Japanese Listeners Are More Likely to Perceive Illusory Vowels in Predictable Contexts

Alexander Kilpatrick[1] Shigeto Kawahara[2] Rikke Bundgaard-Nielsen[34] Brett Baker[1] Janet Fletcher[1]

[1]University of Melbourne [2]Keio University [3]Western Sydney University [4]MARCS Institute for Brain, Behaviour and Development

INTRODUCTION: Studies examining the role of predictability on linguistic behaviour show that speakers typically produce speech less robustly (shorter, quieter and/or with more centralized vowels) in highly predictable contexts (see e.g. Hall et al., 2018; Jaeger & Buz, 2016; Shaw & Kawahara, 2017). We propose a perceptual corollary to this behaviour and demonstrate that listeners are less attentive to incoming acoustic signals when they are more predictable. We present an AXB discrimination experiment testing the ability of Japanese listeners to discriminate between real word and nonsense tokens that adhere to Japanese phonotactics and phonologically similar tokens that contain non-homorganic consonant clusters. Japanese does not allow non-homorganic consonant clusters, and when listeners are exposed to such sequences, they have been reported to perceive an illusory vowel (Dupoux et al., 1999 *et seq*.). Extending the Perceptual Assimilation model (Best, 1994) which predicts and accounts for the influence of L1 transitional probability on L2 perception, Kilpatrick et al., (in press) propose that illicit non-homorganic consonant clusters are assimilated to their perceptually nearest and most transitionally probable match. This perceptual assimilation reduces, or even obliterates, the difference between the cluster and the licit sequence, resulting in the illusory vowel effect. The current experiment tested the influence of word length on perceptual epenthesis in real and nonce word tokens, allowing us to assess the influence of predictability because listeners have not previously encountered nonce words and are therefore unable to predict upcoming input (at least not directly). We predicted that the longer the word, the more predictable the following segments are for real words (cf., Piantadosi et al., 2011). The findings show that word length has an influence on the discriminability of all AXB contrasts but that this effect is greater in contrasts that contain real words. These results suggest that listeners attend less to predictable sequences, making use of top-down information based on L1 transitional probability, instead of attending to raw acoustic information.

METHOD: The stimuli are shown in Tables 1 and 2. Within each pair, one token contained a consonant cluster, whereas in the other token, the cluster was split by a high vowel. Token length—measured in mora—were varied. The stimuli were produced by a female Australian English speaking phonetician. Recordings were taken at a recording studio at the University of Melbourne with a bit depth of 64kb/sec and a sampling rate of 48kHz. ABX tokens were spaced with an inter-stimulus interval of 1000 ms. 20 (17 Female) L1 Japanese speakers living in Melbourne, Australia, (LOR two weeks-four months: M = 2.7 months, SD = .7 months) participated. Participants had not previously travelled outside of Japan for a period greater than 1 month. The experiment was conducted in a quiet room located at the University of Melbourne. Tokens were split into three blocks, each block consisted of 192 AXB discrimination trials, for a total of 578 trials (48 per contrast).

RESULTS: The results show a very clear influence of word length on discriminability and an even greater influence if licit tokens represent extant Japanese words. This is most evident in the differences between long and short contrasts in real and nonce word trials. For example, participants were able to successfully discriminate between both /**suk**i/-/**sk**i/ and /**sup**i/-/**sp**i/ contrasts 98% of the time but were far less accurate at discriminating between the real word contrast /dai**suk**i/-/dai**sk**i/ (88%) compared to the nonce word contrast /gai**sup**i/-/gai**sp**i/ (96%). A MANOVA examined token length (mora count) and status (real/nonce) as independent variables and discrimination accuracy as a dependent variable. It showed a significant univariate effect of token length on discrimination accuracy (p = < 0.001) and an insignificant

univariate effect of status on accuracy (p = 0.946). Importantly, the MANOVA revealed a significant multivariate interaction between token length and status (p = 0.018). A Bonferroni correction calculated on token length with accuracy as the dependent variable revealed a significant difference between all length contrasts (p = < 0.002) except for 5 and 7 mora contrasts (p = 1).

Real Word Contrasts	Gloss	Mora	Mean Acc.	St. Dev.
/tabesugima**ʃi**ta/-/tabesugima**ʃ**ta/	"over ate"	7	80%	10%
/tabema**ʃi**ta/-/tabema**ʃ**ta/	"ate"	5	81%	11%
/ma**ʃi**ta/-/ma**ʃ**ta/	"directly under"	3	83%	9%
/**ʃi**ta/-/**ʃ**ta/	"under"	2	93%	7%
/dai**suk**i/-/dai**sk**i/	"really like"	3	88%	7%
/**suk**i/-/**sk**i/	"like"	2	98%	3%

Table 1. Accuracy results, gloss and mora count of real word contrasts.

Nonce Word Contrasts	Mora	Mean Acc.	St. Dev.
/padezukina**ʃip**a/-/padezukina**ʃp**a/	7	80%	14%
/padena**ʃip**a/-/padena**ʃp**a/	5	77%	15%
/na**ʃip**a/-/na**ʃp**a/	3	87%	8%
/**ʃip**a/-/**ʃp**a/	2	88%	7%
/gai**sup**i/-/gai**sp**i/	3	96%	5%
/**sup**i/-/**sp**i/	2	98%	3%

Table 2. Accuracy results and mora count of nonce word contrasts.

CONCLUSION: Token length has a clear influence on the accuracy of all trials. This is likely due to limitations of working memory whereby longer tokens place a greater demand on cognitive resources, making listeners more error prone. Importantly, the results show that token length effect is greater with trials that contain real words. Unlike nonce words, when real words get longer, the identity of the incoming signal becomes more predictable due to top down knowledge. This increase in predictability results in listeners assigning less attentional resources to the incoming acoustic information so that illicit sequences are more likely to be assimilated to their most transitionally probable match. This assimilation reduces or eliminates the perceptual distance between the contrast pairs, resulting in reduced discrimination accuracy.

REFERENCES

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words, 167*(224), 233-277.

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.

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

Jaeger, T. F., & Buz, E. (2017). Signal reduction and linguistic encoding. In E. M. Fernandez & H. S. Cairns (Eds.), *The handbook of psycholinguistics* (pp. 38-81). Chichester, UK: Wiley-Blackwell

Kilpatrick, A., Bundgaard-Nielsen, R., & Baker, B. (In Press). Japanese Co-occurrence Restrictions Influence Second Language Perception. *Applied Psycholinguistics*.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences, 108*(9), 3526-3529.

Shaw, J. A., & Kawahara, S. (2017). Effects of surprisal and entropy on vowel duration in Japanese. *Language and speech*, 0023830917737331.