Phonotactics and Alternations in the Acquisition of Japanese High Vowel Reduction

James Whang and Frans Adriaans

1. Introduction
1.1. Japanese high vowel reduction

In standard modern Japanese, high vowels /i, u/ undergo gestural reduction when flanked by two voiceless obstruents (e.g., /tʃikʲuː/ → [tʃi̥kʲuː]). High vowel reduction is a nearly obligatory and highly productive process in standard modern Japanese (Fujimoto 2015). While the phenomenon is more commonly called high vowel *devoicing*, the term reduction is used based on empirical evidence suggesting that high vowel reduction can range from simple loss of voicing (Han 1994) to complete deletion of the vowel, resulting in clusters on the surface (Vance 2008).

While high vowels are commonly reduced in Japanese, the language is also well-known for having a strong phonotactic preference for CV structure that prohibits consonant clusters (Kubozono 2015). The CV preference is evident in loanwords, where consonant clusters and codas are repaired through high vowel epenthesis (e.g., [kɹiːm] → [ku̲.riː.mu̲] 'cream'; Smith 2006), and also in psycholinguistic studies, where Japanese listeners report hearing a high vowel between consonant clusters (e.g., [eb.zo] → /e.bu̲.zo/; Dupoux et al. 1999).

The occurrence of consonant clusters as a result of vowel reduction leads to a puzzling situation for the acquisition of Japanese. Learners need to acquire strict CV phonotactics, but the input to the learner contains a substantial amount of consonant clusters. The current paper uses computational modeling to investigate how these two seemingly contradictory aspects of Japanese phonology, high vowel reduction and CV phonotactics, might be acquired. We present a model which combines phonotactic learning and alternation learning mechanisms. The model is trained on a subset of the Corpus of Spontaneous Japanese (Maekawa 2003). A preliminary evaluation of the model is presented, using production data on high vowel reduction.

1.2. Acquisition of high vowel reduction in Japanese children

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 (0 year; 6 months), but this sensitivity is noticeably

*James Whang, New York University, james.whang@nyu.edu. Frans Adriaans, Utrecht University, f.w.adriaans@uu.nl.

 Proceedings of the 41st annual Boston University Conference on Language Development, ed. Maria LaMendola and Jennifer Scott, 730-743. Somerville, MA: Cascadilla Press.

diminished by the age of 1;0, and even more so by the age of 1;6. Japanese infants, therefore, have learned to ignore the difference between C_1C_2 and C_1VC_2 sequences by the age of 1;0. There are very few studies that have looked at the *production* of high vowels by 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 all Japanese children show an initial bias towards CV structure with low reduction rates during production before reaching adult-like levels around the age of five. Japanese children thus learn to ignore the difference between C_1C_2 and C_1VC_2 sequences with a bias towards CV structure quite early in development, and mastery of producing reduced high vowels is acquired much later.

To what extent do Japanese children get evidence for high vowel reduction or CV phonotactics in their input? Studies on infant-directed speech often report a number of differences between infant-directed speech (IDS) and adult-directed speech (ADS), such as an expanded vowel space and F0 range (e.g., Kuhl et al. 1997). With regards to the acquisition of high vowel reduction, one possibility is that Japanese IDS presents the learner with canonical, unreduced forms to facilitate the learning of CV phonotactics. This would, however, obscure the high vowel reduction process that is an integral part of adult speech. Another possibility is that adults present infants with adult-like vowel reductions, which would obscure the CV phonotactics. Fais et al. (2010) report that while some differences in prosodic cues are present between IDS and ADS in Japanese, the rates of high vowel reduction in both IDS and ADS were around 85% for lexical words and around 20% for nonce words. Another study by Martin et al. (2014) reports sightly lower rates of high vowel reduction in IDS relative to ADS (77% and 90%, respectively). Both studies show that Japanese caretakers provide infants with a substantial amount of reduced vowels, and as a consequence, Japanese infants are likely to encounter consonant clusters in their input.

1.3. Training data for the computational model

The subset of the Corpus of Spontaneous Japanese used for the current study is the CSJ-Core (Maekawa 2003). CSJ-Core contains approximately 45 hours of speech recorded from 200 speakers (500,000 words). The recordings were segmented and labeled both phonemically and sub-phonemically. The most relevant annotations for the model are the word level, phoneme level, and phone level transcriptions.

For each word level transcription, which is represented in Unicode characters of Japanese orthography, there are phoneme level and phone level transcriptions. Phoneme level transcriptions approximate underlying forms and are largely based on the kana syllabary of Japanese. Phone level transcriptions approximate surface forms and indicate phonetic variations including vowel reduction and consonant allophony. For example, the word ‘fox’ is represented as <キツネ>, <kicune>, and <kjIcune> at the word, phoneme, and phone levels, respectively. In the transcriptions, <kj> is a "phonetically palatalized voiceless velar stop", <c> is an alveolar affricate, and uppercase vowels are devoiced vowels. In IPA, the phoneme level transcription would correspond to /kit͡sune/ and the phone level to [k̟i̥t͡sune].

In phone level transcriptions, reduced vowels were always transcribed as devoiced and never as deleted. However, a number of studies have proposed that high vowels are likely to delete in high-predictability contexts where only one of two high vowels are phonotactically legal (Varden 2010; Whang 2014). To create consonant clusters that result from reduction, vowels transcribed as devoiced in the corpus were deleted using the following probabilities:

- Baseline deletion probability by vowel height (Base):
  - Short high vowels = 0.15
  - Other vowels = 0.05
- Additional probabilities by contexts
  - Non-reducing = Base − 0.05
  - Low-predictability reducing = Base + 0.15
  - High-predictability reducing = Base + 0.65

Deletion probabilities were set arbitrarily due to a lack of direct articulatory data with actual deletion rates, so they should merely be taken as a proof of concept. They were used to introduce a reasonable number of consonant clusters in the input for the model, so that we could examine potential effects on learnability. The deletion probabilities above only applied to vowels already transcribed as devoiced in the corpus. For example, if the word /kita/ 'north' was transcribed as [k̟ita] with no devoiced vowel, it was left unchanged. Conversely, if the same word was transcribed as [k̟i̥ta], the devoiced [i̥] would have a 30% probability of being deleted (15% (short high vowel baseline) + 15% (low-predictability context addition)). Similarly, if the word /suki/ 'to like' was transcribed as [su̥k̟i], the devoiced [u̥] would have an 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 [do̥ko] 'where' would never delete.

2. The model

2.1. Phonotactics and alternations

The proposed model combines phonotactic learning and lexical alternation learning. The model incorporates a simple phonotactic learning mechanism that tracks co-occurrence probabilities of segments in the input, and induces biphone constraints based on statistical patterns in the input (e.g., Adriaans and Kager 2010). The model also includes an alternation learning component. Studies have consistently shown that high vowel reduction rates are generally above 90%

between two voiceless obstruents, while it is well below 10% elsewhere (Fujimoto 2015). Reduced and non-reduced high vowels are thus in near complementary distribution with each other, and we therefore view high vowel reduction as an alternation between reduced and unreduced high vowels. By combining the two learning mechanisms, we aim to explore whether the seemingly contradictory preferences for CV structure and reduction of high vowels in Japanese can be learned from the same input data.

2.2. Phonotactic learning

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. Instead of inducing contiguity constraints, the current model induces what will be called CORRESPONDENCE constraints (hereafter COR) for overrepresented biphones (e.g., COR-*ku* assigns a violation for every instance of *ku* in the input that does not correspond to *ku* in the output). Instead of the strict domination of constraints used in Adriaans and Kager (2010), 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.

Table 1 below illustrates how the input /suki/ is evaluated by the phonotactic constraints learned by the model. Candidates (b-c) each incur a violation of COR-*su* for not remaining faithful to the input sequence /su/. Candidate (c) incurs an additional violation of **sk̟*. Although the faithful candidate (a) violates **uk̟*, it is selected as the winner with the highest total weight of -2.84e-04.

Table 1: No reduction, showing CV bias.

		/suki/	COR-*su* (9.04e-04)	**uk̟* (2.84e-04)	**sk̟* (2.35e-04)	*total weight*
✓	a.	[suk̟i]		−1		-2.84e-04
	b.	[su̥k̟i]	−1			-2.17e-03
	c.	[sk̟i]	−1		−1	-1.14e-03

What does this phonotactic component learn when trained on the CSJ corpus? An inspection of the model’s output shows that out of a total of 1,097 constraints, only three COR constraints involved consonant clusters, namely *ɸk, ɸts, tsk*. In comparison, there were $\sim$160 markedness constraints prohibiting all other heterorganic clusters. The model thus induces a bias against consonant clusters. This is to be expected from the distribution of consonant clusters in Japanese. 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. For example, in a corpus of 500K words, [sk] occurs less than 500 times. In contrast, [su] occurs $\sim$16,000 times and [uk] $\sim$2,700 times.

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 low O/E ratios for clusters. The phonotactic learner, therefore, learns a bias against consonant clusters and towards CV structure. It should be noted that the model in its current form does not create feature-based generalizations. However, given the types of specific constraints that are induced, the inclusion of a feature-based generalization mechanism such as Single Feature Abstraction (Adriaans and Kager 2010) would lead to more general constraints preferring CV, and dispreferring consonant clusters (*CC).

2.3. Alternation learning

The alternation learner is given access to a lexicon, which allows the model to keep track of what input forms correspond to what meaning (Apoussidou 2007), and eventually acquire a paradigm over the lexicon. A lexicon was created by keeping track of orthographic forms in the corpus, along with their corresponding phonemic and phonetic transcriptions. For example, the words '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. Additionally, because the alternation learner has access to both the phonemic and phonetic transcriptions, the model can keep track of one or more surface forms that correspond to one underlying form. This lexicon building process should not be interpreted as a model of lexical acquisition, but rather as a way to explore how alternations might be learned from a given lexicon.

Because the phonetic transcriptions in the corpus were modified to delete a large percentage of the vowels that were originally transcribed as devoiced in the corpus (30%$\sim$80% for high vowels), the model can keep track of surface variations between unreduced, devoiced, and deleted vowels. For example, shown below in Table 2 is a toy lexicon built from all occurrences of the word 'to like', 'an opening', and 'after/over' in the CSJ-Core corpus.

With a lexicon in place, the alternation learning mechanism simply keeps track of environments in which an underlying vowel surfaces as either unreduced or reduced and induces underlying-surface conversion rules. In order to be able to combine the learned rules with phonotactic constraints in a single model, the rules are coded as weighted constraints, where a violation is assigned for every instance of an input-to-output conversion that does not match the conversion rule.

Table 2: Toy lexicon.

Word	*Gloss*	*Underlying* (freq.)	*Surface* (freq.)	
好き	'to like'	/suki/ (140)	[sk̟i] (110), [sk̟i̥] (2), [sk̟] (2),	[su̥k̟i] (24), [su̥k̟i̥] (1), [su̥k̟] (1)
隙	'an opening'	/suki/ (1)	[sk̟i] (1)	
過ぎ	'after/over'	/sugi/ (44)	[sug̟i] (42),	[sg̟i] (2)

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, the weights of which add up to 1. Table 3 shows an example of triphone conversion rules that the model would learn from the toy lexicon, focusing on initial /CVC/ sequences. The model would learn that the underlying sequence /suk/ occurred 141 times, which surfaced as [su̥k̟] 26 times (0.184 = 26 ÷ 141) and as [sk̟] 115 times (0.816 = 115 ÷ 141).

Table 3: Example of triphone conversion rules and weights.

conversion rule	*weight*
/suk/ ⇋ [su̥k̟]	0.184
/suk/ ⇋ [sk̟]	0.816
/sug/ ⇋ [sug̟]	0.955
/sug/ ⇋ [sg̟]	0.045

The sequence length that the learner keeps track of may vary depending on the language being acquired. Pilot simulations were run to test the effectiveness of both triphone and biphone conversion rules, but biphone rules resulted in poorer performance. The advantage of triphone rules is expected for modeling Japanese high vowel reduction, since the processes requires access to both consonants flanking the target vowel. The simulations presented in this paper, therefore, use triphone conversion rules.

Using the conversion rules from Table 3 but with actual weights that the model learned from the corpus, consider the example in Table 4 below. The input is /suki/. The faithful candidate (a) violates /suk/ ⇋ [sk̟] and /suk/ ⇋ [su̥k̟], since the candidate contains neither [sk̟] nor [su̥k̟] that corresponds to the /suk/ in the input. Candidate (a) also violates /uki/ ⇋ [k̟i] for retaining the /u/ vowel. The devoiced candidate (b) violates /suk/ ⇋ [sk̟] and uki/ ⇋ [k̟i]. Candidate (d) violates all of the rules. The deleted candidate (c) is chosen as the winner despite violating /suk/ ⇋ [su̥k̟], with the highest total weight of -0.047. Note that by contrast, the phonotactic learner chose suk̟i] as the output as was shown in Table 1.

Table 4: Correct deleted form selected.

		/suki/	/suk/ ⇋ [sk̟] (0.242)	/suk/ ⇋ [su̥k̟] (0.047)	/uki/ ⇋ [k̟i] (0.355)	*total weight*
	a.	[suk̟i]	−1	−1	−1	-0.644
	b.	[su̥k̟i]	−1		−1	-0.597
✓	c.	[sk̟i]		−1		-0.047
	d.	[sug̟i]	−1	−1	−1	-0.644

2.4. Combining phonotactic constraints and alternations

Like most grammars based on an Optimality Theoretic framework, the model is assumed to consist of EVAL, CON, and GEN mechanisms (Prince and Smolensky 1993/2004). The EVAL mechanism of the model stratifies phonotactic constraints and conversion rules such that given an input, output candidates are evaluated first by conversion rules as a first pass filter, then by phonotactic constraints to further narrow down the choice of candidate if necessary. If more than one candidate remains after both evaluations, one is chosen at random as the final output. The model thus relies primarily on lexical processes over phonotactics (e.g., Shademan 2006; Vitevitch and Luce 1999). The model was set up this way so that for inputs that require reduction, the conversion rules can eliminate non-reduced candidates before the phonotactic grammar can impose a CV preference on the output. This is illustrated in Table 5 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. Although not shown here, when conversion rules return more than one output candidate, the phonotactic grammar helps narrow down the choice of output candidate further.

Table 5: Two-tier grammar selects correct reduced output.

A. Lexical level

		/suki/	/suk/ ⇋ [sk̟] (0.242)	/suk/ ⇋ [su̥k̟] (0.047)	/uki/ ⇋ [k̟i] (0.355)	*total weight*
	a.	[suk̟i]	−1	−1	−1	-0.644
	b.	[su̥k̟i]	−1		−1	-0.597
✓	c.	[sk̟i]		−1		-0.047
	d.	[sug̟i]	−1	−1	−1	-0.644

B. Phonotactic level

		/suki/	COR-*su* (9.04e-04)	**uk̟* (2.84e-04)	**sk̟* (2.35e-04)	*total weight*
✓	c.	[sk̟i]	−1		−1	-1.14e-03

As can be seen in the example, GEN was limited to manipulating the first vowel of the input (e.g., /suki/ → [ski, su̥ki, saki, suki]) to simplify the task and evaluation of the model.

3. Simulations
3.1. Methodology

In what follows, we present a preliminary evaluation of the model, using experimental data obtained in a production experiment (Whang, under revision). Twenty-two monolingual Japanese speakers (12 women, 10 men) were recruited in Tokyo, Japan for the experiment. The participants' ages ranged from 18 to 24 years. The stimuli were all lexical items with reducible high vowels in the first syllable. The stimuli were 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). Stimuli were divided into *low predictability* and *high predictability* contexts. The consonant preceding the target vowel (C_1) in *low predictability* stimuli were /k, ʃ/, after which both /i, u/ can occur. In *high predictability* stimuli, C_1 was /ʧ, ç, ɸ, s/, after which only one of the two high vowels can occur. The stimuli were further divided into *reducing* and *non-reducing* tokens. The consonant following the target vowel (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 was always /b, d, g/, creating a non-reducing environment. Each stimulus token was placed in unique and meaningful carrier sentences of varying lengths, constructed so that no major phrasal boundaries immediately preceded the word containing the target high vowel. There were 10 tokens per C_1V combination, for a total of 160 tokens (80 reducing and 80 non-reducing). Examples of reducing stimuli are shown in Table 6 below.

Table 6: Example of reducing stimuli by C_1 and vowel.

stimulus type	*C_1*	*V*	*example*	*gloss*
low predictability	k	i	kita	'north'
		u	kuki	'twig'
	ʃ	i	ʃika	'deer'
		u	ʃuku	'blessing'
high predictability	ʧ	i	ʧika	'underground'
	ç	i	çiteː	'denial'
	ɸ	u	ɸukoː	'unhappy'
	s	u	sukuː	'rescue'

The stimuli from the experiment were used as test data for the model. The model's EVAL mechanism sometimes returned more than one output candidate as equally optimal. Simulations, therefore, were run 22 times, same as the number of participants in the experiment, to capture the resulting variance in performance. Simulation results reported below are the means from 22 test simulations.

To compare the model's performance to the production experiment, d-prime values were calculated for both the experiment and simulations using the function *dprime.mAFC(Pc, m)* in the *psyphy* package in R (Knoblauch 2014), where *Pc* is the proportion of correct responses and *m* is the number of answer choices. *Pc* is the sum of hits and correct rejections divided by the total number of responses (Knoblauch and Maloney 2012). Since the number of reducing and non-reducing tokens were equal, the calculation was simplified as shown in the equation below:

Proportion correct (Pc):

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

Devoiced and deleted outputs were collapsed as reduced because there is no contrast between devoiced and deleted tokens in Japanese (e.g., [ɸu̥ku] = [ɸku] 'clothes'). The number of answer choices (i.e., *m*) was set at three: reduced, unreduced, and wrong vowel. Although the actual number of possible candidates generated by the model's GEN was 11 (5 voiced vowels, 5 devoiced vowels, 1 vowel deletion), 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. If the vowel in the output is a different vowel from the input, the vowel's identity and reduction status do not matter as it is simply wrong. Also, using a low *m* value returns more conservative d-prime values.

3.2. Production experiment results

Experimental results are summarized in Table 7 below. Reduction rates were 99.4% in reducing environments and 10% in non-reducing environments, the latter being driven mostly by /ʧ/-initial and /s/-initial tokens. This means that the hit and correct rejection rates 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.

Table 7: Reduction rate by token type from 22 Japanese participants.

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

3.3. Model simulation results

The results of the proposed model that combines phonotactic and alternation learning are presented in Table 8 below. In the *reducing* group, the *hit* column shows hit rates, where target high vowels in the input were devoiced or deleted in the output (e.g., /suk/ → [su̥k, sk]). In the *non-reducing* group, the *rejection* column shows correct rejection rates, where the target high vowel in the input was unchanged in the output (e.g., /sug/ → [sug]). 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., /suk/ → [sḁk, sek]).

The overall hit rate and the correct rejection rate of the model were 93.6% and 66.1%, respectively. This gives a *Pc* of 0.799 and a d-prime value of 1.648. Compared to the hit and rejection rates from the experiment (99.4% and 90%, respectively), the model underperforms, particularly in non-reducing tokens.

Table 8: Mean probabilities from 22 test simulations.

C1	reducing		non-reducing	
	hit	wrong vowel	rejection	wrong vowel
/k/	0.950	0.000	0.750	0.000
/ʃ/	1.000	0.000	0.550	0.000
/ʧ/	0.664	0.291	0.444	0.000
/ɸ/	1.000	0.000	1.000	0.000
/s/	1.000	0.000	1.000	0.000
/ç/	1.000	0.000	0.222	0.000
overall	0.936	0.048	0.661	0.000

Additional simulations were run with phonotactic constraints and conversion rules separately, to investigate their respective contributions to the results above and whether combining the two mechanisms provides a performance increase. Only the overall rates are shown below for the sake of space.

Table 9: Overall hit, rejection, and wrong vowel rates for all models.

	reducing		non-reducing	
	hit	wrong vowel	rejection	wrong vowel
Combined	0.936	0.048	0.661	0.000
Phonotactics	0.089	0.464	0.985	0.013
Alternation	0.931	0.065	0.488	0.154

First, compared to the combined model, the phonotactics-only model results show that there is a preference for unreduced CV structure in both reducing and non-reducing environments. In non-reducing environments, the model's performance is near-ceiling with a correct rejection rate of 98.5%. Although a similar CV bias is evident in reducing contexts with a low hit rate of 8.9%, wrong vowel responses are high at 46.4%. A survey of the phonotactic model's output revealed that phonotactic evaluation often failed to eliminate any vowel-ful candidate and chose an output candidate at random. This failure stemmed mainly from two factors. First, the bias against reduction was driven by COR-VC constraints. COR-*Vk* and COR-*Vt* constraints in particular were highly overrepresented in the input due to a high number of CVCV Sino-Japanese roots, where the second consonant is always /k, t/ (Ito and Mester 2015). These constraints kept the phonotactic model from reducing vowels in these contexts. Second, of the ∼1,400 biphones in the input, no constraint was induced for ∼300 biphones due to O/E ratios that fell between the two thresholds of 0.5 and 2.0. Without markedness constraints against certain CV and VC sequences, the model could not eliminate wrong vowel candidates.

Second, the alternation-only model shows that high vowel reduction was learned somewhat successfully through conversion rules. There are some notable problems, however, that are not present in the combined model. Although the overall hit rates are similar between the combined and alternation-only models, the latter also had more wrong vowel errors. A survey of the model's output revealed that the alternation-only model had wrong vowel errors in /k/- and /ʧ/-initial reducing tokens and all non-reducing contexts except for /s/-initial tokens. The combined model in comparison had wrong vowel errors only in /ʧ/-initial reducing tokens. The combined model's wrong vowel errors were limited to test words with an initial /ʧit/ sequence, because there were no words that began with /ʧit/ in the input.

4. Discussion and conclusion

Summarized below in Table 10 are the hit rate, rejection rate, and d-prime values that correspond to the production experiment and each simulation. Confidence intervals are also shown for the simulations. Output candidates were chosen at random when the model failed to narrow down the choice to a single candidate, 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.

Table 10: Hit rate, correct rejection rate, and d' with 95% CI of all models.

simulation	hit	rejection	*d*-prime
Production experiment	0.994	0.900	2.672
Combined 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

Of the three models that were tested, the phonotactics-only model showed the strongest CV preference across all contexts, confirming that statistically induced phonotactic constraints can indeed capture the strong CV preference in Japanese. The combined model's simulation results additionally confirmed that allowing the alternation rules to evaluate output candidates before the phonotactic constraints can overcome the CV preference and produce reduced outputs. 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 CV structure 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. This issue was largely remedied in the combined model. However, the failure of the combined model to handle novel sequences suggests that a generalization mechanism is necessary, such as the one in StaGe (Adriaans and Kager 2010). The generalization mechanism of StaGe utilizes a single feature abstraction mechanism that combines two or more similar 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. For example, /ʧiʧ/ $\leftrightharpoons$ [ʧi̥ʧ] and /ʧits/ $\leftrightharpoons$ [ʧi̥ts] differ only by the place of the final consonant, so a more general rule /x∈{ʧ};y∈{i};z∈{ʧ,ts}/ $\leftrightharpoons$ [x∈{ʧ};y∈{i̥};z∈{ʧ,ts}] can be induced, which means, "For an underlying sequence /ʧ/ followed by /i/ followed by /ʧ/ or /ts/, have a surface sequence [ʧ] followed by [i̥] followed by [ʧ] or [ts]." Ultimately a general rule that reduces high vowels between two voiceless obstruents can be induced this way.

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.

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

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.

Durvasula, Karthik, and Jimin 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. 1994. Acoustic manifestations of mora timing in Japanese. *Acoustical Society of America* 96:73–82.

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.

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, and Armin Mester. 2015. Sino-japanese phonology. In *Handbook of Japanese Phonetics and Phonology*, ed. Haruo Kubozono, chapter 7. Mouton de Gruyter.

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.

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.

Kubozono, Haruo. 2015. Loanword phonology. In *Handbook of Japanese Phonetics and Phonology*, ed. Haruo Kubozono, chapter 8, 313–362. Mouton de Gruyter.

Kuhl, Patricia K., Jean E. Andruski, Inna A. Chistovich, Ludmilla A. Chistovich, Elena V. Kozhevnikova, Victoria L. Ryskina, Elvira I. Stolyarova, Ulla Sundberg, and Fransisco Lacerda. 1997. Cross-language analysis of phonetic units in language addressed to infants. *Science* 277:684–686.

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.

Maekawa, Kikuo. 2003. Corpus of Spontaneous Japanese: Its design and evaluation. *Proceedings of the ISCA & IEEE workshop on spontaneous speech processing and recognition (SSPR)* .

Martin, Andrew, Akira Utsugi, and Reiko Mazuka. 2014. The multidimensional nature of hyperspeech: Evidence from Japanese vowel devoicing. *Cognition* 132:216–228.

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.

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.

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.

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.

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.

Tesar, Bruce, and Alan Prince. 2007. Using phonotactics to learn phonological alternations. In *CLS 39*, volume 2, 209–237.

Vance, Timothy J. 2008. *The sounds of Japanese*. New York: Cambridge University Press.

Varden, J. Kevin. 2010. Acoustic correlates of devoiced Japanese vowels: velar context. *The Journal of English and American Literature and Linguistics* 125:35–49.

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.

Whang, James. 2014. Acoustic effects of predictability and gender on Japanese high vowel reduction. Poster presentation at the 14th Conference of Laboratory Phonology.

Whang, James. under revision. Recoverability-driven gestural coordination: Acoustic evidence from Japanese high vowel reduction.

Proceedings of the 41st annual Boston University Conference on Language Development

edited by Maria LaMendola
and Jennifer Scott

Cascadilla Press Somerville, MA 2017

Copyright information

ISSN 1080-692X
ISBN 978-1-57473-076-0 (2 volume set, paperback)
ISBN 978-1-57473-176-7 (2 volume set, library binding)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com