Predictability, Word Frequency and Japanese Perceptual Epenthesis

Alexander J. Kilpatrick[1], Shigeto Kawahara[2], Rikke L. Bundgaard-Nielsen[3], Brett J. Baker[1], Janet Fletcher[1]

[1]University of Melbourne, [2]Keio University, [3]MARCS Institute for Brain, Behaviour, and Development, Western Sydney University
alex.kilpatrick@unimelb.edu.au

ABSTRACT

Speakers typically invest less effort in the articulation of sounds and words that are highly predictable from their contexts. Recent research reveals a perceptual corollary to this behaviour, showing that listeners pay less attention to acoustic signal in predictable contexts. The present paper expands on this finding by testing the acceptability and discriminability of sequences of speech with varying levels of predictability. Stimuli are contrast pairs and are either phonotactically attested or else contain an illicit non-homorganic consonant cluster. Such clusters violate Japanese phonotactics and have been found to elicit perceptual epenthesis in Japanese listeners. The results show that unattested consonant clusters are perceived as more acceptable in high-frequency sequences than in low-frequency sequences.

Keywords: Japanese, Perceptual Epenthesis, Vowel Devoicing, Predictability.

1. INTRODUCTION

Speakers typically produce segments and words less robustly (shorter, quieter and/or with more centralized vowels) in predictable contexts (see [10]). Studies that examine these phenomena often frame their findings within Zipf's [12] principle of least effort: humans tend to use the least amount of energy required to accomplish a task successfully. This theory assumes that listeners can retrieve meaning from predictable contexts, even with degraded incoming acoustic signals—speakers would not risk reducing sounds if it impeded comprehensibility. Indeed, recent research has shown that—much like speakers—listeners pay less attention to acoustic signal when upcoming input is predictable, making use of top-down knowledge of their native language (L1) experience [4]. These findings are tested further in the present paper by with respect to the problem of 'perceptual epenthesis'.

When listeners are exposed to phonotactically unattested sequences of speech, they sometimes misperceive the signal as adhering to native phonotactics, a process sometimes described as perceptual repair. A well-described example of this is perceptual epenthesis which involves the repair of unattested consonant clusters with epenthetic vowels [1]. Japanese generally disallows non-homorganic consonant clusters [11] and is thus an excellent language to study in this regard: Japanese listeners sometimes perceptually repair a phonotactic cluster violation with an epenthetic /u/ [1], although they have also been found to perceive /i/ when the consonant preceding the epenthetic context has a high /Ci/ transitional probability (or is predictable given its context) [6].

A number of studies have examined the role of predictability in perceptual epenthesis. One such study uses real and nonce words of varying length (measured in terms of mora counts) as a proxy for predictability [4]. This study—conducted under similar conditions, although at different times and with different participants as the current experiments—is based on the premise that when real words get longer, the identity of upcoming signal becomes more predictable, because words attain their lexical uniqueness point [9], while no such effect should occur in unfamiliar/nonce words. The experiments revealed a clear effect of predictability on Japanese listeners' ability to discriminate between phonotactically attested sequences and unattested consonant clusters, whereby longer-word contrasts are less discriminable when the epenthetic context is predictable.

The present paper extends [4] by examining the acceptability of contrast tokens which maintain varying levels of predictability. One note on our methodology is in order. Surprisal and Entropy have often been used to quantify predictability in recent linguistic studies (see [10]). However, this was not possible for the present study because an examination of the Corpus of Spontaneous Japanese [7] revealed that the target vowel in /CimaʃVta/ sequences is very rarely anything other than /i/. Therefore, Surprisal and Entropy are inappropriate for the present study. Instead, the current study uses word frequency as a proxy for predictability, while noting that these two measures are not identical (although very often correlated).

The current experiment measures the acceptability of sequences of phonotactically legal or illegal sequences using Goodness of Fit (GoF) scores. However, it is problematic to describe these as legal/illegal contrasts because—in the upcoming experiments—Japanese listeners assign higher GoF

scores to tokens that supposedly violate Japanese phonotactics (see discussion in the Results section). We therefore refer to tokens that contain no evidence of a vowel in the epenthetic position as "reduced tokens" and those that contain clear evidence of a vowel as "unreduced tokens".

Here, listeners are predicted to assign higher GoF scores to reduced tokens when they are more predictable (as determined by word frequency) due to listeners being tuned to the effects of predictability on speech production. This should result in more divergent GoF scores and—in turn—greater discriminability in predictable contexts. However, this is seemingly paradoxical because [4] shows that predictability decreases discriminability. The present paper tests these two effects in two experiments. In Experiment 1, Japanese listeners categorise and assign GoF scores to reduced and unreduced sequences which represent words that occur at varying frequencies. In Experiment 2, listeners distinguish those same sequences in a series of AXB discrimination tests.

2. METHOD

The stimuli for the following experiments are listed in Table 2; [maʃita] and [maʃta] were also recorded. From these additional recordings, [ʃita] and [ʃta] portions were excised and spliced into unreduced sequences. Excising and splicing occurred 10 ms into the medial [ʃ] in all cases. Splicing was conducted to ensure that the epenthetic context was identical across contrasts. All tokens constitute real words in Japanese except for two; [pimaʃita] is a phonotactically legal nonce word, while [simaʃita] either violates the Japanese co-occurrence restriction */si/ or [si] is a mispronunciation of /ʃi/, depending on your view of Japanese phonology. Tokens that represent real words are listed in Table 2 according to the rate that they occur in Japanese discourse as counted in the Corpus of Spontaneous Japanese [7]. Five repetitions of each target word were recorded by the first author (who is a non-native but fluent speaker of Japanese) in a recording studio located at the University of Melbourne and had a bit depth of 64 kb/sec and a sample rate of 48 kHz. The first and fifth repetitions were not used to avoid the effects of list initial unfamiliarity and list final intonation patterns.

19 L1 Japanese speakers (13 Female; *M* age 28 years) living in Murayama, Japan were recruited for the following experiments All participants had had limited exposure to languages other than Japanese, apart from compulsory English instruction. One participant's responses in Experiment 2 suggested that they had not understood the instructions and their data was excluded from the analyses.

Experiments took place in quiet rooms located in Murayama, Japan. Experiments 1 and 2 occurred in succession. Experiment 1 consisted of 108 trials where participants were asked to assign tokens to categories represented in Hiragana (the most basic Japanese transcription system, which offers no way of representing non-homorganic clusters). After each stimulus, participants were re-exposed to the token and asked to assign a GoF rating to indicate how well the token fit to the assigned category. This was presented to participants as a Likert scale ranging from 1 (different) to 7 (identical). If participants failed to respond to either sub-task within 3500 ms, the trial would time-out and both the categorisation and goodness of fit judgment sub-tasks would be replayed at a random time later in the experiment.

Table 1: Stimuli used in Experiments 1 and 2, with gloss and frequency of occurrence in the Corpus of Spontaneous Japanese (Count).

Unreduced	Reduced	Gloss	Count
[ʃimaʃita]	[ʃimaʃta]	Did	7305
[kimaʃita]	[kimaʃta]	Came/Wore	4130
[mimaʃita]	[mimaʃta]	Saw	1450
[nimaʃita]	[nimaʃta]	Resembled	6
[pimaʃita]	[pimaʃta]	*Nonce*	--
[simaʃita]	[simaʃta]	*Nonce - Violation*	--

Experiment 2 consisted of 2 identical blocks of 144 AXB discrimination trials (288 trials in total). To avoid phone sequence bias, tokens were organised into AXB triads by way of a Latin square; within each block, tokens were presented at random. Participants responded to the task by pressing either "1" or "3" on a keyboard to indicate whether the middle token (X) was a better match to the first (A) or third (B) token. Stimuli were spaced with a 1000 ms interval and trials were spaced with a 1500 ms interval. If participants failed to respond to trials within 2000 ms from the onset of the B token, the trial would time-out and be replayed at a random time later in the experiment.

3. RESULTS

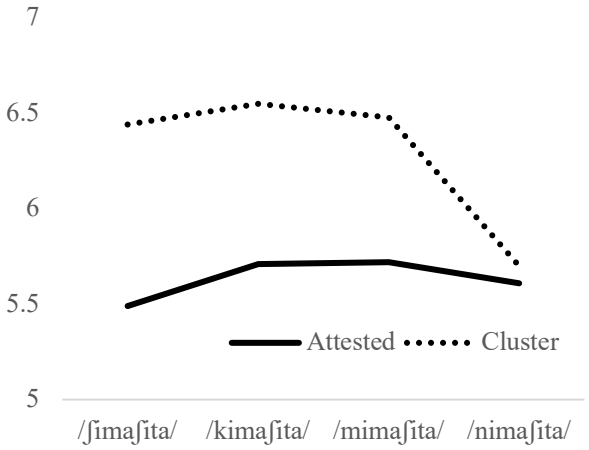
In the categorisation phase of Experiment 1, participants assigned both reduced and unreduced tokens to their predicted categories almost all of the time. There was, however, some confusion with nasal onset sequences where tokens that begin with [m] (/#m/ tokens) were occasionally assigned to /nimaʃita/ and vice versa. This confusion is likely only observed in nasal onset tokens because nasal consonants are demonstrably perceptually similar to each other, and more likely to undergo place assimilation than oral stops [2]. [#n] tokens were

assigned to the /mimashita/ category (M = 18%, SD = 18%) at a far greater rate than [#m] tokens were assigned to the /nimaʃita/ category (M = 5%, SD = 9%). A paired-sample t-test revealed a significant difference between [#m] token and [#n] token mis-categorisation rates ($t(18) = 5.1$, $p < .001$). This asymmetrical result is likely due to a lexical frequency bias: the lower frequency [#n] tokens are more likely to be misperceived as higher frequency /mimaʃita/ rather than the other way around. Of the nasal onset sequences, reduced tokens were more likely to be assigned to an unexpected category (M = 14%; SD = 9%) than unreduced tokens (M = 9%; SD = 9%). Again, a paired-sample t-test revealed a significant difference between reduced and unreduced token mis-categorisation rates ($t(18) = 7.3$, $p < .001$). This asymmetry is difficult to account for because—while non-homorganic consonant clusters are unattested in Japanese—reduced tokens—which maintain no evidence of a vowel in the target position—achieved higher GoF scores (M = 6, SD = 1.2) than unreduced tokens (M = 5.4; SD = 1.7). Similar studies have shown that tokens that contain consonant clusters achieve higher GoF scores than tokens that adhere to Japanese phonotactics when the consonant preceding the epenthetic context is a voiceless fricative and attribute this behaviour to vowel devoicing behaviour, arguing that vowel-less sequences are a nearer match to devoiced vowels than more robust, fully voiced vowels [5]. A paired-sample t-test revealed a significant difference between reduced token and unreduced token GoF scores ($t(113) = 4.6$, $p < .001$). For existing words, this difference appears to correlate with word frequency (see Table 2) whereby the higher the frequency, the greater the difference between reduced and unreduced tokens. For nonce words, the reduced token is also perceived as more acceptable than the unreduced token, but this difference is ranked somewhere between the medium/low frequency words /mimashita/ and /nimashita/. These differences were assessed in a series of paired sample t-tests (Table 2) and real word differences are presented in Figure 1.

Table 2: GoF scores, GoF differences between tokens that share a category and paired-sample t-test results.

Token	GoF	Category	GoF Diff.	t	Sig.
[ʃimaʃita]	5.49	/ʃimaʃita/	0.95	2.689	.018*
[ʃimaʃta]	6.44				
[kimaʃita]	5.71	/kimaʃita/	0.84	2.347	.034*
[kimaʃta]	6.55				
[mimaʃita]	5.72	/mimaʃita/	0.75	2.133	.051
[mimaʃta]	6.48				
[nimaʃita]	5.61	/nimaʃita/	0.09	0.242	.813
[nimaʃta]	5.7				
[pimaʃita]	5.65	/pimaʃita/	0.68	2.159	.049*
[pimaʃta]	6.33				
[simaʃita]	4.12	/ʃimaʃita/	0.42	1.615	.129
[simaʃta]	4.54				

Figure 1: Experiment 1 GoF differences in real word tokens. IPA transcriptions of the categories appear horizontally along the bottom of the graph, average GoF scores are represented vertically.



An ANOVA was performed to test our hypothesis that predictability influences reduced token GoF scores. The ANOVA—which was calculated on each participants' average GoF score for reduced tokens—revealed a significant effect of contrast on GoF score ($F(3, 75) = 5.15$, $p = .003$). A second ANOVA that was performed on the average GoF scores for unreduced tokens did not reveal a significant effect of predictability on unreduced tokens ($F(3, 75) = .087$, $p = .967$). The unreduced [simaʃita] token GoF scores do appear to be an outlier here, however, previous studies have shown that Japanese listeners assign low GoF scores to [si] sequences that are assigned to /ʃi/ categories [3]. These findings suggest that unreduced sequences are perceived as equally acceptable regardless of predictability while reduced sequences are perceived as more acceptable when they occur at a higher frequency.

Results for Experiment 2 are presented in Table 3. An ANOVA on discrimination accuracy did not reveal a significant difference between contrasts ($F(5, 107) = 0.43$, $p = .83$); however a second ANOVA on response time showed a significant effect ($F(5, 5183) = 5.25$, $p = < .001$).

Table 3: Experiment 2 AXB Accuracy and Response Times. RT (ms) is the average response time in milliseconds.

Contrast	Accuracy %	RT (ms)
[ʃimaʃita] - [ʃimaʃta]	91	1097
[kimaʃita] - [kimaʃta]	93	1031
[mimaʃita] - [mimaʃta]	92	1046
[nimaʃita] - [nimaʃta]	91	1045
[pimaʃita] - [pimaʃta]	91	1065
[simaʃita] - [simaʃta]	91	1032

Response time in experiments similar to the one reported on here is typically seen as a reflection of task difficulty (e.g., [8]). A post hoc test with Bonferroni correction was conducted to provide a more fine-grained measure of task difficulty. The test revealed a significant difference of response time only between the [ʃimaʃita] - [ʃimaʃta] contrast and the following contrasts: [kimaʃita] - [kimaʃta] ($p < .001$), [nimaʃita] - [nimaʃta] ($p = 0.01$), [simaʃita] - [simaʃta] ($p < .001$).

4. CONCLUSION

The current experiments examined the effect of predictability on speech perception by exposing participants to sequences of phones that vary in their production and predictability. The results demonstrate that listeners are tuned to the influence of predictability on reduction whereby the acceptability of words with reduced sounds appear to correlate with increased frequency. While unreduced tokens maintain reasonably stable GoF scores regardless of predictability, GoF scores assigned to reduced tokens vary significantly depending on how often they occur in spontaneous Japanese discourse. This effect was surprisingly strong with the GoF scores of the most frequently occuring sequences—[ʃimaʃita] (5.49) and [ʃimaʃta] (6.44)—showing almost a full point difference along a 7-point scale. This contrasts with the much less frequent sequences—[nimaʃita] (5.61) and [nimaʃta] (5.7)—which were not significantly different. Interestingly, nonce words—sequences which listeners are presumably never or very rarely exposed to—performed like contrasts of medium to low level predictability. This is an interesting finding that deserves further investigation: it may be due to listeners sometimes parsing nonce words at the level of /maʃita/ or /ʃita/, which would explain why they behave as tokens that are expected.

The current experiments follow a well-established experimental paradigm whereby the GoF scores assigned to sequences of phonotactically legal or illegal phones that share a category are a useful predictor of their discriminability. The GoF scores in Experiment 1 suggest that sequences that occur more frequently should be easily discriminable; however, this was not reflected in Experiment 2. Participants were not significantly more accurate at discriminating between high and low predictability contrasts. This may be the result of the effect of predictability on attention to phonetic detail [4]—which should reduce discriminability of predictable contrasts—cancelling out the effect of divergent GoF. Here, we hypothesise that if stimuli were manipulated to be more similar (e.g., by reducing the duration or intensity of target vowels in unreduced tokens), the influence of predictability would decrease in terms of GoF score divergence and increase in terms of discrimination accuracy.

5. ACKNOWLEDGEMENTS

We would like to thank our participants.

6. REFERENCES

[1] Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of experimental psychology: human perception and performance, 25*(6), 1568.

[2] Kawahara, S., & Garbey, K. 2014. Nasal Place Assimilation and the Perceptability of Place Contrasts. *Open Lingusitics 1*(1), 17-36.

[3] Kilpatrick, A. J., Bundgaard-Nielsen, R. L., & Baker, B. J. (2019). Japanese co-occurrence restrictions influence second language perception. *Applied Psycholinguistics, 40*(2), 585-611.

[4] Kilpatrick, A. J., Kawahara, S., Bundgaard-Nielsen, R. L., Baker, B. J., & Fletcher, J. 2018. Japanese Listeners Are More Likely to Perceive Illusory Vowels in Predictable Contexts. Poster presented at: *The 6th Annual Meeting on Phonology*.

[5] Kilpatrick, A. J., Kawahara, S., Bundgaard-Nielsen, R. L., Baker, B. J., & Fletcher, J. 2018. Japanese vowel devoicing modulates perceptual epenthesis, *Proceedings of the 17th Australasian International Conference on Speech Science and Technology*.

[6] Kilpatrick, A. J., Kawahara, S., Bundgaard-Nielsen, R. L., Baker, B. J., & Fletcher, J. Japanese Perceptual Epenthesis is Modulated by Transitional Probability. *Language and Speech*. Under review.

[7] Maekawa, K. 2003. Corpus of Spontaneous Japanese: Its design and evaluation, *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition.*

[8] Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and speech, 38*(3), 289-306.

[9] Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological review, 115*(2), 357.

[10] Shaw, J., & Kawahara, S. 2018. Predictability and phonology: Past, present & future, *Linguistics Vanguard 4*(S2).

[11] Tsujimura, N. 2013. *An introduction to Japanese linguistics*, 3rd ed., John Wiley and Sons: West Sussex, 2013.

[12] Zipf, G. K. 1949. *Human behavior and the principle of least effort: an introduction to human ecology*, New York: Hafner Publisher Company.