predictability group included /k, ʃ/ and the high predictability group included /tʃ, ɕ, φ, s/. The stimuli were divided into two groups: *reducing* where high vowel reduction is expected to occur and *non-reducing* where high vowel reduction should not occur.  $C_2$  was always /p, t, k/ for reducing tokens creating an environment where high vowels are flanked by two voiceless obstruents.  $C_2$  for non-reducing tokens were always /b, d, g/. There were 10 tokens per  $C_1V$  combination, for a total of 160 tokens (80 reducing and 80 non-reducing). Examples of the reducing stimuli are shown in Table 11 below.

| <i>stimulus type</i> | $C_1$ | $V$ | <i>example</i>            | <i>gloss</i> |
|----------------------|-------|-----|---------------------------|--------------|
| low predictability   | k     | i   | <u>kiki</u>               | ‘handedness’ |
|                      |       | u   | <u>kuki</u>               | ‘twig’       |
|                      | ʃ     | i   | <u>ʃiko:</u>              | ‘thought’    |
|                      |       | u   | <u>ʃuko:</u>              | ‘plan’       |
| high predictability  | tʃ    | i   | <u>tʃik<sup>h</sup>u:</u> | ‘earth’      |
|                      | s     | u   | <u>suku:</u>              | ‘rescue’     |
|                      | φ     | u   | <u>φuko:</u>              | ‘unhappy’    |
|                      | ɕ     | i   | <u>ɕite:</u>              | ‘denial’     |

Table 11: Example of reducing stimuli by  $C_1$  and vowel.

Each stimulus token was placed in the context of unique and meaningful carrier sentences of varying lengths, constructed so that no major phrasal boundaries immediately precede the word containing the target high vowel.

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

## 3.4 Results

### 3.4.1 Summary of production experiment

The reduction rates from the experiment are summarized in Table 12 below. For reducing tokens, reduction rates were essentially at ceiling with an overall reduction of 99.4%, while non-reducing tokens had an overall reduction rate of 10%. Among the non-reducing tokens, /tʃ/-initial and /s/-initial tokens had the highest reduction rates, which were both around 20%.

| C <sub>1</sub> | reducing | non-reducing |
|----------------|----------|--------------|
| /k/            | 0.979    | 0.055        |
| /ʃ/            | 0.986    | 0.080        |
| /tʃ/           | 1.000    | 0.191        |
| /ɸ/            | 1.000    | 0.042        |
| /s/            | 1.000    | 0.214        |
| /ç/            | 1.000    | 0.015        |
| <i>overall</i> | 0.994    | 0.100        |

Table 12: Reduction rate by token type from 22 Japanese participants.

### 3.4.2 Simulation 1: 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/ → [a, u]). This last category was included under the assumption that a complete change in vowel category should not happen given the general description of high vowel reduction in Japanese.

To compare each of the models' performance to that of the production experiment,  $d'$  values were calculated for both the experimental results and each of the models. The following metrics were used to evaluate model performance. Hit rate ( $H$ ) is the mean of the reduction rate in reducing tokens and the *unreduced* rate in non-reducing tokens since vowels should not reduce in these tokens. False alarm ( $F$ ) is the average of wrong vowel errors in reducing and non-reducing tokens. A model with a high rate of wrong vowel errors, therefore is penalised even if the reduction rates are otherwise high. Since wrong vowel errors are assumed to never occur in natural speech, the production experiment has a False Alarm rate of 0. Given the metrics below,  $d$ -prime of the experimental results is 5.916.

Hit Rate ( $H$ ):

$$H = \frac{P(\text{reduced}|\text{reducing}) + P(\text{unreduced}|\text{non-reducing})}{2} \quad (1)$$

False Alarm ( $F$ ):

$$F = \frac{P(\text{wrong vowel}|\text{reducing}) + P(\text{wrong vowel}|\text{non-reducing})}{2} \quad (2)$$

$d$ -prime ( $d'$ ):

$$d' = z(H) - z(F) \quad (3)$$

The results for the phonotactics only model are as shown in Table 13 below. With an overall reduction rate of 8.9% for reducing tokens and an unreduced rate of 98.5% for non-reducing tokens, the Hit Rate for the phonotactics only model is 0.537. Furthermore, with a wrong vowel error rate of 46.4% for reducing tokens and 1.3% for non-reducing tokens, the False Alarm rate is 0.289. This yields a  $d$ -prime value of 0.649.

| 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       |
| /j/            | 0.143    | 0.095     | 0.762       | 0.000        | 1.000     | 0.000       |
| /tʃ/           | 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       |
| <i>overall</i> | 0.089    | 0.447     | 0.464       | 0.002        | 0.985     | 0.013       |

Table 13: Phonotactics only: Mean probabilities from 22 test simulations.

The results confirm that the CVCV phonotactic preference in Japanese can be learned from statistically induced constraints. Compared to the 99.4% reduction rate for reducing tokens by Japanese speakers, the reducing tokens had a substantially lower reduction rate for the phonotactics-only model at a mere 8.9% overall. The phonotactic model also selects the wrong vowel as the optimal candidate 46.4% of the time. At 0.2%, the reduction rates for non-reducing tokens are also much lower than the 10% in the experimental results.

### 3.4.3 Simulation 2: Alternation only

The results for the alternation-only model are presented below in Table 14. With an overall reduction rate of 93.1% for reducing tokens and an unreduced rate of 48.8% for non-reducing tokens, the hit rate for the phonotactics only model is 0.710. Also, with a wrong vowel error rate of 6.5% for reducing tokens and 15.4% for non-reducing tokens, the false alarm rate is 0.110. Given these hit rate and false alarm rate, the  $d$ -prime value of the alternation only model is 1.780.

| 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       |
| /j/            | 1.000    | 0.000     | 0.000       | 0.459        | 0.412     | 0.129       |
| /tʃ/           | 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       |
| <i>overall</i> | 0.931    | 0.004     | 0.065       | 0.358        | 0.488     | 0.154       |

Table 14: Alternation only: Mean probabilities from 22 test simulations.

The results show that alternation learning from the CSJ was somewhat successful with some notable problems. Reduction rates 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%. Wrong vowel errors are also low at 6.5% and are limited to /k/ and /tʃ/ tokens specifically. However, the reduction rate for /tʃ/ 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 /tʃit/ contexts. The reason for the error in this context was that there were no word-initial /tʃit/ sequences in the training data. This meant that no conversion rule applied to any input with word-initial /tʃit/, 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<sub>1</sub> contexts had some wrong vowel error with the exception of /s/ tokens, resulting in an overall wrong vowel error rate of 15.4%, and 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/  $\rightleftharpoons$  [çid], leading the model to favor reduction in some non-reducing environments.

### 3.4.4 Simulation 3: Stratified model

The results of the stratified model are presented in Table 15 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 /tʃ/-initial reducing tokens, the stratified model makes zero wrong vowel errors. The elimination of wrong vowel errors also improves the reduced/unreduced rates in general as well. This is confirmed by the *d*-prime value for the stratified model which is 2.815 ( $H = 0.799$ ;  $F = 0.024$ ).

Even in the case of /tʃ/-initial reducing tokens where wrong vowel errors were not completely corrected, the error rates are still lowered slightly, bumping up 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 66.1% non-reduction rate bring it much closer to the experimental results.

| 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       |
| /j/            | 1.000    | 0.000     | 0.000       | 0.450        | 0.550     | 0.000       |
| /tʃ/           | 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       |
| <i>overall</i> | 0.936    | 0.016     | 0.048       | 0.339        | 0.661     | 0.000       |

Table 15: Proposed model: Mean probabilities from 22 test simulations.

## 4 Discussion and conclusion

The main goal of our model was to predict the reduction rates of high vowels in a production. As the summary of model performances in Table 16 shows, the performance of the combined model was closest to the production experiments.



| simulation            | hit rate          | false alarm       | <i>d</i> -prime   |
|-----------------------|-------------------|-------------------|-------------------|
| Production experiment | 0.947             | 0.000             | 5.916             |
| Stratified model      | $0.799 \pm 0.001$ | $0.024 \pm 0.002$ | $2.815 \pm 0.059$ |
| Alternation model     | $0.710 \pm 0.003$ | $0.110 \pm 0.004$ | $1.780 \pm 0.030$ |
| Phonotactic model     | $0.537 \pm 0.007$ | $0.289 \pm 0.009$ | $0.649 \pm 0.046$ |

Table 16: Hit rate, false alarm, and *d*-prime with 95% CI of all models.

Of the three models that were tested, the phonotactics-only model in Simulation 1 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. Based on the assumption that phonotactic learning happens before the acquisition of a lexicon (Hayes 2004; Tesar and Prince 2007), the difference in overall reduction rates between the phonotactic and alternations models leads to the prediction that reduction rates in Japanese children should increase over time. This is because the strong phonotactic preference for CVCV structures would lead younger children to reduce less until they become morphologically aware as their lexicon grows. Empirical work on the production of Japanese high vowel reduction in children are limited relative to the number of perception studies. However, a study by Imaizumi et al. (1999) which compared the high vowel reduction rates of Japanese children from different dialectal regions provides some support that the predicted, gradual increase in reduction rates is indeed what happens in children from the Tokyo area.

Although the alternation model was more successful than the phonotactic model in overall hit rate, it still suffered from wrong vowel errors in the non-reducing tokens. 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. The stratified 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 Tesar and Prince (2007); Pater and Tessier (2003).

A feature of STAGE that was not implemented in the current iteration of our model was a generalization mechanism. The generalization mechanism of STAGE utilizes a single feature abstraction mechanism that combines two or more similar constraints. For example, since IO-|gu| and IO-|bu| differ only by place, a more general constraint IO-| $x \in \{g, b\}; y \in \{u\}$ |, 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 component 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.

The next obvious step is to see how well the combined phonotactic and alternation grammars predicts the repair of consonant clusters during perception. A recent work by (Durvasula and Kahng 2015) investigated illusory vowel epenthesis in Korean speakers, and have argued that knowledge of phonological alternation processes in addition to phonotactics could better explain why different vowels are perceived in different contexts. Because the conversion rules that make up the alternation grammar in our model are bidirectional, testing the model for perceptual accuracy could simply be a matter of giving the model surface forms rather than underlying forms as input. It remains to be seen, however, just how flexible the model is.

## References

- 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.
- 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.
- Bernstein Ratner, N. 1984. Patterns of vowel modification in mother–child speech. *Journal of Child Language* 11:557–78.
- 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.
- Calamaro, Shira, and Gaja Jarosz. 2015. Learning general phonological rules from distributional information: A computational model. *Cognitive Science* 39:647–666.
- Coetzee, Andries W., and Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *NLLT* 26:289–337.
- Daland, R., and J. Pierrehumbert. 2011. Learning diphone-based segmentation. *Cognitive Science* 35:119–155.
- Davidson, Lisa, and Maureen Stone. 2003. Epenthesis versus gestural mistiming in consonant cluster production: an ultrasound study. In *Proceedings of WCCFL 22*.
- 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 Kakehi, Yuki Hirose, Christophe Pallier, and Jacques Mehler. 1999. Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception & Performance* 25:1568–1578.
- Dupoux, Emmanuel, Erika Parlato, Sónia Frota, Yuki Hirose, and Sharon Peperkamp. 2011. Where do illusory vowels come from? *Journal of Memory and Language* 64:199–210.
- Durvasula, K., and J. Kahng. 2015. Illusory vowels in perceptual epenthesis: The role of phonological alternations. *Phonology* 32:385–416.
- 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.
- 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.
- Han, Mieko S. 1962. *Japanese Phonology: An Analysis Based Upon Sound Spectrograms*. Kenkyusha, Tokyo.
- Han, Mieko S. 1994. Acoustic manifestations of mora timing in Japanese. *JASA* 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 Moravcsik, 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, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *LI* 39:379–440.
- Hirose, Hajime. 1971. The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese. *Phonetica* 23:156–170.
- 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.
- Jarosz, Gaja. 2006. Rich lexicons and restrictive grammars – Maximum likelihood learning in Optimality Theory. Doctoral Dissertation, Johns Hopkins University.
- 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.
- Kindaichi, Haruhiko. 1995. *Shin Meikai Nihongo Akusento Jiten [Japanese Accent Dictionary]*. Sanseido.
- Kiparsky, Paul. 2000. Opacity and cyclicity. *Special issue of The Linguistic Review* 17:351–65.
- 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.
- Martin, Andrew, Akira Utsugi, and Reiko Mazuka. 2014. The multidimensional nature of hyper-speech: Evidence from Japanese vowel devoicing. *Cognition* 132:216–228.
- Mattys, S. L., and P. W. Jusczyk. 2001. Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78:91–121.
- Maye, J., J. F. Werker, and L. Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82:B101–B111.
- 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.
- NHK. 1985. *Nihongo Hatsuon Akusento Jiten [Japanese Pronunciation Accent Dictionary]*. Tokyo: Nihon Hoso Kyokai [Japan Broadcasing Corporation].
- Ogasawara, Naomi. 2013. Lexical representation of Japanese high vowel devoicing. *Language and Speech* 56:5–22.
- 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 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*.
- 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. 2004. Learning phonotactic distributions. In *Constraints in phonological acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, 245–291. Cambridge, UK: Cambridge University Press.
- Saffran, J. R., R. N. Aslin, and E. L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274:1926–1928.
- 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/>.
- Tesar, Bruce, and Alan Prince. 2007. Using phonotactics to learn phonological alternations. In *CLS 39*, volume 2, 209–237.
- Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: The MIT Press.

- 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.
- 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.
- Varden, J. Kevin. 1998. On high vowel devoicing in standard modern Japanese. Doctoral Dissertation, University of Washington.
- Varden, J. Kevin. 2010. On vowel devoicing in Japanese. *The MGU Journal of Liberal Arts Studies KARUCHURU* 4:223–235.
- 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.
- White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130:96–115.
- Yoshioka, H. 1981. Laryngeal adjustment in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica* 38:236–351.