Exemplar Averaging of Phonetically Discrete Variants[1]

Daiki Hashimoto

University of Tokyo
daiki@phiz.c.u-tokyo.ac.jp

ABSTRACT

Exemplar Theory hypothesizes that a production target is formed by choosing a target exemplar and averaging the phonetic value of the exemplar with those of the surrounding exemplars (Pierrehumbert [1]; Wedel [2]). The aim of this study is to seek acoustic evidence for exemplar-averaging of phonetically distinct variants by exploring New Zealand English (NZE) loanword phonology. The /r/ sounds in loanwords from te reo Māori may be realized as non-native rhotic sounds [ɾ] or native rhotic sounds [ɹ] in NZE. This study demonstrates that the acoustic characteristics of a speaker's pronunciation of [ɹ] correlate with the number of approximants versus taps she produced. This could be captured by positing that exemplars with [ɹ] are likely to be averaged with those with [ɾ] when they are *less predictable* given a speaker, with the result that the acoustic property of [ɹ] (i.e., lower F3) becomes less clear.

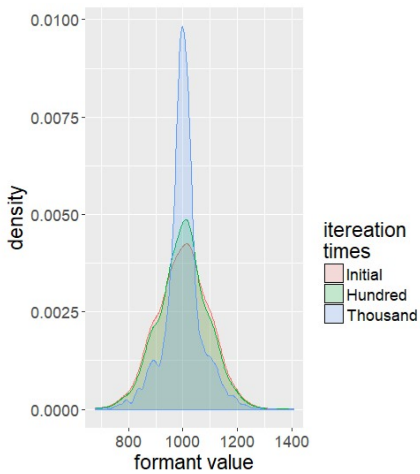
Keywords: approximant, averaging, exemplar, F3, predictability

1. INTRODUCTION

Exemplar theory assumes that a speaker represents exemplars with detailed phonetic information (e.g., formant value and duration) as well as categories (e.g., lexical category and phonological category), the result of which is that a large cloud of memories is formed in the cognitive system. It is posited that a speaker begins speech production by activating a category they want to produce, and chooses exemplars associated with the category, based on which the speaker forms a production target.

Pierrehumbert [1] notes that "a target location in the exemplar cloud is selected at random, and the exemplars in the neighborhood of this location all contribute to the production plan." That is, it is hypothesized that a production target is formed by choosing a target exemplar and averaging the phonetic value of the exemplar with those of the surrounding exemplars. This mechanism is called "averaging." The mechanism is needed to account for a phenomenon known as entrenchment (see Pierrehumbert [1]; Wedel [2]), in which the variance for a category decreases with usage. For example, a L2 learner may produce a particular vowel with a variety of formant values in the early stage of the acquisition, but the learner may come to be able to produce the vowel with more fixed formant values. Figure 1 simulates how this averaging mechanism could make a particular category sharper. Imagine that a phonetic category (e.g., [ʊ]) is represented by 1,000 exemplars with particular formant values in the initial stage (red), and a production target is formed by choosing 10 exemplars in a random way and averaging the formant values. After this category is produced 100 times (i.e., 100 production targets) and they are stored as new exemplars, the distribution becomes a little bit narrower (green). After 1,000 productions, the distribution becomes much narrower (blue). In this way, the exemplar-averaging mechanism allows us to encapsulate entrenchment.

Figure 1: Simulation of averaging exemplars



The aim of this study is to test this hypothesis regarding "exemplar-averaging," by exploring a case where there are two phonetically quite distinct variants that are in variation within a single phonological category. Our main research question is "Are exemplars that are cognitively linked but phonetically distinct averaged in production? **(Research Question)**" This research question is addressed by exploring the phonetic detail of [ɹ] in Māori loanwords in NZE, more specifically, the relationship between the F3 value of adapted [ɹ] and the predictability of choosing adapted structure given a loanword and a speaker. Addressing this question increases our understanding of how linguistic variants are stored in our mind and processed in production.

2. BACKGROUND AND PREDICTION

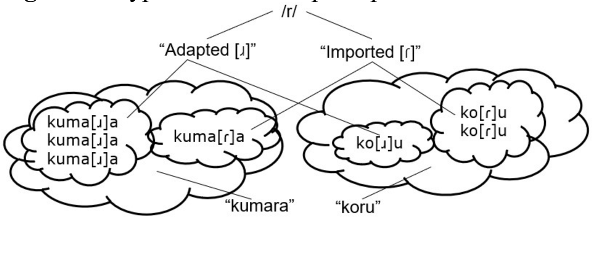
This section reviews previous literature related to the above research question, and puts forward a specific prediction.

2.1. Realization of /r/ in NZE loanword phonology

The loanword phonology in NZE provides an interesting test case for the above question. In the production of a loanword from te reo Māori, the /r/ sounds are sometimes realized as native rhotic sounds or adapted [ɹ] (e.g. *ko*[ɹ]*u* and *kuma*[ɹ]*a*), and sometimes realized as non-native rhotic sounds or imported [ɾ] (e.g., *ko*[ɾ]*u* and *kuma*[ɾ]*a*). That is, there are two variants, [ɾ] and [ɹ], of the variable /r/. My previous work [3] has demonstrated that the likelihood of choosing a variant differs in accordance with loanwords and speakers, i.e., some loanwords are more likely to be produced with adapted [ɹ], and some NZE speakers are more likely to produce adapted [ɹ] in production of a loanword.

Namely, it could be assumed that a phonological category /r/ is represented by exemplars with two phonetically distinct sounds (i.e., [ɾ] and [ɹ]) in the cognitive system of a NZE speaker, and the rate of exemplars with [ɾ] and [ɹ] differs in accordance with lexical categories and cognitive systems. Exemplars with [ɾ] and [ɹ] may be stored in a close distance in the exemplar space, because they are variants of a single category. Figure 2 illustrates how loanwords and phonological categories could be represented in the cognitive system of a given speaker. In this hypothetical system, a lexical category "kumara" is represented by a higher number of exemplars with adapted [ɹ], with the reverse true for another lexical category "koru." As shown in this figure, it is assumed that an exemplar can simultaneously be associated with a lexical category and a phonological category (see Docherty and Foulkes [4]).

Figure 2: Hypothetical exemplar space



2.2. Theoretical hypotheses and prediction

This section reviews a set of hypotheses based on Exemplar Theory, and aims to deduce a prediction regarding the formation of a production target for adapted [ɹ] by averaging exemplars with adapted structure [ɹ] and those with imported structure [ɾ].

As stated in Section 1, Exemplar Theory posits that a production target is formed by choosing a target exemplar and averaging the phonetic value of the exemplar with those of the surrounding exemplars. For example, in production of a vowel [ʊ], one exemplar associated with a category [ʊ] is chosen as a target exemplar. Then, the formant value of the exemplar is averaged with the formant values of the surrounding exemplars, and the production target is formed. Recall that this mechanism allows us to account for entrenchment (see Figure 1).

This raises a question: Is it possible to average the phonetic values of exemplars that are cognitively linked but phonetically distinct? As explained in the preceding section, the current thesis posits that exemplars with imported [ɾ] and adapted [ɹ] are closely stored in the cognitive system of a NZE speaker, as they are variants of /r/-sounds in Māori loanwords. If this is true, when forming a production target for [ɹ], a target exemplar with [ɹ] may be averaged with not only surrounding exemplars with [ɹ] but also those with [ɾ] **(Hypothesis 1)**[2].

We assume that the higher predictability of adapted structure given a *loanword* means that there are higher number of exemplars with [ɹ] stored in relation to the *lexical category*, and vice versa; the higher predictability of adapted structure given a *speaker* means that there are higher number of exemplars with [ɹ] stored throughout the cognitive system. As long as there are a higher number of exemplars with [ɹ] stored, a target exemplar with [ɹ] may be more likely to be averaged with surrounding exemplars with [ɹ] **(Hypothesis 2)**.

We crucially hypothesize that exemplars representing [ɹ] are stored with a lower F3 value, because approximants are known to have lower F3 in comparison with other sounds (Lindau [5]); exemplars representing [ɾ] are stored with null F3 values, because they consist of simply closure and release phases without inherent formant structure. Hence, a production target for [ɹ] has a lower F3 value, when it is formed by averaging a larger number of exemplars with [ɹ] (see Hay and Maclagan [6]); a production target for [ɹ] has a higher F3 value, when it is formed by averaging a larger number of exemplars with [ɾ] **(Hypothesis 3)**.

On the basis of Hypotheses 1-3, the following prediction can be put forward: adapted structure [ɹ] is produced with *lower F3*, when it is *more predictable* given a loanword (i.e., a larger number of exemplars with [ɹ] are stored amongst exemplars associated with the lexical category) **[Prediction A]**, and when it is *more predictable* given a speaker (i.e., a larger number of exemplars with [ɹ] are stored within the cognitive system) **[Prediction B]**.

3. RESEARCH DESIGN

This section illustrates the research design employed to test the above prediction.

3.1. Experiment: passage-reading task

96 NZE speakers participated in a passage-reading task. They are aged between 18-35, and not speakers of te reo Māori. They were interviewed individually by the author in a sound booth.

Their task is to read several short passages including Māori loanwords. Half the passages are about Māori society or history, and the other half are about general life or leisure in New Zealand. The target passages contain 15 Māori loanwords (e.g., *Oamaru*, and *Whangarei*), which are place names because topics can be manipulated naturally. The target loanwords are mentioned twice within a passage, and they appear in a few passages. The loanwords all include word-medial /r/, of which the realizations can be classified into adapted [ɹ] and imported [ɾ]. The order of the passages was pseudo-randomized to avoid order effects that may affect our interpretation of the data.

3.2. Classification of /r/ realizations into [ɹ] and [ɾ]

The current study mainly relies on spectrograms to identify the realization of a /r/ sound in a target Māori loanword as adapted [ɹ] or imported [ɾ]: a /r/ sound was identified as adapted [ɹ] if it was realized with no clear consonantal edges with lowered F3; it was identified as imported [ɾ] if it was realized with clear consonantal edges. The remaining productions (e.g., realization with neither clear consonantal edges nor lowered F3) were impressionistically identified as [ɹ], [ɾ], or others. This is performed by the author.

Table 1: Total observation of /r/ realization

Classification	Number (Ratio)
Acoustically [ɾ]	3,920 (41.5%)
Acoustically [ɹ]	4,474 (47.4%)
Impressionistically [ɾ]	655 (6.9%)
Impressionistically [ɹ]	183 (1.9%)
Others	216 (2.3%)
SUM	**9,448**

3.3. Calculation of predictability of choosing [ɹ]

As in the Predictions, two types of predictability were calculated, i.e., *p*(ɹ|loanword) and *p*(ɹ|speaker). These predictabilities were calculated based on the above total observations excluding tokens classified as others. For statistical reasons, they were transformed into information content (IC) by taking $-\log_2$. Note that higher probability results in lower information content. For instance, 208 tokens of a loanword *Tokoroa* were produced with adapted structure, and 355 tokens of *Tokoroa* were produced with imported structure in the experiment reported in the current study. Then, the probability of [ɹ] given *Tokoroa* *p*(ɹ|Tokoroa) is 36.9% (i.e., 208/563), and the information content *IC*(ɹ|Tokoroa) is 1.36 bits.

3.3. Measurement of F3 in adapted [ɹ]

The current study explores solely acoustically identified [ɹ]-realizations when discussing a F3 value of adapted [ɹ]. The domains of these tokens were identified by consistent F2 around the lowest F3 (Lavoie [7]), although it is arbitrary to some extent to identify the domain of inter-vocalic approximants. The lowest F3 value within the domain was adopted as the formant value of a token identified as [ɹ], using a Praat script (Boersma and Weenink 2016 [8]).

4. RESULT

This section explains the statistical results to assess Predictions A and B. The 4,474 tokens of acoustically identified [ɹ]-realizations were hand-fitted into a mixed-effects regression model with the *lmer* function in the *lme4* library (Bates et al. [9]) and *lmerTest* library (Kuznetsova et al. [10]) implemented in R (R Core Team [11]).

As in the Predictions, a response variable is a z-scored F3 value of adapted [ɹ], and key variables are *IC*(ɹ|loanword) and *IC*(ɹ|speaker), which were centred for statistical reasons. Although our interest lies in this relationship, several factors were fitted as control variables. First, anatomical differences are known to affect formant values (Watt et al. [12]). Hence, *nativeF3* and *gender* were fitted as control variables. *nativeF3* represents a speaker-specific potential F3 value, and was measured by analysing inter-vocalic /r/ approximants in some native words that appeared in the reading passages. We also explored some factors that might affect F3 as control variables such as phonological factors (e.g., stress and prosodic break), passage factors (e.g., topic and speech rate), word-specific factors (e.g., word frequency and word length), and speaker-specific factors (e.g., relationship with Māori language and culture).

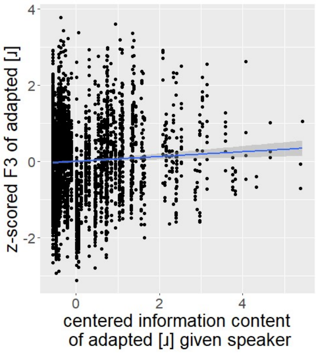
In order to find the best-fitted model, we started with a model with all the above variables and two random intercepts for participants and words. Backward elimination was performed manually through pairwise model comparisons using ANOVA. Interactions between significant main effects were also considered. If a model comparison shows no significance (p>.05), then the smaller model was adopted. This comparison was performed until all predictors reached significant levels, with the result that the following model was selected (Table 2). The VIF scores are below 4, suggesting that there is no multicollinearity in this model. The other assumptions of linear model were checked visually by inspecting a residual plot and histogram (Winter [13]).

Table 2: Model summary of the best-fitted model

	β	*SE*	*t*	*p*
(intercept)	0.2207	0.0807	2.43	**
IC(ɹ|speaker)	0.1497	0.0567	2.64	*
nativeF3	0.4599	0.0589	7.8	***
gender male	−0.28	0.1255	−2.2	*
NofSegment	0.324	0.0706	4.59	***
wdFreq	−0.053	0.0232	−2.3	*

As the dependent variable is a z-scored F3 value, positive coefficients indicate higher F3 values, and vice versa. It was found that *IC*(ɹ|speaker) significantly affects the F3 value, i.e., adapted [ɹ] is produced with *lower F3*, when it is *more predictable* (i.e., it has *lower IC*) given a speaker (see Figure 3). On the other hand, *IC*(ɹ|loanword) is not significant (p=0.25), but the direction is in line with *IC*(ɹ|speaker). As for control variables, it was found that adapted [ɹ] is produced with *higher F3* by speakers potentially producing higher F3; by female speakers; in loan-words with more segments and with low frequency.

Figure 3: Relation between F3 and *IC*(ɹ|speaker)

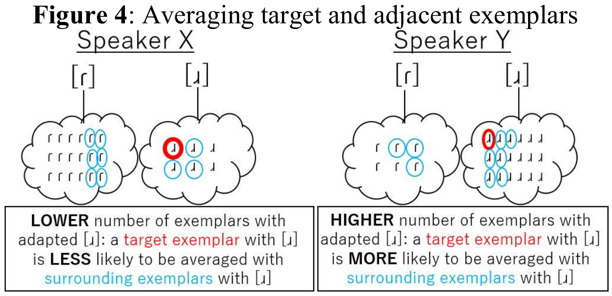


5. DISCUSSION AND SUMMARY

The effects of *gender* and *nativeF3* are as expected, that is, F3 values of adapted [ɹ] in loanwords are affected by anatomical differences. The effect of word length may be because a segment is likely to be produced with shorter duration in longer words (Lindblom [14]). This is why the F3 could not be lowered well in longer words. The effect of word frequency could be interpreted to mean that high frequency loanwords are incorporated into the English lexicon more fully than lower frequency loanwords, and they are more likely to be pronounced in the same way as native words.

Finally, let us discuss the two key variables, *IC*(ɹ|speaker) and *IC*(ɹ|loanword). The effect of *IC*(ɹ|speaker) is statistically significant in our dataset (β=0.14, t=2.64, p<0.05). This supports our Prediction B: adapted [ɹ] is produced with *lower F3*, when it is *more predictable* (has lower IC) given a speaker. Our exemplar-based interpretation is as follows: when forming a production target for [ɹ], a target exemplar with [ɹ] may be averaged with not only surrounding exemplars with [ɹ] but also those with [ɾ] (Hypothesis 1). When there are higher number of exemplars with [ɹ] represented in the cognitive system of a speaker (i.e., the predictability of [ɹ] given a speaker is higher), a target exemplar with [ɹ] may be more likely to be averaged with surrounding exemplars with [ɹ] (Hypothesis 2), see Figure 4. A production target for [ɹ] is produced with lower F3 value, when it is formed by averaging a larger number of exemplars with [ɹ] (Hypothesis 3).

Figure 4: Averaging target and adjacent exemplars



Our results show that *IC*(ɹ|loanword) is not statistically significant (p=0.25), and therefore Prediction A was not well-supported in the current study. However, we would like to note that the direction of the effect of the word-specific predictability is in line with our averaging-based Prediction. The null result might be due to the narrow set of loanwords employed in the passage-reading task, which might not enable us to explore a variety of values of the predictability given a loanword. Recall that the target loanwords are all place names. The future exploration of this predictability effect using a larger dataset including common nouns might strengthen the hypothesis regarding exemplar-averaging in relation to a lexical category.

In summary, this study provides evidence that exemplars that are cognitively linked but phonetically distinct can be averaged in production. The result reported in the current study can neatly be captured by positing that exemplars with [ɾ] and [ɹ] are stored closely, and they are averaged when forming a production target for [ɹ]. The findings develop our understanding how phonetically distinct variants are stored in our mind and processed in production.

6. REFERENCES

[1] Pierrehumbert, J. 2001. Exemplar dynamics, word frequency, lenition, and contrast, In : Bybee, J., and Hopper, P. (eds), *Frequency effects and the emergence of linguistic structure*. 135–157. Amsterdam: John Benjamins.

[2] Wedel, A. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23, 247–274.

[3] Hashimoto, D. 2019. *Loanword phonology in New Zealand English: Exemplar activation and message predictability*. Christchurch: University of Canterbury doctoral thesis.

[4] Docherty, G., and Foulkes, P. 2014. An evolution of usage-based approaches to the modelling of sociophonetic variability. *Lingua* 142. 42–56.

[5] Lindau, M. 1985. The story of /r/, In: Fromkin, V.A. (ed), *Phonetic Linguistics*. 157–168. Orlando: Academic Press.

[6] Hay, J., Maclagan, M. 2012. /r/ sandhi in early 20th century New Zealand English. Linguistics 50 (4). 745–763.

[7] Lavoie, L. 2001. *Consonant strength: Phonological patterns and phonetic manifestation*. New York: Routledge.

[8] Boersma, P., Weenink, D. 2016. Praat: Doing phonetics by computer (version 6.0.18). www.praat.org

[9] Bates, D., Maechler M., Bolker, B., Walker, S. 2015. Fitting linear mixed effects models using lme4. *Journal of statistical software* 67(1), 1–48.

[10] Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. 2016. lmerTest: Tests in linear mixed effects models (R package version 2.0-32) https://CRAN.R-project.org/ package=lmerTest

[11] R Core Team. 2016. R: A language and environment for statistical computing (version 3.2.4). R Foundation for statistical computing, Vienna, Austria. https://www.R-project.org/.

[12] Watt, D., Fabricius, A., Kendall, T. 2011. More on vowels: plotting and normalization, In: Paolo, M.D., Yaeger-Dror, M. (eds), *Sociophonetics: a student's guide*. 107–118. New York: Routledge.

[13] Winter, B. 2013. Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. http://arxiv.org/pdf/1308.5499.pdf

[14] Lindblom, B. 1968. Temporal organisation of syllable production. *Speech transmission laboratory, Quarterly progress and status report*, 1–5.

---

[1] This study is a part of my doctoral thesis submitted to the University of Canterbury. I deeply appreciate my supervisory team for their guidance: Jen Hay (Canterbury), Beth Hume (Ohio), and Jeanette King (Canterbury). This study also benefits from three anonymous reviewers.

[2] It is also worth exploring whether a production target for imported [ɾ] is formed by averaging a target exemplar with [ɾ] with surrounding exemplars with [ɹ]. This will be reported somewhere else.