Rory Turnbull*

Patterns of probabilistic segment deletion/ reduction in English and Japanese

https://doi.org/10.1515/lingvan-2017-0033
Received May 2, 2017; accepted July 5, 2018

Abstract: Probabilistic phonetic reduction is widely attested in a variety of languages, acoustic domains, and interpretations of predictability. Less well-studied is the categorical effect of probabilistic segment deletion, which in principle is subject to similar pressures. This paper presents the results of an exploratory study into patterns of segment deletion in corpora of spontaneous speech in English and Japanese. Analysis at the word level reveals that words with more phonemes and higher-frequency words tend to have more of their segments deleted. Analysis at the phoneme level reveals that high-probability phonemes are more likely to be deleted than low-probability phonemes. For Japanese only, this analysis also shows effects of word length, frequency, and neighborhood density on deletion probability. Taken together, these results suggest that several large-scale patterns of probabilistic segment deletion mirror the processes of phonetic reduction and apply to both languages. Some patterns, though, appear to be language-specific, and it is not clear to what extent languages can and do differ in this regard. These findings are discussed in terms of our understanding of the universality of proposed predictability effects, and in terms of probabilistic reduction more broadly.

Keywords: English; Japanese; segment deletion; predictability; cross-linguistic studies.

1 Introduction

Lexical predictability factors such as word frequency (Aylett and Turk 2004; Gahl 2008) and phonological neighborhood density (Munson and Solomon 2004; Wright 2004) are known to influence phonetic reduction, with more predictable words tending to be phonetically reduced relative to less predictable words. The results of phonetic reduction processes are often gradient differences in the magnitude of some acoustic feature, on the scale of a few Hertz or milliseconds. However, reduction processes can also have (near) categorical effects, such as wholesale deletion of phonemes. Johnson (2004) identified a number of such "massive reductions" in a subset of the Buckeye Corpus of Spontaneous Speech, where words had several phonemes deleted in production, such as [pɚɾĩ] for *apparently*, [hlɛɹɛs] for *hilarious*, and [ptɪkɚ] for *particular*. Johnson (2004) observed greater reduction of (segmentally) longer words than shorter words, and higher rates of deletion in function words relative to content words.

Although it may seem reasonable to consider segment deletion to be an instance of the extreme end of a spectrum of phonetic reduction, phonetic reduction and segment deletion are conceptually orthogonal. For example, it is possible for a word with a deleted segment to have the same or longer duration than a word with no deleted segments; and it is possible for a word with no deleted segments to undergo temporal reduction. These distinctions are not merely curiosities of production, but are potentially relevant for perception too. Ernestus and Baayen (2007) presented evidence suggesting that, while segment deletion delays word recognition, shorter durations may speed up word comprehension. Therefore, while it may be tempting to expect

***Corresponding author: Rory Turnbull,** Department of Linguistics, University of Hawaiʻi at Mānoa, Honolulu, HI, USA,
E-mail: rory.turnbull@hawaii.edu

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

that the conditions that give rise to segment deletion also necessarily lead to temporal reduction, the reality of the situation may be considerably more complex.[1]

Given the similarity but also the differences between segment deletion and phonetic reduction, it is reasonable to ask whether predictability factors affect segment deletion. For example, coronal consonants in English are more likely to be deleted in words with high conditional probability than in low-probability words (Raymond et al. 2006; Coetzee and Kawahara 2013); an effect also established for continuous phonetic reduction (Jurafsky et al. 2001). Addressing this topic, Cohen Priva (2015) analyzed deletion patterns in close to 70,000 consonant segments in the Buckeye corpus, and found that highly predictable segments are more likely to undergo deletion than less predictable segments. Crucially, segment predictability was defined both globally and locally, suggesting that multiple levels of probability are relevant to these processes. Seyfarth (2014) examined some 40,000 word tokens from the Buckeye corpus and found proportionally more deletions for segmentally longer words.

Theories of probabilistic reduction are purportedly universal (Lindblom 1990; Aylett and Turk 2004; Tyrone and Mauk 2010; Gahl et al. 2012; Hall et al. 2018), and should therefore be equally applicable to all languages. In contrast, regular phonological segment deletion processes are language-specific; a deletion rule in one language does not necessarily apply to another language. For probabilistic segment deletion, then, we may expect some universal tendencies but also language-specific processes.

In the current paper, we present results of an exploratory study of segment deletion patterns in corpora of spontaneous speech in English and Japanese. This investigation encompasses over two million data points, several orders of magnitude more than previous studies. The analysis is necessarily coarse-grained and ignores some linguistically-relevant dimensions, such as morphological structure, but nevertheless can provide insight into overall patterns both within and between these languages (Foulkes et al. 2018).

2 Corpora

2.1 American English

The Buckeye Corpus of Conversational Speech (Pitt et al. 2007) is an orthographically and phonetically transcribed audio corpus of roughly 40 hours of speech. It consists mostly of spontaneous monologues in a conversational style. The audio recordings are transcribed in terms of standard orthography, (dictionary) phonemic transcriptions, and broad phonetic transcriptions.

Since each word in the corpus has both a phonetic and phonological transcription, it is possible to compare the two transcriptions and determine if any phonemes were deleted in the word's pronunciation. For example, the phonological transcription of *Columbus* /kəlʌmbʌs/ can be aligned with an attested phonetic transcription [klʌbʌs] as follows:

k	ə	l	ʌ	m	b	ʌ	s
k		l	ʌ		b	ʌ	s

It can be seen that two phonemes (/ə, m/) were deleted. These alignments were computed using a modified string edit distance algorithm implemented in BioPython (Cock et al. 2009). Allowable substitutions were: any vowel could be substituted for any other vowel; sonorants could be substituted for their syllabic counterparts; and /t/ and /d/ could be substituted for /ɾ/. All other mismatches between the phonemic and phonetic transcriptions were resolved through insertion and deletion operations.

1 The segment deletion considered here is necessarily *probabilistic* deletion, a variable process which cannot be predicted with accuracy. This is distinguished from predictable segment deletion as a result of regular phonological processes such as elision. This paper only considers probabilistic segment deletion.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

Therefore, the production of *getting* /gɛtɪŋ/ as [gɛɾɪŋ] aligns perfectly, with no deletions. On the other hand, the pronunciation of *holding* /hoʊldɪŋ/ as [hoʊŋ̩] involves the deletion of 3 phonemes: /l, d, ɪ/, and the substitution of [ŋ̩] for /ŋ/. These forms are aligned as follows:

h	oʊ	l	d	ɪ	ŋ
h	oʊ				ŋ̩

One advantage of this approach is that it provides a sensible notion of which phonemes were deleted. A disadvantage is the method's reliance on a one-to-one correspondence between the two levels of representation. For example, in the pronunciation of *holding* /hoʊldɪŋ/ as [hoʊŋ̩], it is plausible to suppose that elements of both the /ɪ/ and the /ŋ/ are present in the [ŋ̩]; that is, these two sounds coalesced into a single sound. From this perspective, it is not strictly correct to say that the /ɪ/ (or the /ŋ/) was deleted, but this approach forces us to make that determination. Due to the types of substitutions permitted with the current algorithm, there is a bias toward tagging the vowel as deleted in cases of sonorant syllabification.

From these alignments, the number of deletions per word was calculated, and the identity of the deleted phonemes recorded. The neighborhood density of each word was calculated from the CMU Pronouncing Dictionary, and the frequency of each word was measured from a concatenation of the Buckeye corpus with the Fisher corpus (Cieri et al. 2005). The biphone probability of each phoneme (the probability of the phoneme given the previous phoneme) was estimated from the CMU Pronouncing Dictionary.

2.2 Japanese

The Corpus of Spontaneous Japanese (Furui et al. 2000; Maekawa et al. 2000; Maekawa 2003; Maekawa 2004: CSJ) is an orthographically transcribed audio corpus of over 650 hours of spontaneous speech. Forty-five hours of this corpus include phonological transcriptions which are time-aligned with the acoustic signal. Unlike the Buckeye corpus, however, the alignments in the CSJ do not allow us to unequivocally determine if a phoneme was deleted. Instead, phoneme *reduction* can be detected. Deletion is a type of reduction, but not all reductions are deletions.

In the case of phoneme reduction in the CSJ, two segments are aligned to the same portion of the acoustic signal. For example, Figure 1 depicts a waveform and spectrogram of *hiken sha desu*, "[he] is a test subject", produced as [çkẽɕaes]. It is clear that the first /i/ and last /ɯ/ have been deleted,[2] the /n/ has been merged into the end of the first word, and the /d/ of *desu* has been completely lost without even a trace of a fricative or approximant. As can be seen in the alignment in Figure 1, the deleted or reduced segments have been aligned to the same portion of acoustic signal as the neighboring segments.

From these annotations it is possible to compare the number of segments in the phonological transcription with the number of aligned portions of the acoustic signal, to calculate how many phonemes were deleted in a given word. It is also possible to record, on a per-phoneme basis, if the phoneme was involved in a process of reduction. That is, in Figure 1, both the /e/ and the /n/ are tagged as being involved in a reduction. As noted, this tagging is slightly different from the deleted/not-deleted tagging of the Buckeye corpus: in the Buckeye

2 The case of the reduction of /ɯ/ in words like *desu* is commonly referred to as vowel devoicing, rather than true deletion in its own right (for review, see Fujimoto 2015). The devoiced vowel often remains in the signal as an acoustic trace on adjacent consonants (Faber and Vance 2000). In some cases, however, the vowel is clearly deleted, as in the production of *sentakuki* "washing machine" /sentakɯki/ as [sentakːi], where the deletion of the /ɯ/ leads the two identical stop consonants to coalesce into a geminate (Arai 1999). Complicating this matter of diagnosing devoicing versus deletion is the fact that native language can influence perceptions of vocoid segments in consonant clusters (Dupoux et al. 1999), meaning that even phonetically trained linguists may disagree on whether a phoneme has actually been deleted or simply "reduced". In any case, regardless of whether the vowel is "truly" deleted or "just" devoiced, the outcome is a phonetic reduction which reduces the prominence of a single segment, which is implemented in a probabilistic fashion. Additionally, this paper examines deletions of *all* segments of Japanese, not just the high vowels. If the observed results are simply a consequence of devoicing patterns, then the high vowels /ɯ/ and /i/ ought to be the most-deleted segments; however, as Table 3 shows, this is not the case.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

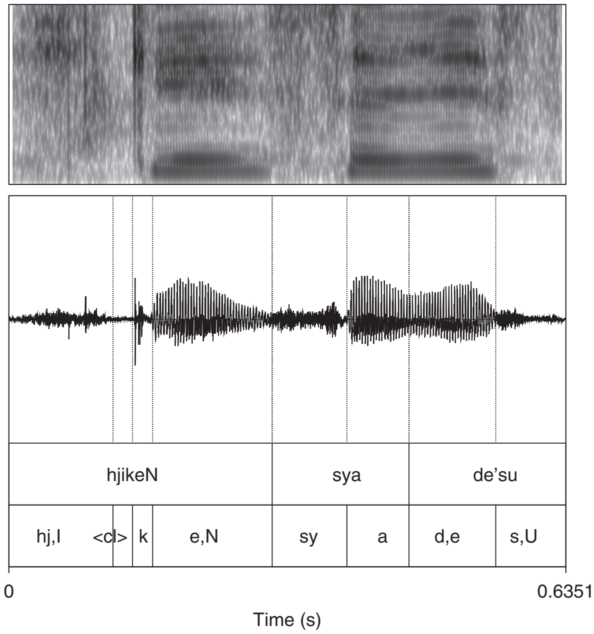

Figure 1: Waveform and spectrogram of *hiken sha desu* [çkẽɕaes] from the phrase *hiken sha desu keredo mo*, "although [he] is a test subject". The first tier lists the phonological transcription of the words. The second tier lists the phonemes as produced, with particular annotations: <cl> represents stop closure, and I and U represent devoiced /i/ and /ɯ/ respectively.

corpus, we can definitively state whether a phoneme was deleted. In the CSJ, we know that a phoneme was reduced in some way, but not necessarily completely deleted. For this reason, we will refer to *deletions* for the English data and *reductions* for the Japanese data.

For each word, frequency and neighborhood density were estimated from the entirety of the CSJ, and for each phoneme, biphone probabilities were estimated from a phonological lexicon derived from the CSJ.

2.3 Summary

The Buckeye corpus has around 40 hours of transcribed speech. The transcriptions allow us to determine if a phoneme is present in the dictionary form but not in the transcription, i.e. was deleted. This definition of deletion also encompasses cases of coalescence, for example the production of /an/ as [ã] would be counted as a deletion of /n/.

The CSJ has around 45 hours of aligned speech. Examination of alignment overlap allows us to determine if a phoneme overlaps with another, i.e. was reduced in some way. This definition of reduction includes the types of deletion covered in the Buckeye corpus above, and also cases where two sounds were produced with an unclear boundary between them (such as a sequence of vowel and sonorant), and therfore is a somewhat more liberal measure than in the Buckeye corpus.

Ideally for such an investigation, the corpora used would involve the same transcription and alignment schema; unfortunately such a resource does not exist with the magnitude of data required. Nevertheless, the conceptual similarity of these measures is apparent, and meaningful comparison is still possible.

3 Modeling

Two sets of modeling were carried out. The word-level analyses predicted, for a given word, how many of its phonemes will be deleted. The phoneme-level analysis predicted, for a given segment, how likely is it to be deleted/reduced.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

3.1 Word-level analysis

For each language, a statistical model was constructed to predict the number of phonemes deleted from each word. This process was modeled using mixed effects Poisson regression. Fixed effects were log-transformed word frequency, neighborhood density, and log-transformed number of phonemes in the phonological transcription. These factors were strongly correlated with each other, and residualization was carried out to parse out the relative contribution of each factor. After z-transforming, frequency and density were independently residualized on number of phonemes. This led to measures of frequency and density which were not correlated with each other or with number of phonemes. These factors can be interpreted as frequency and density after controlling for word length, although care must be taken in their interpretation (York 2012; Wurm and Fisicaro 2014). Random intercepts of word identity were also included. All words with two or more phonemes were included in the model; this led to the inclusion of 259,491 data points in the English model and 375,264 in the Japanese model.

3.2 Phoneme-level analysis

For each language, another statistical model was constructed to predict the likelihood of an individual phoneme token deletion. This process was modeled using mixed effects logistic regression.

Fixed effects were log-transformed word length in number of phonemes, log-transformed word frequency, neighborhood density, and biphone probability. Again, for both languages, frequency, neighborhood density, and word length were significantly correlated with each other, so density and frequency were independently residualized against word length. All factors were z-transformed before being entered into the model. Random intercepts of word identity and phoneme identity were included. In total, 890,213 data points were included in the English model and 1,231,846 in the Japanese model.

As described, phoneme identity was included as a random intercept in the binomial regression models. These intercept values represent the variance in deletion rates for each phoneme, which is not accounted for by the fixed effects of word length, word frequency, neighborhood density, or biphone probability. These values can therefore be regarded as a measure of "deletion rate" after controlling for these confounding factors.

4 Results

4.1 Number of deletions per word

The model output of the Poisson regressions predicting the number of deletions per word for English and Japanese are shown in Table 1. In both languages, an effect of the number of phonemes was observed, such

Table 1: Model output for regression model predicting number of phoneme deletions per word in English and Japanese.

Effect	β	SE	z	p-Value
Language: English				
Intercept	−2.356	0.044	−53.28	<0.001
Number of phonemes	0.744	0.017	43.36	<0.001
Residualized frequency	0.540	0.027	19.97	<0.001
Residualized density	−0.225	0.035	−6.48	<0.001
Language: Japanese				
Intercept	−2.868	0.040	−72.35	<0.001
Number of phonemes	0.840	0.020	42.95	<0.001
Residualized frequency	0.140	0.023	5.97	<0.001
Residualized density	0.133	0.032	4.20	<0.001

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

that words with more phonemes are likely to have more deletions, concordant with Johnson's (2004) findings. This effect is graphed in Figure 2. There is a hard limit on the number of possible deletions for a given word, and that limit is larger for longer words than for shorter words. Also for both languages, more deletions were observed for words with higher (residualized) frequency, depicted in Figure 3. This result demonstrates that word frequency has an influence on phoneme deletion, independently of word length.

Both languages also exhibited significant effects of residualized neighborhood density, but in opposite directions. In English, higher residualized density led to fewer deletions, while in Japanese, higher residualized density led to more deletions. These effects are depicted in Figure 4. Inspection of this figure reveals that

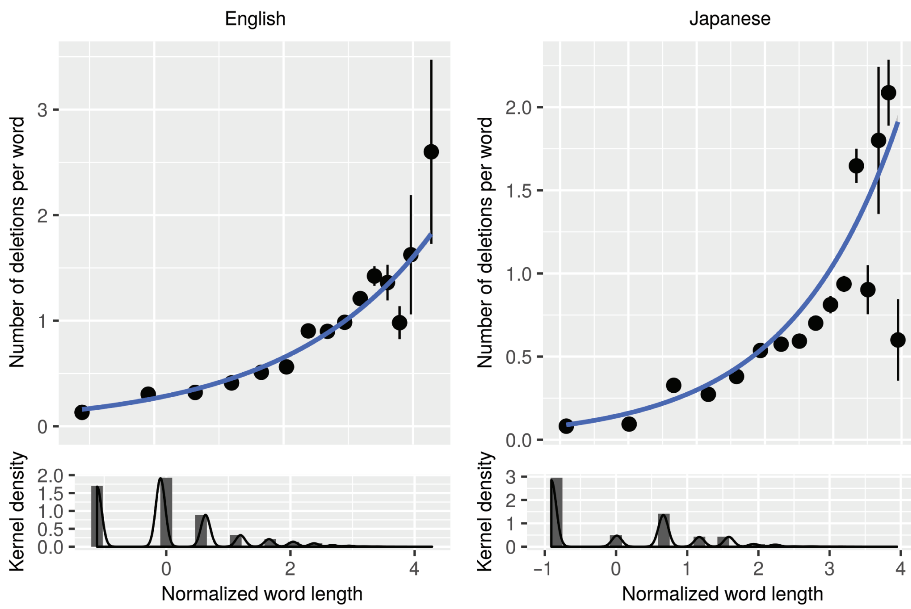

Figure 2: Number of deletions per word as a function of normalized word length in phonemes for English and Japanese. Poisson curve overlaid. Histogram and density plot shown in lower panel.

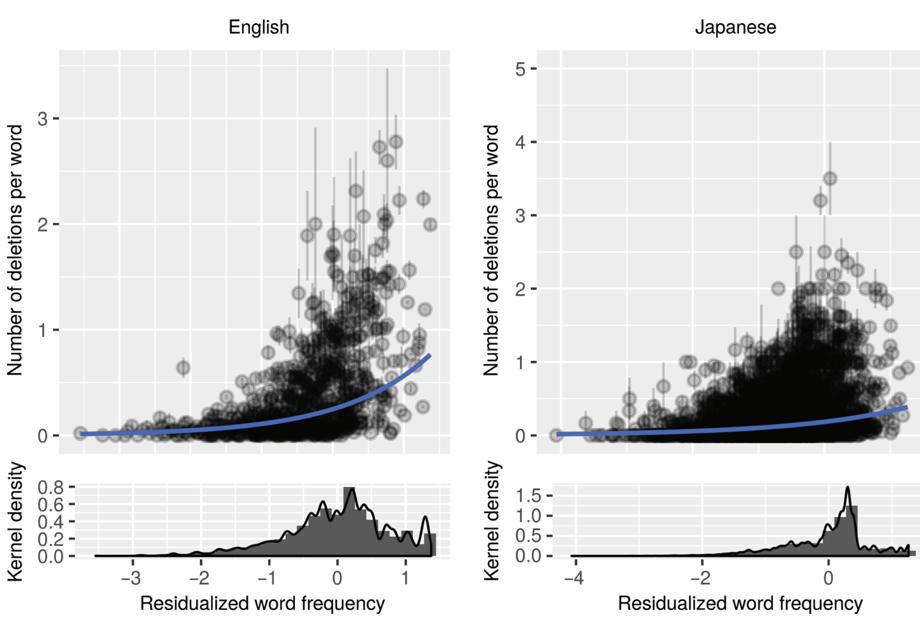

Figure 3: Number of deletions per word as a function of residualized word frequency for English and Japanese. Poisson curve overlaid. Histogram and density plot shown in lower panel.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

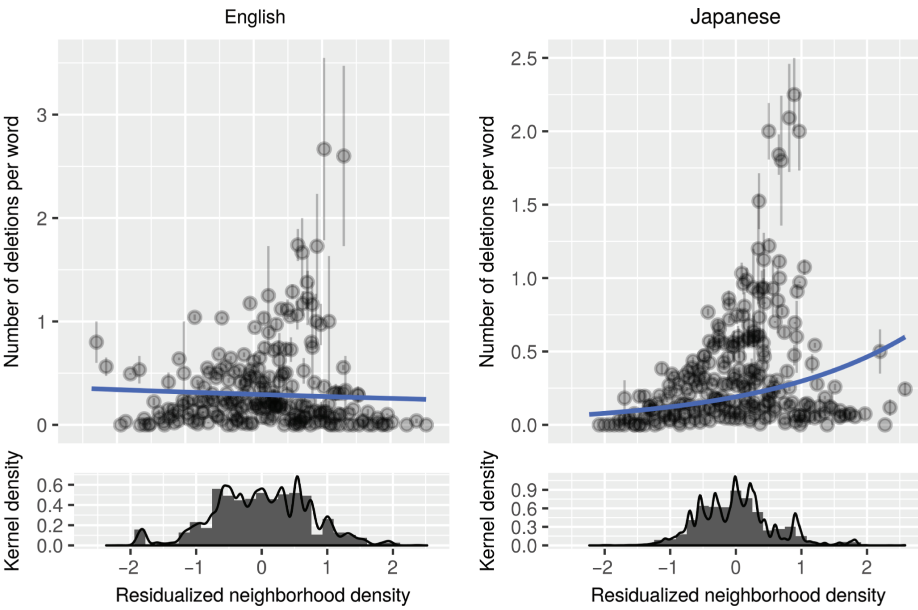

Figure 4: Number of deletions per word as a function of residualized neighborhood density for English and Japanese. Poisson curve overlaid. Histogram and density plot shown in lower panel.

the modeled slope is in fact a poor fit for this kind of data, and that the effect of neighborhood density on deletions is non-linear. Deletions appear to be common for words with a residualized density value between 0 and 1, and generally low outside of this window. English appears to have a relatively high number of deletions for some low-density words, while Japanese has a high number of deletions for some high-density words, causing the differences in signs between the coefficients in the model. In any case, though, the majority of the deletions for both languages is constrained to the window between 0 and 1. The pattern that emerges is that words with *extreme* density values – either extremely high or extremely low – are likely to have fewer deletions than words with intermediate density values.

4.2 Likelihood of deletion per phoneme

Table 2 shows the model output of the binomial models predicting probability of deletion per phoneme for English and Japanese. Both languages show an effect of biphone probability, such that phonemes which are

Table 2: Model output for regression model predicting probability of deletion per phoneme in English and Japanese.

Effect	β	SE	z	p-Value
Language: English				
Intercept	−3.522	0.149	−23.69	<0.001
Number of phonemes	0.009	0.012	0.77	0.440
Residualized frequency	−0.001	0.012	−0.11	0.909
Residualized density	0.012	0.016	0.76	0.448
Biphone probability	0.025	0.006	3.84	<0.001
Language: Japanese				
Intercept	−1.785	0.043	−41.69	<0.001
Number of phonemes	0.115	0.022	5.25	<0.001
Residualized frequency	0.192	0.020	9.70	<0.001
Residualized density	−0.145	0.026	−5.52	<0.001
Biphone probability	0.197	0.005	36.32	<0.001

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

highly probable given the previous phoneme are more likely to be deleted than phonemes which are less probable given the previous phoneme. This effect is graphed for both languages in Figure 5.

Word length (measured as number of phonemes), word frequency, and neighborhood density were observed to have significant effects on probability of phoneme deletion in Japanese, but not in English. These effects are depicted in Figures 6, 7, and 8, respectively. In Japanese, phonemes in longer words were more likely to be deleted than phonemes in shorter words; phonemes in higher-frequency words were more likely to be deleted than phonemes in lower-frequency words; and phonemes in words with fewer neighbors were more likely to be deleted than phonemes in words with more neighbors.

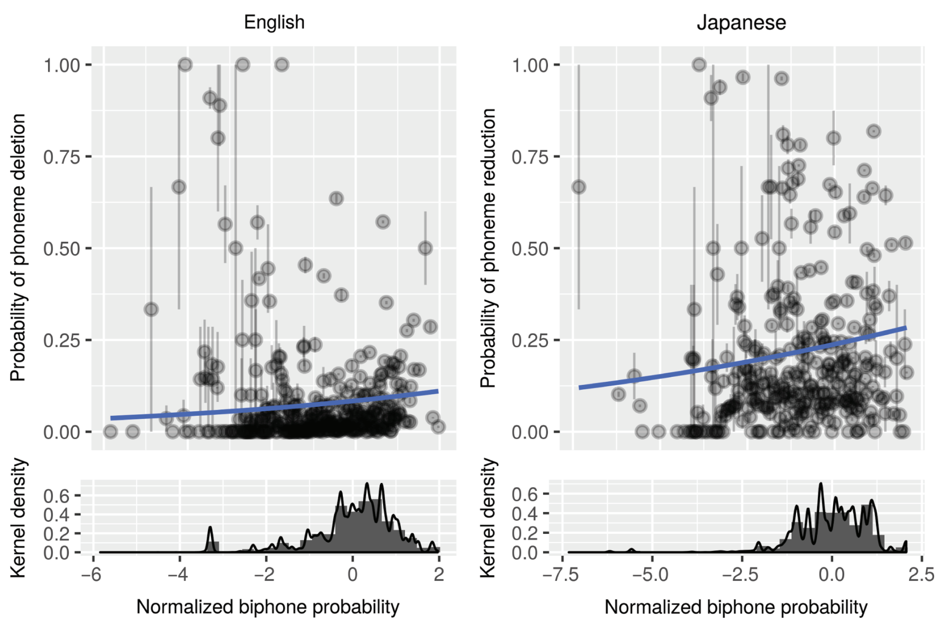

Figure 5: Probability of deletion or reduction per phoneme as a function of normalized biphone probability for English and Japanese. Binomial curve overlaid. Histogram and density plot shown in lower panel.

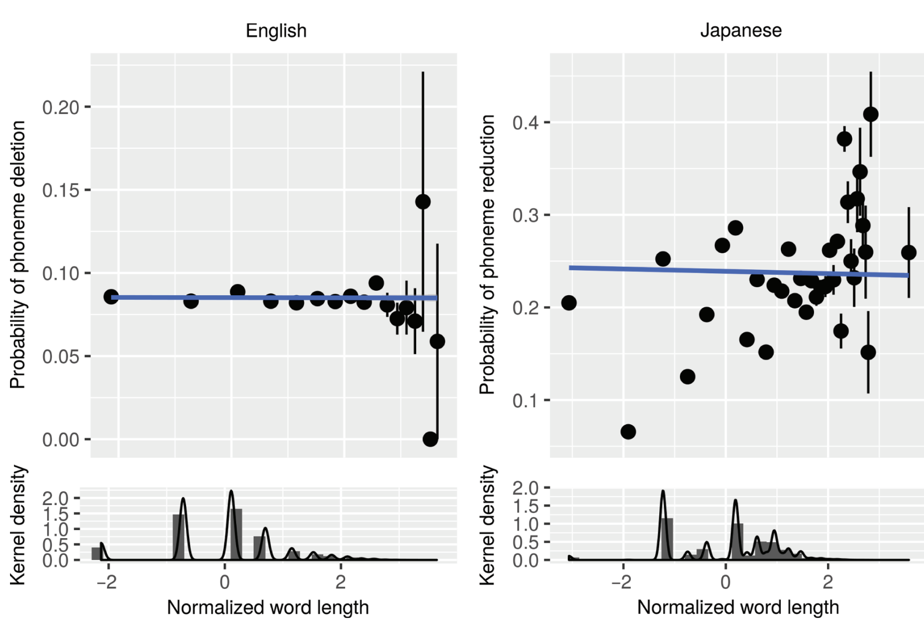

Figure 6: Probability of deletion or reduction per phoneme as a function of normalized word length for English and Japanese. Binomial curve overlaid. Histogram and density plot shown in lower panel.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

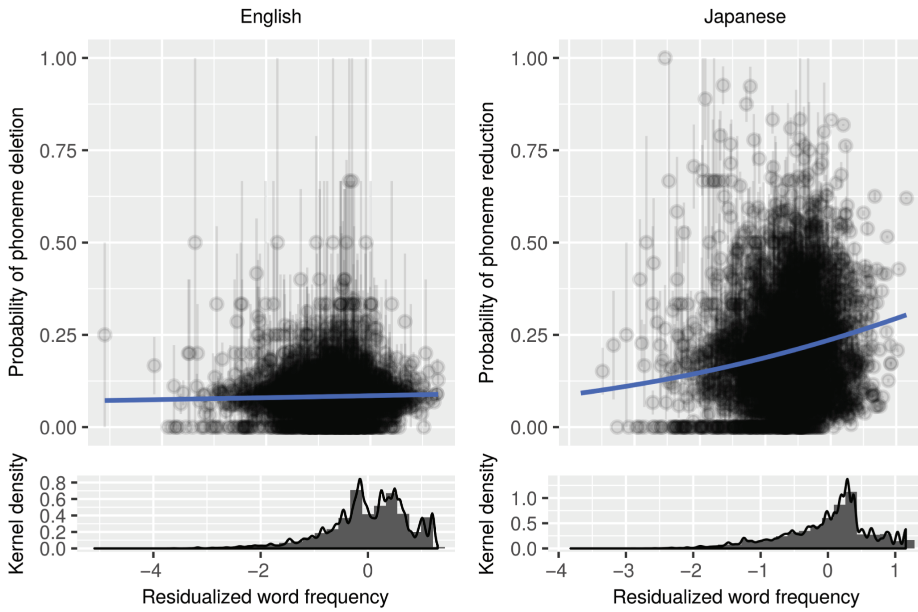

Figure 7: Probability of deletion or reduction per phoneme as a function of residualized word frequency for English and Japanese. Binomial curve overlaid. Histogram and density plot shown in lower panel.

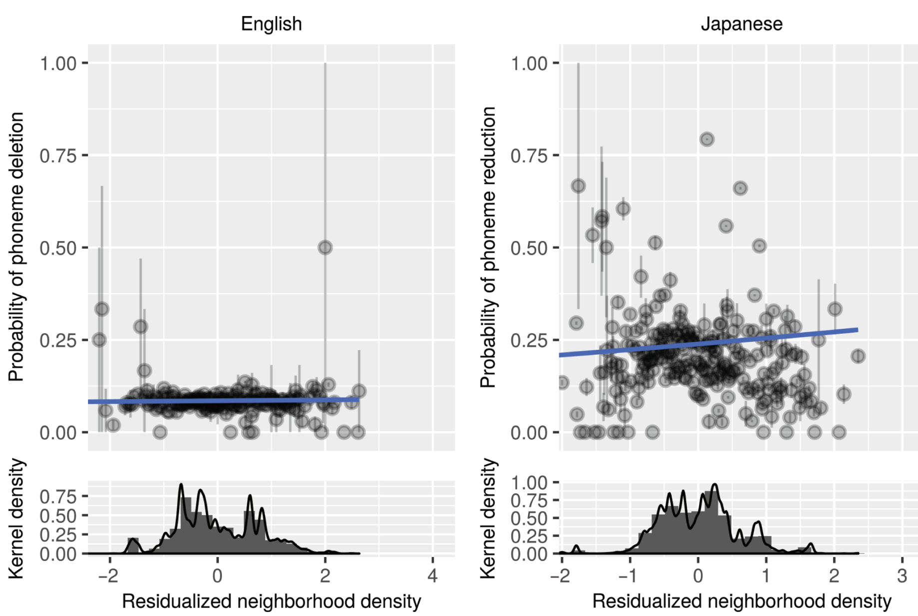

Figure 8: Probability of deletion or reduction per phoneme as a function of residualized neighborhood density for English and Japanese. Binomial curve overlaid. Histogram and density plot shown in lower panel.

4.3 Deletion rates

Random intercept values for each phoneme in English and Japanese are listed in Table 3. Intercepts are listed in descending order; the larger the value, the more likely that phoneme is to be deleted. Examination of these values reveals some expected patterns. The English phonemes which are most prone to deletion are /d, ɹ, h, t, v/, which is concordant with previous research about phoneme deletion in spontaneous English (Shockey 2003; Johnson 2004; Turnbull 2015). Among the phonemes which are least prone to deletion are the diphthongs /aɪ, aʊ, ɔɪ/.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

Table 3: Random by-phoneme intercepts for each English and Japanese phoneme.

Japanese		English	
Phoneme	**Intercept**	**Phoneme**	**Intercept**
g	1.468	d	2.451
b	1.386	ɹ	2.183
d	0.862	h	2.104
ɯ	0.774	t	2.012
h	0.745	v	1.951
ŋ	0.662	n	1.841
s	0.631	ð	1.524
z	0.574	l	1.174
tː	0.375	ə	1.106
i	0.352	b	1.038
w	0.341	æ	0.980
e	0.270	ɪ	0.896
p	0.152	w	0.837
o	0.104	j	0.797
n	0.064	ŋ	0.489
k	0.045	θ	0.355
t	0.030	ʊ	0.235
aː	−0.024	u	0.219
oː	−0.113	k	−0.139
pː	−0.143	ʒ	−0.166
j	−0.149	m	−0.191
sː	−0.258	ɛ	−0.223
a	−0.284	tʃ	−0.240
r	−0.438	dʒ	−0.474
iː	−0.489	z	−0.539
ɯː	−0.504	g	−0.653
kː	−0.956	oʊ	−0.770
m	−1.144	p	−1.067
eː	−1.384	ɚ	−1.068
		i	−1.104
		eɪ	−1.108
		s	−1.190
		aʊ	−1.270
		f	−1.287
		ɑ	−1.479
		aɪ	−2.393
		ɔɪ	−2.414
		ʃ	−2.679

Higher values indicate higher probabilities of deletion.

In Japanese, the phonemes which are most prone to deletion are /g, b, d, ɯ, h/. There is no relevant literature on Japanese spontaneous speech to compare these results to. However, they make sense from the point of view of Japanese phonology (Vance 1987): the voiced stops /b, d, g/ are common in intervocalic position (and rare in initial position); /ɯ/ is commonly devoiced or even somewhat blended with the preceding consonant (see also Fujimoto 2015). Among the least common phonemes to be deleted in Japanese are the long vowels and geminate consonants. (See also Sano 2018, on Japanese geminates.)

The consistent pattern in both languages is that segments with a long intrinsic duration (long vowels, geminate consonants, diphthongs) are generally less likely to be deleted. Conversely, the most commonly deleted segments appear to be related to language-specific patterns.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

5 Discussion

The regression models demonstrate several factors influencing segment deletion in English and Japanese. Some of these factors are relevant for both languages, while others are restricted to Japanese alone.

The word-level models, which predicted the number of deletions per word, show effects of word length. In both languages, words with more phonemes have more deletions. This finding replicates that of Seyfarth (2014) for English, and extends it to Japanese. Frequency was also found to be relevant: in both languages, higher-frequency words have more deletions than low frequency words. Frequency was residualized against word length, so this effect is independent of any linear effects of word length. It is also consistent with the notion that high-frequency words tend to be subject to phonetic reduction (e.g. geminate devoicing; see Coetzee and Kawahara 2013) to a greater degree than low-frequency words. The residualized density effects are significant for both languages, but in different directions. Visual inspection of the patterns, depicted in Figure 4, suggests that the effect of density on deletions is actually nonlinear. Deletions are more common in words of average density (given word length), while in words in extremely dense or extremely sparse neighborhoods, deletion is less common. Phoneme deletion in a dense neighborhood could lead to segmental overlap with a lexical competitor, and thus lead to transmission errors in speech; however, there is no clear theoretical motivation for inhibition of deletion in *sparse* neighborhoods.

The phoneme-level analyses, predicting probability of deletion for individual phonemes, reveal an effect of biphone probability. In both languages, phonemes with a higher conditional probability are more likely to be deleted than those with a lower conditional probability. However, the other effects in the phoneme-level analyses show differences between the English and Japanese data. In Japanese, phonemes in longer words, higher-frequency words, and lower-density words are all more likely to be deleted than phonemes in shorter words, lower-frequency words, and higher-density words. These effects are consistent with our understanding of predictability effects more generally. In English, however, none of these effects were obtained. Notably, Cohen Priva (2015) also failed to find an effect of word frequency in predicting phoneme deletion in a subset of the Buckeye corpus. It is not known why Japanese phonemes should be sensitive to word frequency when English phonemes are not.

A possible source for these differences between English and Japanese is in how deletion was defined for the phoneme-level analysis. Recall from Section 2.3 that for English, we are measuring deletion, while for Japanese, we are measuring reduction, of which deletion is a subtype. This more liberal measure for Japanese then includes phonetic reductions which are not "true" deletions, such as lenition of voiced stops into approximants in intervocalic position. This difference artificially inflates the number of deletions in Japanese relative to English – note the higher model intercept and biphone probability coefficient for Japanese relative to English in Table 2. For example, it could be the case that word frequency does not influence phoneme deletion, but that it does influence non-deletion reductions (such as lenition). Such a scenario would give rise to the pattern observed in Table 2, where a frequency effect is observed for Japanese but not for English. However, this interpretation is impossible to confirm without details on the precise realization of the Japanese reduced phonemes, information which is currently unavailable.[3]

An alternative source of differences between the languages is word structure. Specifically, English has a relatively flexible phonotactic system, and a single syllable can range from 1 to 7 phonemes in length. Japanese, on the other hand, has maximally 3 phonemes per syllable, and the great majority of syllables are biphonemic. If segments are more likely to be deleted in words with more syllables, then this tight connection between number of phonemes and number of syllables could lead to an observed effect in Japanese, but not in English, where the relationship between number of syllables and number of phonemes per word is much more loose.

3 A total of 294,411 phonemes are tagged as reduced. On a conservative estimate, we can assume it takes a phonetically-trained transcriber around 5 seconds on average to listen to and tag each token into a reduction category. Tagging every reduced token in this way would take at least 400 hours.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

Nevertheless, the differences between the languages give rise to questions of what counts as a "universal" effect. For example, suppose some effect *X* is observed for language *A*, but not for language *B*. Further suppose that it turns out that language *B* has some process *Y* which makes *X* impossible. To what extent is it then reasonable to claim that effect *X* is universal? One option is to claim that *X* applies in language *B*, but it is blocked by *Y*, and thus never "surfaces". Alternatively, one could claim that *X* only applies in language *A*, and never in *B*. This interpretational freedom, coupled with the often scant formal details of theories of predictability effects, can lead to a great deal of flexibility in the falsifiability of a particular hypothesis. Are predictability effects universal? Some patterns in the present data were observed for both English and Japanese, while others were observed solely for Japanese.

A consensus-seeking interpretation is that predictability effects *in general* are universal, but each language may realize these effects in different ways (Turnbull et al. 2015), and that different languages may be sensitive to different kinds of probability (e.g. global, local, or contextual; see Cohen Priva 2015; Turnbull 2017). This interpretation is weak in terms of prediction and falsifiability, and does not constrain well the possible hypothesis space. Moving forward, researchers must take care when choosing between theories which are general and unsophisticated or complex and specific, particularly when considering theoretical integration with other theories of aspects of speech communication and language processing (Cohen Priva and Jaeger 2018; Ferrer-i-Cancho 2018).

While the data in this current paper do not necessarily adjudicate between particular theoretical disputes, they do provide a solid empirical basis upon which future work can be developed. Simply visualizing these patterns of segment deletion is enormously helpful in generating specific hypotheses, which can be subjected to targeted experimental validation. The volume of data considered here is several orders of magnitude larger than previous studies of segment deletion (Cohen Priva 2015; Johnson 2004).

6 Conclusion

This paper has presented results of an exploratory study of segment deletion patterns in corpora of spontaneous speech in English and Japanese. Several patterns are observed to be common to both languages: more deletions are observed in words that are longer and higher in frequency. Words with extremely high or extremely low neighborhood density tend to have few deletions. Phonemes within high-probability phoneme sequences are more likely to be deleted than those in low-probability sequences. Table 3 lists the relative deletion proportions of each phoneme in each language, after controlling for lexical effects. The phonemes with longer intrinsic duration (long vowels, geminate consonants, diphthongs) generally had deletion rates below the median.

On the other hand, some patterns were inconsistently observed: Japanese phonemes are more likely to be deleted in words that have high frequency, few neighbors, and many phonemes, relative to words with low frequency, many neighbors, and few phonemes. These effects were not observed in English. Likewise, there is substantial mismatch between the shared phonemes in Table 3: for example, Japanese /s/ is in the top 10 of most-deleted phonemes, while English /s/ is in the bottom 10.

The shared effects may well be due to universal properties of language, while others may be due to language-specific patterns such as word prosody, phonetic differences in phoneme inventory, or methodological differences in the parameterization of deletion. Nevertheless, the consistency of the effects over such a large volume of data is suggestive.

Future work should consider language-specific factors, such as the structure of the English and Japanese lexicons, which may influence the role of predictability, particularly neighborhood density. Moreover, controlling for higher-level factors, such as speech rate, morphological structure, syntax, discourse and narrative context, and speech style, is essential in taming the variability intrinsic to spontaneous speech corpora. Extending this work to other languages is likewise an important goal, particularly to evaluate any claims of universality.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

Acknowledgements: I am grateful to Mary Beckman, Cynthia Clopper, Shigeto Kawahara, Rebecca Morley, Shari Speer, and three anonymous reviewers for feedback on previous versions of this paper, and to Adriana Guevara Rukoz for discussions on this topic. All errors and shortcomings are my own.

Funding: This work was partially supported by an Ohio State University Presidential Fellowship.

References

Arai, T. 1999. A case study of spontaneous speech in Japanese. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (eds.), *14th International Congress of Phonetic Sciences,* 615–618. San Francisco, CA: International Phonetic Association.

Aylett, M. & A. E. Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1). 31–56.

Cieri, C., D. Graff, O. Kimball, D. Miller & K. Walker. 2005. *Fisher English Training Part 2*. Philadelphia, PA: Linguistic Data Consortium.

Cock, P. J. A., T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski & M. Hoon. 2009. Biopython: Freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25(11). 1422–1423.

Coetzee, A. W. & S. Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 31. 47–89.

Cohen Priva, U. 2015. Informativity affects consonant duration and deletion rates. *Laboratory Phonology* 6(2). 243–278.

Cohen Priva, U. & T. F. Jaeger. 2018. The interdependence of frequency, predictability, and informativity. *Linguistics Vanguard* 4(S2).

Dupoux, E., K. Kakehi, Y. Hirose, C. Pallier & J. Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25(6). 1568–1578.

Ernestus, M. & R. H. Baayen. 2007. The comprehension of acoustically reduced morphologically complex words: The roles of deletion, duration, and frequency of occurrence. In J. Trouvain & W. J. Barry (eds.), *Proceedings of the Sixteenth International Congress of the Phonetic Sciences,* 773–776. Saarbrücken: International Phonetic Association.

Faber, A. & T. J. Vance. 2000. More acoustic traces of "deleted" vowels in Japanese. In M. Nakayama & C. J. Quinn (eds.), *Japanese/Korean linguistics: volume 9,* 100–113. Stanford, CA: CSLI Publications.

Ferrer-i-Cancho, R. 2018. Optimization models of natural communication. *Journal of Quantitative Linguistics* 25(3). 207–237. https://doi.org/10.1080/09296174.2017.1366095.

Foulkes, P., G. Docherty, S. Shattuck-Hufnagel & V. Hughes. 2018. Three steps forward for predictability. Consideration of methodological robustness, indexical and prosodic factors, and replication in the laboratory. *Linguistics Vanguard* 4(S2).

Fujimoto, M. 2015. Vowel devoicing. In H. Kubozono (ed.), *The handbook of Japanese language and linguistics: Phonetics and phonology*, 167–214. Berlin: de Gruyter.

Furui, S., K. Maekawa & H. Isahara. 2000. A Japanese national project on spontaneous speech corpus and processing technology. *Proceedings of ISCA ITRW ASR2000.* 244–248.

Gahl, S. 2008. *Time* and *thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3). 474–496.

Gahl, S., Y. Yao & K. Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66(4). 789–806.

Hall, K. C., E. V. Hume, T. F. Jaeger & A. Wedel. 2018. The role of predictability in shaping phonological patterns. *Linguistics Vanguard* 4(S2).

Johnson, K. 2004. Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (eds.), *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th International Symposium.* Tokyo: The National Institute for Japanese Language.

Jurafsky, D., A. Bell, M. Gregory & W. D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure,* 229–254. Amsterdam: John Benjamins.

Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modelling,* 403–439. Dordrecht: Kluwer.

Maekawa, K. 2003. Corpus of spontaneous Japanese: Its design and evaluation. In *Proceedings of ISCA and IEEE workshop on spontaneous speech processing and recognition,* 7–12.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM

Maekawa, K. 2004. Design, compilation, and some preliminary analyses of the corpus of spontaneous Japanese. In K. Maekawa & K. Yoneyama (eds.), *Spontaneous speech: Data and analysis,* vol. 3, 87–108. Tokyo: The National Institute of Japanese Language.

Maekawa, K., H. Koiso, S. Furui & H. Isahara. 2000. Spontaneous speech corpus of Japanese. *Proceedings of LREC 2000*. 947–952.

Munson, B. & N. P. Solomon. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47. 1048–1058.

Pitt, M. A., L. C. Dilley, K. Johnson, S. Kiesling, W. Raymond, E. V. Hume & E. Fosler-Lussier. 2007. *Buckeye corpus of conversational speech*. Columbus, OH: Department of Psychology, Ohio State University.

Raymond, W. D., R. Dautricourt & E. Hume. 2006. Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* 18. 55–97.

Sano, S. 2018. Durational contrast in gemination and informativity. *Linguistics Vanguard* 4(S2).

Seyfarth, S. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133. 140–155.

Shockey, L. 2003. *Sound patterns of spoken English*. Oxford: Wiley-Blackwell.

Turnbull, R. 2015. *Assessing the listener-oriented account of predictability-based phonetic reduction*. Columbus, OH: Ohio State University Dissertation.

Turnbull, R. 2017. The role of predictability in intonational variation. *Language and Speech* 60(1). 123–153.

Turnbull, R., R. S. Burdin, C. G. Clopper & J. Tonhauser. 2015. Contextual information and the prosodic realization of focus: A cross-linguistic comparison. *Language, Cognition and Neuroscience* 30(9). 1061–1076.

Tyrone, M. E. & C. E. Mauk. 2010. Sign lowering and phonetic reduction in American Sign Language. *Journal of Phonetics* 38(2). 317–328.

Vance, T. J. 1987. *An introduction to Japanese phonology*. Albany, NY: SUNY Press.

Wright, R. 2004. Factors of lexical competition in vowel articulation. In J. Local, R. Ogden & R. Temple (eds.), *Papers in laboratory phonology IV: Phonetic interpretation,* 75–85. Cambridge: Cambridge University Press.

Wurm, L. H. & S. A. Fisicaro. 2014. What residualizing predictors in regression analyses does (and what it does *not* do). *Journal of Memory and Language* 72. 37–48.

York, R. 2012. Residualization is not the answer: Rethinking how to address multicollinearity. *Social Science Research* 41. 1379–1386.

Authenticated | rory.turnbull@hawaii.edu author's copy
Download Date | 10/8/18 8:10 AM