The role of predictability in shaping human sound systems

LabPhon 16 satellite workshop

June 23, 2018 • Universidade de Lisboa
http://labphon16.labphon.org/se-05.html

Schedule

9:00am	**Workshop introduction**	
9:05am	*The indirect effect of low information content on lenition processes*	U. Cohen-Priva & E. Gleason
9:25am	*Emergent and categorical differences between English nouns and verbs*	C. Cohen & M. Carlson
9:45am	*Effects of social message predictability on probabilistic reduction*	D. Hashimoto
10:05am	**Break**	
10:15am	*Phonological grammars evolve to preserve information at word beginnings*	A. Wedel & A. Ussishkin
10:35am	*Predictability and boundary strength in English compound nouns*	M. Bell, S. Ben Hadia, & I. Plag
10:55am	*Coordinative mode, phasing, and predictability across languages*	K. Franich
11:15am	*Japanese perceptual epenthesis is modulated by transitional probability*	A. Kilpatrick, S. Kawahara, R. Bundgaard-Nielsen, B. Baker, & J. Fletcher
11:35am	**Break**	
11:45am	**Roundtable introduction**	
11:55am	**Roundtable discussion:** Defining and measuring predictability	All participants
12:30pm	Workshop ends	

Organizers: Scott Seyfarth (seyfarth.2@osu.edu) and Beth Hume (hume.3@osu.edu)

The indirect effect of low information content on lenition processes

Uriel Cohen Priva and Emily Gleason
Brown University
uriel_cohen_priva@brown.edu, emily_gleason@alumni.brown.edu

1 Overview

Consonant lenition refers to a several phonological processes than tend to occur in similar environments (e.g. intervocalically), and under similar conditions (e.g. in faster speech). These typically include degemination, voicing, spirantization, approximantization, tapping, debuccalization, and deletion. Different theories argue for different causal mechanisms that lead to lenition (Kirchner, 1998; Bauer, 2008; Kingston, 2008).

There is growing evidence that low information content (high contextual predictability, low informativity, few minimal pairs) leads to phonetic and phonological reduction (Aylett and Turk, 2004; Levy and Jaeger, 2007; Jaeger, 2010; van Son and Pols, 2003; Wedel et al., 2018). The interpretation of this relationship in current research tends to assume a direct causal route, exemplified in the 2nd theme of the workshop:

1. Low information content causes lower tolerance of articulatory effort or higher tolerance for imprecise production.
2. These conditions make the production of reduced articulatory targets possible or desirable, unlike high information content conditions.

That is, reduction due to low information is a form of optimization. We argue that when lenition is triggered by low information it deviates from the predictions a simple optimization account would predict. Instead, we show that a more likely causal route is that low information content leads to durational reduction, which as a byproduct results in the emergence of lenited forms. We make this hypothesis based on the following reasoning:

1. Lenited forms often require more careful articulation (Kingston, 2008). Imprecise articulation could result in e.g. occlusion, a fortition process.
2. Fast speech results in greater lenition rates (Cohen Priva, 2015), though everything else being equal, fast speech would involve *higher* information rates (cf. Cohen Priva, 2017a).
3. Lenited forms are often *more* informative cross-linguistically (Study 1).
4. Lenition processes involve reduction in duration (Study 2), while processes typically seen as "gaps" in lenition typology do not (Study 3).

2 Studies

2.1 Study 1: Cross linguistic informativity of spirantization and voicing

Cohen Priva (2017b) shows that low informativity correlates with the actuation of lenition. However, if lenition is the result of an information optimization pressure, lenition processes should result in segments that *typically* have lower information. Following Cohen Priva (2017b), we calculated informativity profiles for 46 languages using word frequency counts from the Crúbadán corpus (Scannell, 2007). Phonemic representations of words were deduced from alphabets.

We compared the informativity of same-place segments, contrasting stops with fricatives and voiceless obstruents with voiced obstruents (/t/, /d/, /s/, /z/, /p/, /b/, /f/, /v/, /k/, /g/). If lenition is an optimizing process, the voiced obstruents and fricatives should have **lower informativity**. However, almost all the comparisons resulted in fricatives and voiced obstruents having **higher informativity** than same-place stops and voiceless obstruents (only the informativity of /v/ was higher than /f/'s more often than not). Voicing and spirantization are therefore not optimizing from an information-based perspective.

2.2 Study 2: Are lenition processes duration-reducing?

Information-theoretic accounts have consistently shown that duration-reduction correlates with low information (Bell et al., 2009; Seyfarth, 2014). Reduced duration, in turn, has been argued to cause the perception of voicing and spirantization, as well as passive voicing (Lavoie, 2001, ch. 6, and references herein). Lavoie (2001) and Katz (2016) argue that lenition processes entail reduced duration or increased intensity. This study checks whether lenition processes in American English correspond to reduced duration.

We compared the duration of the surface forms of underlying and output segments for all processes affecting more than 5% of the surface outputs of a given segment in the Buckeye corpus (Pitt et al., 2007). For each process we used a mixed effect linear regression to determine whether the binary distinction (e.g. surface [t] and surface [ʔ]) predicts a change in duration. Surface forms were compared because some forms do not occur underlyingly ([ɾ], [ʔ]). Random intercepts included word, speaker, and the phonological environment (where applicable). Fixed controls included speech rate, word frequency, and distance from both word edges. For lenition processes this comparison always significantly entailed **reduced duration** (all $p<.01$).

The results are exemplified in Figure 1. Negative values mean the duration of the surface form was shorter than the reference form. Error bars are two standard errors from the mean. Color and shape indicate the process type.

2.3 Study 3: Lenition gaps

If lenition is triggered by reduced duration, and only indirectly by low information, then perhaps known gaps in lenition typology should be attributed to the durational factors. We considered three gaps in lenition typology: /ʔ/-spirantization (Hock, 1986), spirantization to stridents (Kirchner, 1998),

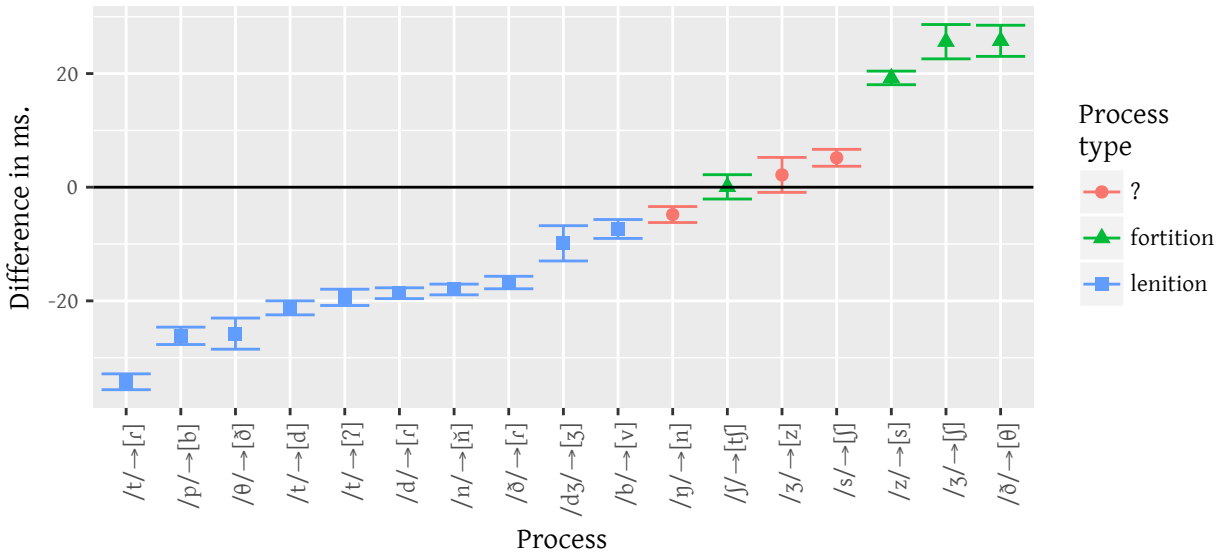


Figure 1: Comparison between reference and surface duration

and nasalization (Kingston, 2008). We used the method described in Study 2, but applied it to the missing processes (e.g. contrasting surface [h] and [ʔ]). Indeed, both /ʔ/-spirantization and spirantization to stridents would have resulted in durational *lengthening* (all $p<10^{-15}$). Voiced stops (/d/ and /b/) were not durationally longer than same-place nasals ($p>.05$), and /g/ was only 5ms longer than /ŋ/, explaining why lenition to nasals is unlikely for voiced stops. Voiceless stops arguably lenite to voiced stops and not nasals due to perceptual similarity (as in P-Map, Steriade, 2001).

We also attempted to see whether a direct causal explanation would have predicted the gaps as well (using informativity comparison as in Study 1). Though stridents typically have higher informativity then coronal stops, nasals typically had *lower* informativity than same-place voiceless and voiced stops. An optimization account could therefore wrongly predict that nasalization is likely to affect lower-information segments.

3 Summary

There are multiple reasons to doubt a direct causal path from resource reduction to lenition. Study 1 demonstrates that lenition does not behave as if it follows from an information-reduction optimization process. Study 2 shows, however, that lenition does correlate with duration-reduction in every case of lenition in the Buckeye corpus. Duration-reduction also predicts gaps in lenition typology (Study 3). Duration-reduction has been repeatedly demonstrated to follow from low information, suggesting that an indirect causal path through duration reduction, rather than articulatory effort / resource reduction, is a more likely explanation for the relationship between low information and lenition.

References

Aylett, Matthew and Turk, Alice. 2004. The smooth signal redundancy hypothesis. *Language and Speech*, 47(1):31–56.

Bauer, Laurie. 2008. Lenition revisited. *Journal of Linguistics*, 44(3):605–624.

Bell, Alan, Brenier, Jason, Gregory, Michelle, Girand, Cynthia, and Jurafsky, Daniel. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1):92–111.

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

Cohen Priva, Uriel. 2017a. Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition*, 160:27–34.

Cohen Priva, Uriel. 2017b. Informativity and the actuation of lenition. *Language*, 93(3):569–597.

Hock, Hans Henrich. 1986. *Principles of historical linguistics*. Mouton de Gruyter, Berlin; New York.

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1):23–62.

Katz, Jonah. 2016. Lenition, perception and neutralisation. *Phonology*, 33(1):43–85.

Kingston, John. 2008. Lenition. In *Selected proceedings of the 3rd conference on laboratory approaches to Spanish phonology*, pages 1–31. Cascadilla Press Somerville, MA.

Kirchner, Robert Martin. 1998. *An Effort-Based Approach to Consonant Lenition*. PhD thesis, UCLA.

Lavoie, Lisa M. 2001. *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. Psychology Press.

Levy, Roger and Jaeger, T. Florian. 2007. Speakers optimize information density through syntactic reduction. In Scholkopf, Bernhard, Platt, John, and Hofmann, Thomas, editors, *Advances in Neural Information Processing Systems (NIPS)*, volume 19, pages 849–856, Cambridge, MA. MIT Press.

Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E., 2007. Buckeye corpus of conversational speech (2nd release). Department of Psychology, Ohio State University.

Scannell, Kevin P. 2007. The Crúbadán project: Corpus building for under-resourced languages. In *Building and Exploring Web Corpora: Proceedings of the 3rd Web as Corpus Workshop*, volume 4, pages 5–15.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1):140–155.

Steriade, Donca. 2001. The phonology of perceptibility effects: the P-map and its consequences for constraint organization.

Son, R. J. J. H. van and Pols, L. C. W. 2003. How efficient is speech? *Proceedings of the Institute of Phonetic Sciences*, 25:171–184.

Wedel, Andrew, Nelson, Noah, and Sharp, Rebecca. 2018. The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, 100:61–88.

Emergent and categorical differences between English nouns and verb

Clara Cohen[1,2], Matthew Carlson[2]
[1] *University of Glasgow,* [2] *The Pennsylvania State University*

Introduction Words can be predictable for many reasons. If a word is likely to be used in a particular utterance context, it has high *contextual* probability (e.g., Jurafsky et al., 2001; Bell et al., 2009). If it is likely to be selected from a set of related words, such as a morphological paradigm, it has high *paradigmatic* probability (Cohen, 2014, 2015; Kuperman et al., 2007). A third measure, inflectional entropy, describes not the word form itself, but its paradigm. Entropy captures how much uncertainty is associated with the inflectional paradigm as a whole (Tabak et al., 2005, 2010), and increases when inflectional paradigms are large, or when usage frequencies are evently distributed across the paradigm. All of these measures can interact with each other in words containing agreement morphology: The agreement inflection is governed by the surrounding utterance (contextual probability), while the use of different inflectional forms reflects the morphological paradigm of the word in question (paradigmatic probability, inflectional entropy). Here, we ask how the pronunciation of unsuffixed and suffixed English nouns and verbs reflects these measures of probability, and what that can reveal about mechanisms of lexical retrieval during spontaneous speech.

We focus in particular on the *-s* suffix, because it has identical morphophonology in nouns and verbs, and yet is distinct in every other way. On present-tense verbs (e..g, *run-s*, *think-s*), *-s* encodes an obligatory agreement relationship with a third-person singular subject, and hence has an extremely high contextual probability. On plural nouns (e.g., *dog-s*, *book-s*), however, *s* reflects an inherent number feature. It can induce agreement on dependents of the noun, such as determiners (e.g., *these, those*), but it is not dictated by anything other than the speaker's intended meaning. It therefore has much lower contextual probability. With paradigmatic probability, however, the situation is reversed: the *-s*-suffixed verb is one of as many as five forms in the inflectional paradigm, with extremely restricted usage; its paradigmatic probability therefore tends to be low. With nouns, the suffixed form is one of only two forms in the paradigm, and therefore tends to have higher paradigmatic probability. More generally, entropy in verbal paradigms tends to be higher than in noun paradigms, because verbal paradigms are larger. If the relationship between pronunciation and probability is similar for both nouns and verbs, then that would imply that mechanisms of lexical retrieval and articulatory encoding are sensitive to probability, but do not make any other fundamental distinctions between lexical category. In this case, some categorical differences between nouns and verbs may well be reducible to differences in their probabilistic relationships to other words. If, on the other hand, nouns and verbs differ in how their pronunciation reflects probability, that would imply that distinctions such as lexical category go beyond such probabilistic relations.

Methods We extracted from the Buckeye Corpus of Conversational English (Pitt et al., 2007) all nouns and verbs that carried either the suffix *-s* (e.g., *dog-s, run-s*) or appeared in their unsuffixed form (*dog, run*). In three separate analyses, we explored how stem duration and stressed vowel dispersion reflected contextual probability, paradigmatic probability, and inflectional entropy. Because the Buckeye corpus is not syntactically parsed, it was not possible to characterize contextual probability in terms of syntactic structure.

We therefore approximated contextual probability with the conditional probability of the verb or noun, given the preceding word. Paradigmatic probability was calculated as the log-tranformed frequency ratio of the target word to the summed frequencies of all non-target words in the inflectional paradigm—in other words, the log-odds of selecting the target word form from that word's inflectional paradigm. Inflectional entropy was calculated as the negative sum of the log-transformed relative frequency of each member of the inflectional paradigm, where each member was weighted by its relative frequency.

Stem duration and vowel dispersion were analyzed with mixed effects linear regression modeling, with random intercepts for speaker and word, and random slopes for key probabilistic predictors. In the first stage of model building, we added control predictors such as part of speech (noun vs. verb), proximity to pauses, lexical frequency, phonological neighborhood density, word length (in segments), phonotactic probability, surrounding phonetic features. After the best-fitting baseline model was determined, the key probabilistic predictors of contextual probability, paradigmatic probability, and entropy were included, and—crucially—allowed to interact with part of speech.

Results Contextual probability affected verbs more strongly than nouns, but for both higher contextual probability yielded shorter duration. Vowel dispersion also reacted similarly, with both nouns and verbs showing equivalent centralization as entropy increased. However, paradigmatic probability and inflectional entropy affected noun and verb stem duration quite differently. Increased paradigmatic probability lengthened noun stems ($\beta = 0.029,\ SE(\beta) = 0.007,\ p < .001$), but shortened verb stems, with the differences most distinct when contextual probability was high ($\beta = -0.121,\ SE(\beta) = 0.013, p < .001$, see Figure 1). On the other hand, the effects were reversed for inflectional entropy, which shortened noun stems ($\beta = -0.026,\ SE(\beta) = 0.013,\ p < .05$) and lengthened verb stems ($\beta = 0.101,\ SE(\beta) = 0.028,\ < .001$, see Figure 2).[1]

Conclusion Nouns and verbs showed similar effects of contextual probability, but reacted distinctly to paradigmatic probability and entropy. These results show categorical distinctions in how lexical retrieval and articulation reflect probability for nouns and verbs—but, crucially, only with respect to paradigm structure. Although pronunciation reflects the surrounding syntactic structure similarly for nouns and verbs, different paradigmatic effects on noun and verb pronunciation cannot be reduced to differently organized inflectional paradigms.

References

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1):92–111.

Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement. *Morphology*, 24(4):291–323.

[1]Note that these apparently reversed effects do not reflect negative correlations between entropy and paradigmatic probability: Suffixed forms have low paradigmatic probability and unsuffixed forms have high paradigmatic probability, yet within a paradigm they both have identical entropy measures. Yet there was no interaction with suffixation in these effects.

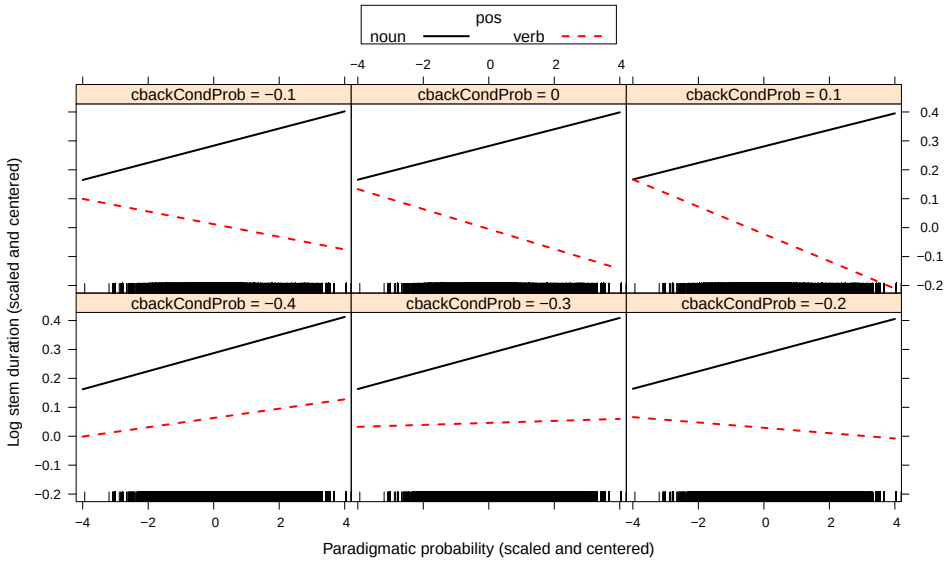


Figure 1: Higher paradigmatic probability lengthened noun stems and shortened verb stems, with strongest effects when contextual probability was high.

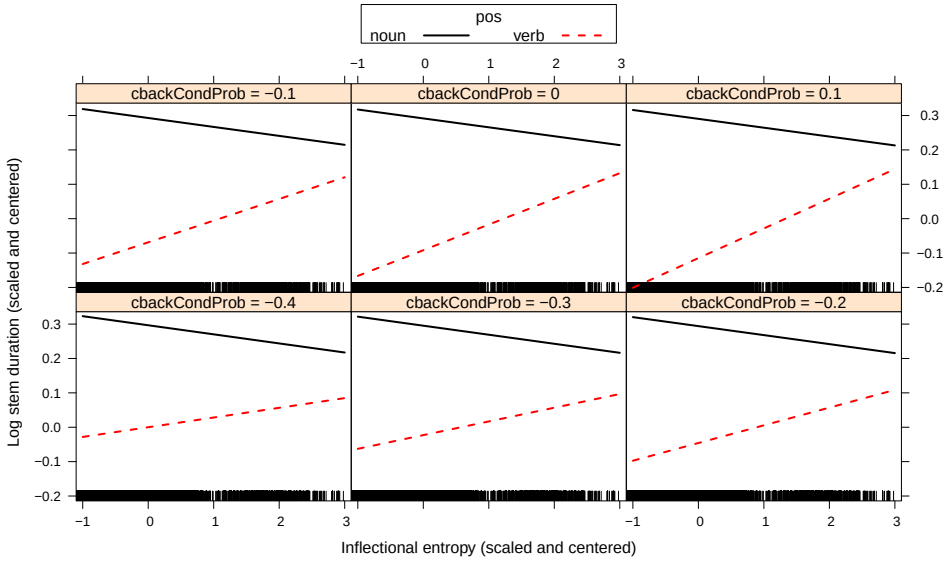


Figure 2: Higher inflectional entropy shortened noun stems and lengthened verb stems, with stronger effects when contextual probability was high.

Cohen, C. (2015). Context and paradigms: Two patterns of probabilistic pronunciation variation in Russian agreement suffixes. *The Mental Lexicon*, 10(3):313 – 338.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. *Frequency and the emergence of linguistic structure*, pages 229–254.

Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America*, 121(4):2261–2271.

Pitt, M. A., Dilley, L. C., Johnson, K., Kiesling, S., Raymond, W. D., Hume, E., and Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech (2nd release).

Tabak, W., Schreuder, R., and Baayen, R. H. (2005). Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in Dutch. *Linguistic evidenceEmpirical, theoretical, and computational perspectives*, (1993):529–555.

Tabak, W., Schreuder, R., and Baayen, R. H. (2010). Producing inflected verbs: A picture naming study. *The Mental Lexicon*, 5(1):22–46.

Effect of Social Message Predictability on Probabilistic Reduction

Daiki Hashimoto

University of Canterbury

The effect of predictability on linguistic signals has been studied at various levels of linguistic structure, and it has been reported that messages tend to be realized with reduced signals when they are more predictable given the context (see Jaeger and Buz 2017). In order to account for these probabilistic reduction phenomena, Message-Oriented Phonology (MOP) hypothesizes that speakers use "the right sort of redundancy" to avoid "inefficient redundancy," and thereby augment "the likelihood of sufficiently accurate and cost-effective message transmission" (Hall et al. unpublished). Namely, this approach hypothesizes that effective communication involves balancing resource cost and message transmission accuracy, and the trade-off between the two forces shapes sound patterns. Consequently, redundancy should be added in a message that is important to successful communication, while less redundancy should be invested in a message that is not important. The importance of messages is determined by information content. Information content is measured by taking the negative log of the predictability of the intended message (Shannon 1948), i.e., a message with lower predictability has higher information content, whereas that with higher predictability has lower information content. In this way, MOP can neatly predict that phonological units are produced with reduced signals for contextually predictable messages.

Linguistic communication involves a variety of message transmissions including lexical and social messages. Lexical message refers to the sense of a word as it appears in a dictionary, for example, a nominal word *penguin* refers to a certain set of animals. Social message is group-associational, and it is expressed by employing the same linguistic variant as employed by a social group (Bell 2001). For instance, a speaker may associate herself to a social group of middle-class by realizing /θ/ as [t] in New York City (Labov 1972).

Although previous literature has demonstrated that the predictability of lexical messages influences the degree to which their signals are reduced, less is known about the effect of predictability of social messages. The aim of this paper is to explore the effect of social message predictability on linguistic signals. The main research question is "Is the production of a linguistic variant influenced by social message predictability?" By addressing this question, we can also address an abstract research question: "Is the predictability of a social message represented in our mind?"

The loanword phonology in New Zealand English (NZE) provides an interesting test case for our research question, because the realization of /r/ is considered to carry a social message and the selection of the variants is probabilistic in accordance with speakers and loanwords. NZE borrows a large number of loanwords from te reo Māori, which has a tap sound [ɾ] as a rhotic phoneme. NZE speakers sometimes adapt the nonnative rhotic to the native rhotic [r] (e.g., *ko*[r]*u* and *ma*[r]*ae*), and sometimes import the nonnative rhotic [ɾ] without modification (e.g., *ko*[ɾ]*u* and *ma*[ɾ]*ae*). Crucially, we assume that the nonnative rhotic [ɾ] is a socially meaningful variant and carries a social message associated with Māori, because it is more likely to be produced by a NZE speaker strongly associated with Māori and in speech strongly

associated with Māori (Hashimoto 2018). My previous work has shown that the importation rate differs in accordance with loanwords and speakers. The following table shows adaptation and importation rates for some example words and speakers:

p(r|loanword) and *p*(ɾ|loanword)			*p*(r|speaker) and *p*(ɾ|speaker)		
Loanword	Rate of [r]	Rate of [ɾ]	Speaker	Rate of [r]	Rate of [ɾ]
Timaru	72%	28%	*Speaker6*	51%	49%
koru	22%	78%	*Speaker10*	16%	84%

In this data, speakers are more likely to produce tap sounds in *koru* than *Timaru*, i.e., social messages associated with Māori are more predictable given *koru* than *Timaru*.

Previous studies have demonstrated that a linguistic unit with higher lexical message predictability is more likely to be reduced. If the predictability effect can be extended to the social message predictability, our prediction would be that, when social messages associated with Māori are contextually predictable (i.e., production of nonnative rhotics is predictable), the linguistic signals (i.e., nonnative rhotics) should be more likely to be reduced or shorter. The social message predictability is defined as the predictability of the selection of the imported structure [ɾ] against the adapted structure [r], and the context is a word or a speaker. For example, the importation predictability given a loanword *Timaru* *p*(ɾ|Timaru) is 28%, and that given a speaker *Speaker6* *p*(ɾ|Speaker6) is 49% in the above table.

In order to test this prediction, 32 non-bilingual NZE speakers were asked to pronounce 36 target Māori loanwords with /r/ sounds plus 84 filler words within a carrier sentence. The stimuli were presented in random order, and the same task was performed twice. After the experiment, the realizations of /r/ and durations of [ɾ] were determined on the basis of lowered F3 and clear consonantal edges (Fig.1). After the annotation, two types of the probability of [ɾ] were calculated given a loanword and a speaker, *p*(ɾ|loanword) and *p*(ɾ|speaker), and they were transformed into information content, $-\log_2 p$(ɾ|loanword) and $-\log_2 p$(ɾ|speaker).

A mixed-effects linear regression analysis was fit on the data, with log-transformed duration of [ɾ] as the response variable. The model was selected using backwards elimination, and it included random intercepts for speakers and words, and random slopes for all the variables showing significant effects. The predictors included main-stress, word-initiality, and centred information content of [ɾ] given a loanword (i.e., $-\log_2 p$(ɾ|loanword)). It was found that $-\log_2 p$(ɾ|loanword) is significant in the predicted direction, whereas $-\log_2 p$(ɾ|speaker) is not significant. That is, a nonnative rhotic sound is produced with shorter duration, when the selection of the nonnative rhotic against the native rhotic is more predictable (i.e., social messages associated with Māori are more predictable) given a loanword (Fig.2).

These results suggest that the predictability of a social message partially affects the duration of a linguistic variant carrying the social message, and the probabilistic reduction is not limited to the predictability of lexical messages. This finding also indicates that the predictability of social messages is represented in our mind.

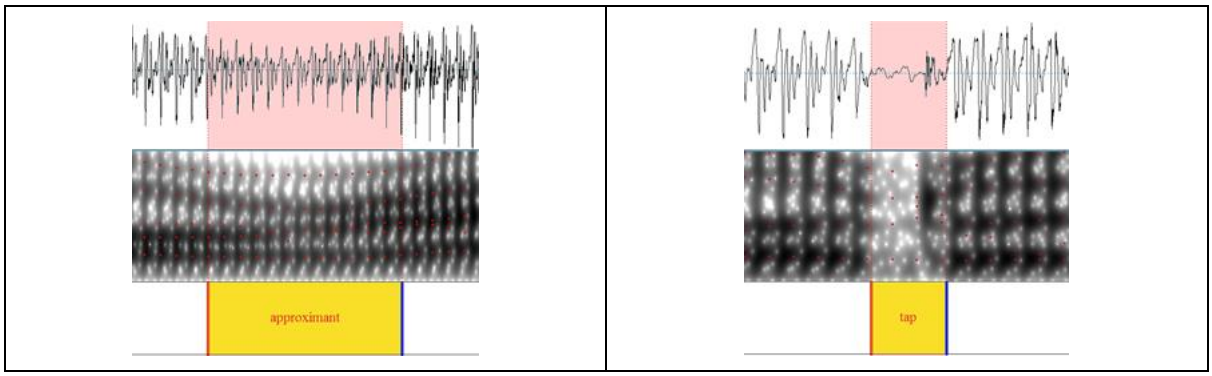


Figure 1. *Classification of /r/ realizations in Māori loanwords into adapted structure [r] and imported structure [ɾ]*

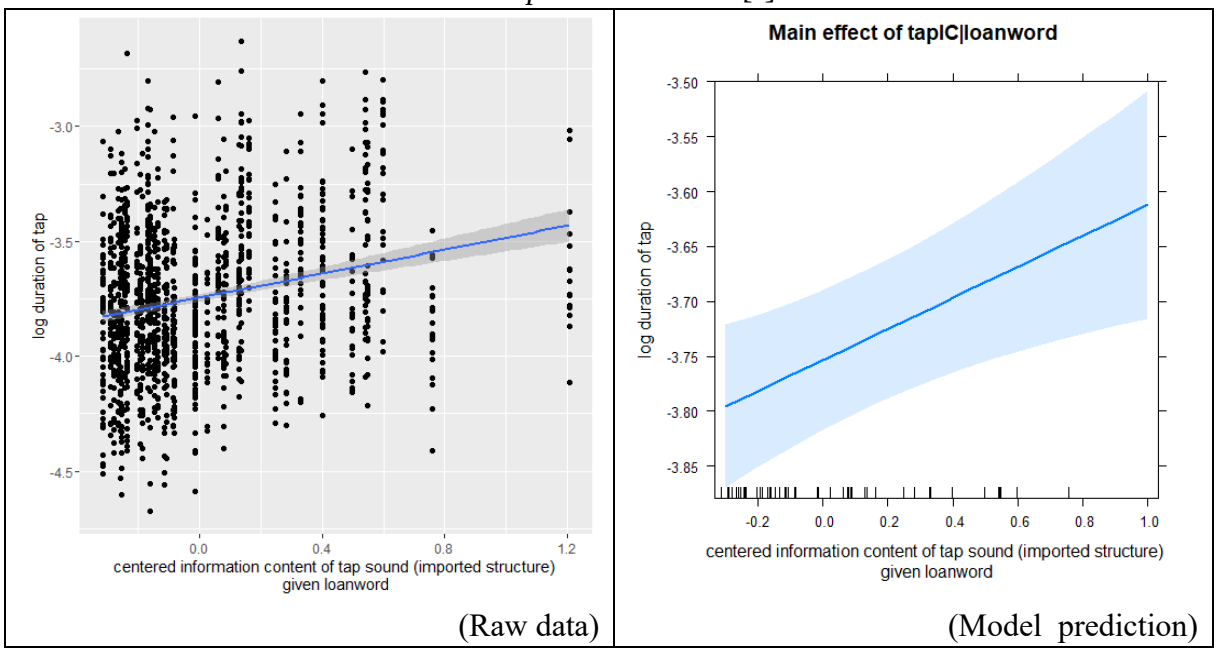


Figure 2. *Correlation between duration of tap sound and predictability of tap sound given loanword (i.e. predictability of social message associated with Māori given loanword)*

Bell, Allan. 2001. Back in Style: Reworking Audience Design. In: Penelope Eckert and John Rickford (eds.) *Style and Sociolinguistic Style*. 136–169. Cambridge: Cambridge University Press.

Hall, Kathleen Currie, Elizabeth Hume, T. Florian Jaeger and Andrew Wedel. unpublished. The Message Shapes Phonology.

Hashimoto, Daiki. 2018. Topic-oriented variation in loanword phonology. Paper presented at LabPhon 16 (Universidade de Lisboa, Portugal).

Jaeger, T. Florian, and Esteban Buz. 2017. Signal Reduction and Linguistic Encoding. M. Fernández and Helen Smith Cairns (eds.) *The Handbook of Psycholinguistics*. 38–81. New Jersey: Wiley-Blackwell.

Labov, Williams. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Shannon, Claude E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27. 329–423.

Phonological grammars evolve to preserve information at word-beginnings

Andrew Wedel, Adam Ussishkin

Listeners identify words incrementally at the sub-lexical level, continually updating hypotheses about the identity of the word as its phonetic signal unfolds. Because information is integrated incrementally, phonetic cues earlier in the word contribute more information on average to lexical identification than phonetic cues later in the word (e.g. Van Son & Pols 2003), and listeners have been shown experimentally to pay preferential attention to word-beginnings (e.g., Nooteboom 1981). For example, the final [m] in the English word *vacuum* is less informative than the initial [v], because the probability of the word *vacuum* is already high given that the listener has processed the preceding signal [vækju-]. Recent work suggests that lexicons evolve in response to this bias, preferentially allocating more informative segments toward word-beginnings where they can contribute more to lexical access, especially in words that are more informative on average (King 2017; King & Wedel in prep; Meylan & Griffiths 2017).

The phonetic form of words is not static however: all languages are characterized by phonological grammars of rules describing predictable modifications of pronunciation in context. As an example, all stop consonants in German are devoiced word-finally, such that 'Hund' *dog* is pronounced 'Hun[t]'. Evidence suggests that phonological rules may develop through context-dependent shifts in the expected range of phonetic variation, which may arise in turn through the action of consistent biases on how speakers pronounce sounds in that context. Because phonetic cues are on average most informative at word beginnings, and speakers appear to pronounce informative phonetic cues more carefully than less informative cues (e.g., Aylett & Turk 2004; Buz, Tanenhaus & Jaeger 2016, Wedel, Nelson & Sharp 2018), we predict that languages should be less likely to evolve phonological rules at the beginnings of words, particularly those that reduce lexical contrast (Houlihan 1975). Here, we investigate this question through a statistical analysis of a cross-linguistic dataset of phonological rules.

We assembled a genetically and areally-diverse dataset of phonological rules coding whether the rule modifies the beginning or end of a word. The dataset contains 266 rules from 50 languages (Figure 1). We dichotomously coded each rule in the dataset for whether the rule is phonemically neutralizing (i.e., whether it potentially creates homophones). Two patterns emerge in this dataset: (i) there are significantly fewer rules overall that modify the beginnings of words, and (ii), this bias is significantly stronger for neutralizing rules (Figure 2). Both patterns are statistically significant in a mixed-effects regression model including Language, Family and Area as random intercepts. We further investigated two potential confounds that could explain this data: (i) a well-known bias toward rules that modify syllable-final consonants, and (ii) the fact that suffixing-dominant languages are more common: given that phonological rules often arise at stem-affix boundaries, the suffixing preference could potentially create an apparent end-bias in a random sample of languages. However, we show that neither of these alternative explanations can account for the observation that contrast-reducing rules preferentially target the ends of words.

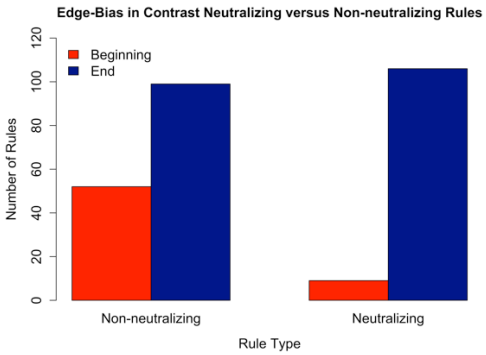
Two general hypotheses can account for the overall lower number of rules word initially. First, if speakers produce informative material with greater precision, we expect beginnings of words to exhibit an impoverished pool of variation from which rules can develop; second, production or processing of a phonological alternation in low predictability positions may entail greater resource cost on the part of the speaker or listener, which could inhibit alternations from developing in those positions, or hasten their loss.

These results provide the first statistical evidence supporting the hypothesis that languages evolve phonological grammars which preferentially preserve initial lexical material. This edge-based asymmetry supports the idea that, in order to account for the evolution of phonological rules, our measures of predictability must explicitly take prior sub-lexical

context into account. More broadly, this finding contributes to the increasingly sophisticated body of evidence that language structures evolve under conflicting biases toward accurate transmission of meaning and effort reduction (e.g., Zipf 1949; Lindblom 1990; Piantadosi et al. 2011; Wedel et al. 2013; Futrell et al. 2015, Hall et al. to appear).

Figure 1. Areal language distribution

Figure 2. Rule type by word edge



References

Aylett, M. & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47: 31-56.

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, 89, 68–86.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112: 10336-41.

Houlihan, K. (1975). The role of word boundary in phonological processes. Austin TX: University of Texas dissertation.

King, A. (2017). Word Informativity Affects Segmental Informativity. Unpublished Masters Thesis. University of Arizona, USA.

King, A. & Wedel, A. (in prep). Greater word informativity predicts greater word-initial segment informativity.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modelling*, ed. by W.J. Hardcastle & A. Marchal, 403-39. Dordrecht: Kluwer.

Meylan, S. C. & Griffiths, T. L. (2017). Word forms - not just their lengths - are optimized for efficient communication. Unpublished manuscript.

Nooteboom, S. G. (1981). Lexical retrieval from fragments of spoken words: beginnings vs. endings. *Journal of Phonetics* 9(4). 407-424.

Piantadosi, S. T, Tily H. & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108: 3526-29.

Van Son, R. J. J. H. & Louis C. W. Pols. (2003). How efficient is speech? *Proceedings of the Institute of Phonetic Sciences* 25: 171-84.

Wedel, A., Kaplan, A. & Jackson, J. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128: 179-86.
Wedel, A., Nelson, N.R., & Sharp R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language* 100: 61-88.
Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley.

Predictability and boundary strength in English compound nouns
Melanie J. Bell, Sonia Ben Hedia and Ingo Plag

This study explores the extent to which consonant duration at compound internal boundaries in English can be explained in terms of different types of predictability measure, and tests two competing hypotheses about the relationship between fine phonetic detail and morphological structure. On the one hand, complex words are hypothesised to show more phonetic reduction if they are less easily decomposable into their constituent morphemes (Hay 2001, 2004). In this vein, Ben Hedia & Plag (2017) find that, in prefixed English words, consonant duration at the morphological boundary varies according to the prefix involved. They hypothesise that these differences reflect differences in segmentability: consonants at morphological boundaries are longer in more semantically transparent, hence more decomposable, complex words. On the other hand, Kuperman et al. (2007), find that interfixes in Dutch compounds have longer durations when they are more probable relative to alternative interfixes, given the first element of the compound. They explain this in terms of the Paradigmatic Signal Enhancement Hypothesis, according to which lengthening is associated with greater paradigmatic support for a given element and hence with reduced uncertainty.

Both the hypotheses in the preceding paragraph can be related to paradigmatic probability, and for noun-noun compounds, the relevant paradigms are usually taken to be the positional morphological families of the constituent nouns (N1 and N2). For example, the compound *time machine*, has an N1 family including *time limit*, *time period*, *time scale*, *etc.*, while the N2 family includes *washing machine*, *fax machine*, *drum machine*, *etc.* The sizes of the constituent families, i.e. the number of different compounds that share a particular element, can be seen as type-based estimates of the entropy of the paradigms. The two hypotheses make opposing predictions about how N1 family size in particular will be related to consonant durations at the compound-internal boundary. On the one hand, the larger the N1 family, the more productive is N1 as a compound modifier, and greater productivity has been show to be associated with greater decomposability of complex words (Hay & Baayen 2003). Thus, if the segmentability hypothesis is correct, we would expect boundary consonant durations in compounds to be positively correlated with N1 family size. On the other hand, greater N1 family size also means that more possible compound heads can follow the modifier, so, at a paradigmatic type-based level, the larger the N1 family, the greater the uncertainty about N2. If the Paradigmatic Signal Enhancement Hypothesis is correct, we might therefore expect boundary consonant durations to be negatively correlated with N1 family size.

From the British National Corpus, we extracted 141 compounds with one of the consonants /n/, /m/, /l/ or /s/, either at the end of the first word (CV) e.g. *steam engine*, the start of the second word (VC) e.g. *media men* or both (CC) e.g. *team members*. Thirty-one adult native speakers of British English were recorded reading the compounds in carrier sentences. Consonant duration was extracted using Praat and used as the dependant variable in linear mixed effects regression modelling, with speaker and item as random effects. As well as N1 family size, we included a second measure assumed to correlate with compound segmentability, namely spelling ratio (compound frequency with unspaced orthography/ compound frequency with spaced orthography). The assumption is that writers are more likely to write a compound with a space the more decomposable they perceive it to be. To control for syntagmatic predictability, which is known to affect the duration of linguistic units at all levels of granularity, we used the conditional probability of N2 given N1, calculated as the frequency of the compound divided by the frequency of N1. Other predictors initially in the model included compound and constituent frequencies, N2 family size, various phonetic measurements, speech rate, consonant type, consonant position (final, initial or geminate) and the position of the compound in the sentence.

In the final model, after step-wise elimination of non-significant predictors, the only surviving frequency-based measure was N1 family size. Other significant predictors were sentence position, speech rate and the durations of the preceding and following vowels, which interact with consonant type and position. The partial effect of N1 family size on boundary consonant duration is shown in figure 1:

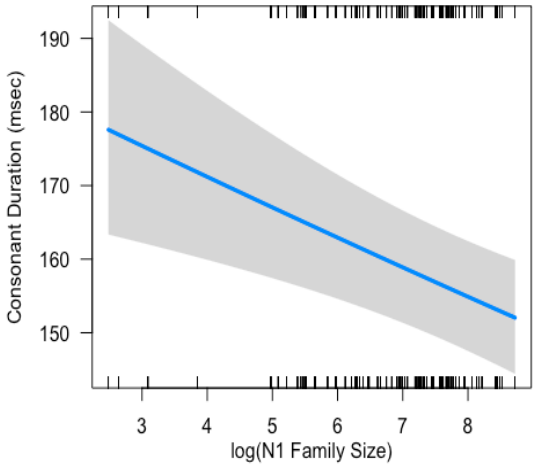


Figure 1: Partial effect of N1 Family size on Consonant Duration

It can be seen that, in keeping with the Paradigmatic Signal Enhancement Hypothesis, and against the segmentability hypothesis, consonant duration is inversely correlated with N1 family size. Further evidence against the segmentability hypothesis comes from the fact that spelling ratio is not a significant predictor of consonant duration. It is also striking that, in the presence of the type-based predictor, the token-frequency based conditional probability of N2 given N1 does not survive in the model (although the random effects structure includes a random slope for conditional probability by participant). Interestingly, in the models on which they based the Paradigmatic Signal Enhancement Hypothesis, Kuperman et al (2007) also used type-based measures of probability, which they found performed better than token-based measures. Hay (2001, 2004), on the other hand, used token-based measures in her work on decomposability. Our results therefore raise the questions as to whether type-based measures detect different effects from token-based measures, and/or whether compound boundaries are different from other word-internal boundaries regarding paradigmatic probability effects.

References

Ben Hedia, Sonia, and Ingo Plag. 2017. Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics* 62: 34-49.

Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39.6: 1041-1070.

Hay, Jennifer. 2004. *Causes and consequences of word structure*. Routledge.

Hay, Jennifer, and Harald Baayen. 2003. Phonotactics, parsing and productivity. *Italian Journal of Linguistics* 15: 99-130.

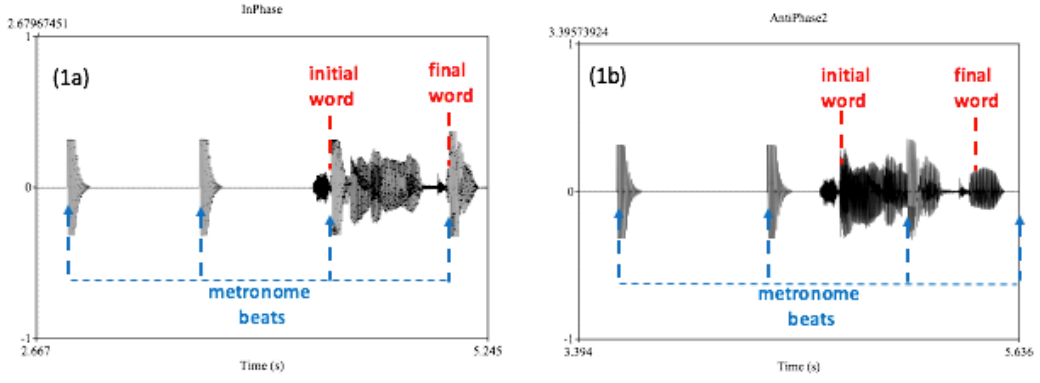
Kuperman, Victor, Mark Pluymaekers, Mirjam Ernestus, and Harald Baayen. 2007. Morphological predictability and acoustic duration of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America* 121.4: 2261-2271.

Coordinative Mode, Phasing, and Predictability Across Languages

Kathryn Franich
University of Chicago

Overview: Prosodic structure is argued to be implicated in models of phonological predictability both in structuring the speech signal to be maximally efficient with respect to information content (Aylett & Turk 2004; Turk & Shattuck-Hufnagel 2014) and through its role in helping listeners to dynamically generate expectations for upcoming structures during online processing (Breen et al. 2014, Brown et al. 2011; Dilley & McAuley 2008). Dynamic expectations of the latter sort have also been shown to influence listeners' processing of phonetic detail (Shaw 2016), suggesting that phonological predictability is reflected in a complex relationship between acoustics, speech timing, and grammatical structure. This paper further explores this relationship by investigating how prosodic expectations are influenced by the coordinative strategies used by speakers in aligning prosodically prominent events in speech with an external stimulus. Specifically, we examine how speaker-comprehenders of English and Medʉmba and Gã, two Niger-Congo languages, perceive speech in the context of a metronome played either *in-phase* (in synchrony) with metrically prominent words/syllables or *anti-phase* (on the 'offbeat'/'upbeat') of such words. Results indicate that English speakers' lexical identification is facilitated by in-phase presentation of these words with the metronome beat, while Medʉmba and Gã speakers' identification is facilitated by anti-phase presentation. Results suggest that speakers of different languages utilize different coordinative strategies when perceiving language in real-time, and that coordinative mode is a variable which should be considered in our conceptualization of predictability as it relates to speech perception/processing and communicative efficiency.

Background: This work builds off of a production study (Franich 2017) in which speakers of Medʉmba repeated sentences in time to a metronome set to different rates. While previous work examining metronome alignment patterns with English speakers (Cummins & Port 1998) showed a strong preference among speakers to align the vowel onsets of metrically-strong initial and final syllables of their utterances (e.g. the words 'Dig' and 'duck' in the sentence *Dig for a duck* in English) in-phase (or in synchrony) with the metronome beat (Fig. 1a), Medʉmba speakers showed a preference for aligning metrically-strong words on the 'offbeat' or 'upbeat' of the metronome (Fig. 1b). All 14 subjects in the original study by Franich (2017) showed similar patterns of alignment with the metronome beat. This was despite instructions and encouragement to produce the first word in sync with the metronome beat.



In-phase pattern **Anti-phase pattern**

Figure 1: Alignment of first and last syllables (as measured from vowel onsets) in repetitions of *Mén lɛ̌n nàʧú* ('The child knows how to enter') by a Medʉmba speaker with the third and fourth of four evenly-timed metronome beats (Franich 2017). Figure 1b represents the prevalent pattern produced by speakers in which the initial word was produced after the third beat, around 1/2 of the way through the cycle of the metronome beat. Figure 1a shows the

sentence adjusted so that the first word aligns with the metronome beat.

The present study examines how coordinative patterns between speech and metronome may also influence listeners' perception of words in similar sentential contexts.

Participants: Participants (data collection ongoing) include 20 speakers of English from the United States, 2 speakers of Medʉmba, a Grassfields Bantu language spoken in Cameroon, and 3 speakers of Gã, a Kwa language spoken in Ghana. Medʉmba and Gã are similar in their tonal properties (both include High/Low tonal contrasts) and morphosyntactic properties (both include largely mono- and disyllabic words/isolating morphology).

Method: A speeded lexical identification experiment was run in which the final, metrically-prominent syllable[1] in a four-syllable sentence was randomly varied, and speakers were asked to indicate the final word they heard. Sentences were presented in one of two conditions: *in-phase*, in which the initial/final words of the spoken sentence were presented in synchrony with a metronome beat, and *anti-phase*, in which the initial/final words were presented on the 'offbeat' of the metronome, 1/2 of the way through the beat cycle, as indicated in Fig. 1b. In all trials, subjects heard four beats leading up to the target beat/sentence pairings. Sample sentences are provided in Fig. 2. Each of 6 sentences per language was played 4 times per condition, for a total of 6 x 2 x 4 = 48 utterances heard per speaker.

English	*We learned to teach*
Medʉmba	*Mén lɛ̌n nə̀-ʧú* child know inf-enter 'The child knows how to enter.'
Gã	*Wɔ́ kásé nú* we learn water 'We learn water'

Figure 2: Sample stimuli

Results: As shown in Fig. 3-4, English speakers were faster at identifying the target word when it was presented in-phase with the metronome beat ($p < 0.05$), whereas speakers of the two African languages showed the opposite effect of faster identification in the anti-phase condition ($p < 0.05$). The facilitation effect for in-phase patterning was found across all English-speaking subjects, and the anti-phase affect was found across Medʉmba and Gã speakers.

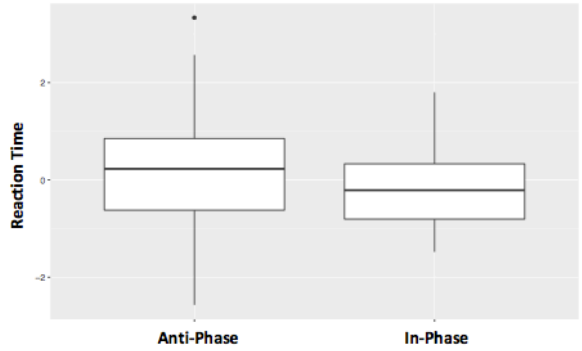


Figure 3: Results for English Speakers

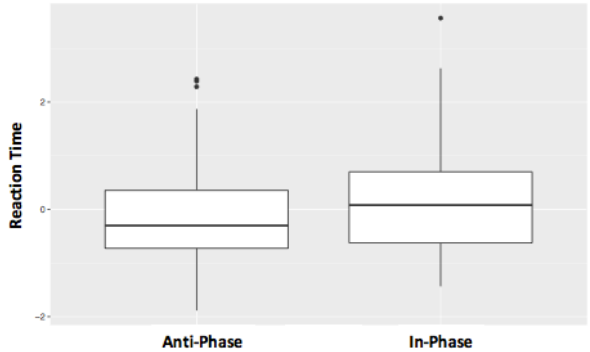


Figure 4: Results for African Language Speakers

Discussion/Conclusion: Why should speech-to-metronome alignment patterns affect lexical perception differently across languages? We propose that this may be an areal phenomenon

[1] See Franich (2018) for phonological evidence that stem-initial syllables in at least some Niger-Congo languages are metrically-prominent.

relating to the different types of exposure that speakers have to anti-phase versus in-phase patterns in linguistic contexts as well as through other coordinative activities, as in dancing and singing to music. This allows for the quantifiable expectation that reduced exposure to certain phasing patterns (namely, anti-phase patterns, which are relatively less stable than in-phase patterns; Haken et al. 1985) will lead to changes in the types of coordinative modes that speakers utilize in perception, and concomitant 'retuning' of speech perception to environmentally-relevant coordinative modes. We discuss specific ways in which this hypothesis might be tested.

The relative importance of in-phase and anti-phase coordinative modes in this study is posited based on the idea that coordination with an external stimulus (including with a conversation partner in turn-taking; see Wilson & Wilson 2005) can be modeled in terms of entrainment between adaptive coupled oscillators (Large & Jones 1999; McAuley 1995; Port 2003; Salzman & Byrd 2000). This modeling approach makes the additional prediction that in-phase and anti-phase coordination should be preferred over any other coordinative patterns, a prediction which is again readily testable with an expanded stimulus set using the current experimental paradigm. Finally, these results speak more generally to the need for understanding the role of conversational interaction in shaping predictive processes in speech perception and production, and how interactional factors may relate to previously-identified talker- and listener-oriented sources of phonetic variation (e.g. from Lindblom 1990).

References:

Aylett, M., Turk, A. (2004) The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence and Duration in Spontaneous Speech. Language and Speech, Volume 47(1), 31-56.

Breen, M., Dilley, L.C., McAuley, J.D., and Sanders, L.D. (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Language, Cognition and Neuroscience*. 29 (9), 1132-1146.

Brown, M., Salverda, A. P., Dilley, L. C, & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin and Review*, 18(6).

Cummins, F. and Port, R. F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26(2):145-171.

Dilley LC, McAuley JD. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*. 59:294–311

Franich, K. (2018). Evidence for metrical prominence asymmetries in Medumba. In Gallagher, Gillian, Gouskova, Maria, and Sora Yin (eds.), *Proceedings of the 2017 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America. URL.

Garrod, S. & Pickering, M.J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, 6, 751.

Haken, H., Kelso, J.A.S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51, 347 356.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1), 119-159.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory, In W. J. Hardcastle & A. Marchal: *Speech Production and Speech Modeling*, 403-439, Kluwer Academic Publishers, Dordrecht.

McAuley, J.D. (1995). *Perception of time as phase: Toward an adaptive-oscillator model of rhythmic pattern processing.* Unpublished doctoral dissertation, Indiana University, Bloomington.

Port, Robert (2003). Meter and speech. *Journal of Phonetics*, 31, pp. 599-611.

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, 19, 499-526.

Shaw, J. (2016). Dynamic listening: temporal expectations guide perception of phonetic detail. *Presentation at LabPhon 15*, Cornell University.

Turk, A. and Shattuck-Hufnagel, S. (2014), Timing in talking: What is it used for, and how is it controlled? Philosophical Transactions of the Royal Society B: Biological Sciences 369(1658) DOI: 10.1098/rstb.2013.0395

Wilson, M. & Wilson, T.P. (2005). An oscillator model of timing in turn-taking. *Psychonomic Bulletin and Review*, 12(6), 957-968.

Japanese Perceptual Epenthesis is Modulated by Transitional Probability

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: Perceptual epenthesis occurs when a listener is exposed to speech that violates the phonotactics of their native language. For instance, Japanese does not allow non-homorganic consonant clusters and Japanese listeners sometimes report hearing an illusory /u/ when exposed to sequences that violate this restriction (e.g., /ebzo/→/ebuzo/) [1]. The /u/—as opposed to other vowels—is epenthesised because /u/ is the most perceptually minimal vowel in Japanese (shortest and most quiet), making it the best match to the non-existent sound [1]. However, studies on Korean have shown that different illusionary vowels can be perceived when the epenthesis of the minimal vowel would violate native co -occurrence restrictions [2]. Using an extension of the Perceptual Assimilation Model [3] which takes into account the frequency at which sounds co-occur within a native language [4], we propose that these patterns arise because listeners rely on transitional probability, in addition to perceptual similarity, to assimilate unfamiliar sequences to those that they are frequently exposed to in L1 discourse.

METHOD: To test the effects of transitional probabilities on perceptual epenthesis, we conducted two experiments which compare the perceptual epenthesis of two Japanese vowels—the phonetically minimal vowel /u/ and the next most minimal vowel /i/ [5]—in illicit VCCV tokens, by examining different consonant environments with different transitional probabilities. To determine transitional probability, we employ an existing corpus analysis [6] to quantify how frequently /u/ and /i/ occur in CV sequences with /tʃ/, /ʃ/ and /g/; these consonants were chosen due to them having a higher transitional probability with /i/ than with /u/. Shannon's [7] Surprisal (/Ci/ Surprisal: /tʃi/ = 0.66, /ʃi/ = 0.61, /gi/ = 3.32; /Cu/ Surprisal: /tʃu/ = 0.32, /ʃu/ = 0.31, /gu/ = 4.05) and Entropy (/tʃ/ = 1.38, /ʃ/ = 1.47, /g/ = 1.76) were calculated to provide a measure of transitional probability and to quantify the uncertainty about vowels associated with each consonantal environment. In Experiment 1, 30 native Japanese listeners from Keio University in Tokyo categorised illegal VCCV tokens (e.g., [egpo]) into one of two VCVCV categories (e.g., /egupo/ or /egipo/). In Experiment 2, the same participants discriminated between illicit VCCV tokens and licit VCVCV tokens (using the stimuli from Experiment 1) which vary between /u/ and /i/ in the epenthetic medial position in AXB trials.

RESULTS: The results of Experiment 1 appear in Table 1. Of the illicit tokens, [egpo] was categorised as /egupo/ almost 100% of the time, [eʃpo] was categorised as /eʃupo/ 69% of the time and [etʃpo] was categorised as /etʃupo/ 54% of the time. A similar pattern emerged in Experiment 2 (Table 2): listeners were more successful at discriminating between VCCV and VCiVC tokens when the consonant preceding the epenthetic position maintained a high Entropy. Listeners successfully discriminated between /egpo/ and /egipo/ 93% of the time but were only able to discriminate between /egpo/ and /egupo/ 82% of the time, a difference of 11%, and the likewise, /ʃ/ condition exhibited a difference of 5%. The /tʃ/ condition, which has the lowest Entropy, showed no significant difference. Additionally, these results are reflected in the response times for AXB tests where low Entropy tests exhibit less difference.

DISCUSSION: The results suggest that perceptual epenthesis is partly shaped by erroneous predictions made based on the listener's prior linguistic experience. Speech is filtered through the lens of the listener's native language which considers the perceptual similarity of sounds and the transitional probability of those sounds co-occurring. The results also suggest that perceptual similarity—in the case of perceptual epenthesis, perceptual minimality—is the most

important factor in these predictions, as /u/ is preferred even when the preceding consonant is more likely to co-occur with /i/ rather than /u/. However, the differences observed between the three consonants demonstrate that transitional probability is also an important factor in determining the quality of epenthetic vowels. The Surprisal of /gi/ as a sequence is far higher—meaning it is less probable—than that of /ʃi/ or /tʃi/ and the Entropy of /g/ is lower than both /tʃ/ and /ʃ/; the results suggest that /g/ very infrequently elicits /i/ epenthesis. The Surprisal of /ʃi/ is higher than /tʃi/, yet /tʃ/ is significantly more likely to elicit /i/ epenthesis than /ʃ/. Here, we propose that this difference can be at least partly explained by considering the Entropy differences: /ʃ/ has a greater uncertainty as to which vowel may follow it in CV sequences. Because /tʃ/ has a higher certainty as to which vowel may follow—and the most likely vowel in these sequences is /i/—we observe a significant difference in the quality of epenthetic vowels following /tʃ/ and /ʃ/. This suggests that when the overall co-occurrence predictability of a consonant is high, listeners are more likely to rely upon co-occurrence frequency, while they are more likely to rely upon perceptual similarity when the predictability is low. One possible interpretation of these results is listeners rely on varying levels of transitional probability information, depending on the level of uncertainty.

Token	[etʃpo]		[eʃpo]		[egpo]	
Category	/etʃupo/	/etʃipo/	/eʃupo/	/eʃipo/	/egupo/	/egipo/
Categorisation Rate	54%	46%	69%	31%	100%	0%

Table 1. Results of Categorisation Task (Experiment 1)

AXB Contrast	**Accuracy**	**ACC. Diff.**	**Response Time**	**RT. Diff.**
/etʃipo/-/etʃpo/	73%	1%	1201 ms	-5 ms
/etʃupo/-/etʃpo/	72%		1196 ms	
/eʃipo/-/eʃpo/	77%	5%	1160 ms	42 ms
/eʃupo/-/eʃpo/	72%		1202 ms	
/egipo/-/egpo/	93%	11%	1040 ms	117 ms
/egupo/-/egpo/	82%		1157 ms	

Table 2. Accuracy and Response Time results of Discrimination Task (Experiment 2)

[1] Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from?. *Journal of Memory and Language*, *64*(3), 199-210.

[2] Durvasula, K., & Kahng, J. (2015). Illusory vowels in perceptual epenthesis: the role of phonological alternations. *Phonology, 32*(3), 385-416.

[3] 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.

[4] Kilpatrick, A. J., Bundgaard-Nielsen, R. L., & Baker, B. J. (In Press). Japanese phonotactics influence second language perception. *Applied Psycholinguistics*.

[5] Arai, T., Warner, N., & Greenberg, S. (2001). OGI tagengo denwa onsei koopasu-ni okeru nihongo shizen hatsuwa onsei no bunseki. *The spring meeting of the acoustical society of Japan, 1*, 361-362.

[6] Tamoka, K., & Makioka, S. (2004). Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. *Behaviour Research Methods, Instruments & Computers, 36*(3), 531-547.

[7] Shannon, C. E. (1948). *A mathematical theory of communication*. MA thesis, MIT.