



López Assef, Belén & Strahm, Stephanie & Boyce, Keara & Page, Michael P. A. & Zamuner, Tania S. 2023. Production benefits recall of novel words with frequent, but not infrequent sound patterns. *Glossa: a journal of general linguistics* 8(1). pp. 1–21. DOI: <https://doi.org/10.16995/glossa.8582>



Open Library of Humanities

Production benefits recall of novel words with frequent, but not infrequent sound patterns

Belén López Assef, Department of Linguistics, University of Ottawa, CA, mlope075@uottawa.ca

Stephanie Strahm, School of Audiology and Speech Sciences, University of British Columbia, CA, stephaniestrahm@gmail.com

Keara Boyce, School of Communication Sciences and Disorders, Dalhousie University, CA, K.Boyce@dal.ca

Michael P. A. Page, Department of Psychology, University of Hertfordshire, UK, m.2.page@herts.ac.uk

Tania S. Zamuner, Department of Linguistics, University of Ottawa, CA, tania.zamuner@uottawa.ca

The production effect is influenced by various factors, including cognitive and linguistic-related variables. Previous studies found that the production effect varies when stimuli have native versus non-native speech sounds, but to date, no studies have investigated whether the effect is also modulated by the frequency of sound patterns within a language. Adults were taught novel words in two training conditions: Produced or Heard. These items were comprised of English sound patterns that varied in frequency. Participants trained on frequent English patterns recalled more Produced than Heard items. In contrast, participants trained on infrequent English patterns showed no difference in recall rates between conditions. The strength and direction of the production effect is modulated not only by native versus non-native speech sounds, but can also vary depending on the frequency of the sound patterns within a speaker's native language. Thus, the production effect is linked to previously established, long-term phonological knowledge.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

OPEN ACCESS



1. Introduction

Language learning involves the creation of representations. As language develops, these representations are built on previously established knowledge, which can in turn influence the learning of new material. This is the case with phonotactic probabilities. Experience with language provides information about which patterns are more frequent in a language and this information can later be a factor in word learning. For example, learners show higher recall of novel words comprised of frequently occurring sounds in the ambient language (Thorn & Frankish 2005). In addition to linguistic elements, other factors have been shown to have an impact on the learning of items, such as the manner in which words are trained. Adults show improved memory for words when they have overtly produced during training relative to words that are read silently, known as *The Production Effect* (Hopkins & Edwards 1972; MacLeod et al. 2010). The effect of production is subject to various factors which vary in their nature, ranging from cognitive variables (e.g., attention, task difficulty) to language-related ones (e.g., language experience). For example, results from studies looking at linguistic factors have found that characteristics of the stimuli, whether they are comprised of native or non-native sound structures, can change the outcome of the learning task: the production effect is reversed (disadvantage for produced items) when learning non-native sound structures (Kaushanskaya & Yoo 2011; Thorin et al. 2018; Baese-Berk 2019). Identifying the variables that influence the strength and direction of the production effect is central to understanding how learners build representations. The aim of this research was to further investigate the role of linguistic factors on the production effect by determining whether the within-language familiarity of the stimuli also modulates the production effect. Participants were taught novel words composed of English sounds, but their sounds and sequencing varied in their frequency of occurrence within the English language. We predicted that adults would show a production effect with novel words comprised of frequent English sound patterns, as this is the kind of stimuli for which the production effect has been found in the past. If the production effect is sensitive to within-language effects, then novel words with infrequent English sound patterns could be treated more similarly to non-native items and show a reversal or attenuation of the production effect. On the other hand, if the production effect is only sensitive to native vs. non-native distinctions, then we expect novel words with infrequent English sound patterns to show the classic production effect.

2. Background

Studies have documented the effect of production on adults' retention and learning of words in their native language (Gathercole & Conway 1988; MacLeod et al. 2010; Forrin et al. 2012; Zamuner et al. 2016; Icht et al. 2020;), in their second language with phonologically familiar material (e.g., Ellis & Sinclair 1996; Kaushanskaya & Yoo 2011; Icht & Mama 2019) and in their native language with familiar accents (Grohe & Weber 2018).

Across these studies, adults typically show better recall and recognition for words that are overtly produced compared to words that are read-silently or only heard during a training or study phase (Hopkins & Edwards 1972; MacLeod et al. 2010; Ozubko & MacLeod 2010; Forrin et al. 2012; Mama & Icht 2016; among others). This effect is also seen when participants learn novel words (MacLeod et al. 2010; Krishnan et al. 2017). To illustrate, English-speaking adults who were learning Welsh as a second language showed improved retention of words and phrases that were overtly produced while learning compared to a learning condition without overt production (Ellis & Sinclair 1996). Similar results were found between a production learning group and a comprehension learning group in a study using an artificial language paradigm (Hopman & MacDonald 2018). This advantage for produced items has been explained by appealing to distinctiveness: producing a word aloud offers additional information that makes it more distinct from other words that have been only read or heard, creating a distinctive representation (MacLeod et al. 2010). This explanation for distinctiveness falls within the levels of processing framework (Craik & Lockhart, 1972). Distinctiveness is explained by the number of processing or encoding levels entailed by different actions. For example, only hearing a word (and not producing it aloud) involves one encoding process: auditory processing from hearing the word. Whereas producing a word involves additional processes: auditory processing and also articulatory processing (making articulators move to produce the word). At test, when learners have to retrieve these words, there is more information available to aid in retrieval for produced words than for silently read or heard words due to the higher number of processes involved when learning or memorizing the word. Furthermore, produced words not only have the additional articulatory information, but also the memory of having actively produced the word at training, making them more distinctive compared to other words which lack this additional information.

The studies cited above report a production advantage, measured in a variety of ways. However, other studies have reported effects where production appears to disrupt learning (Thorin et al. 2018; Zamuner et al. 2018; Baese-Berk 2019; López Assef et al. 2021). In these studies, cognitive load and availability of processing resources have been identified as a possible source for the learning disruption. Under a resource-sharing hypothesis (Baese-Berk 2019), learners have a limited number of cognitive resources that need to be split between tasks when performing more than one action, such as producing aloud in a perception learning task. In this scenario, resources have to be split between speech production and listening. On the other hand, without the production task all the resources can be solely focused on perception. When adult speakers are learning stimuli that is familiar to them, this dual task scenario does not pose a negative effect, thus allowing the production effect to emerge. This changes once manipulations create a learning task with higher difficulty, for example by testing non-native stimuli (Thorin et al. 2018; Baese-Berk 2019). In these cases, having to produce stimuli aloud can lead to increased cognitive load and thus, hinder learning.

One factor that has resulted in learning disruptions is phonological familiarity. Kaushanskaya & Yoo (2011) investigated whether the production effect could be expanded to non-native stimuli. Adult English speakers learned novel words that either followed or not the phonological structure of English. In both cases, half the words were trained under a production condition (repeating word aloud), and the other half were trained in a sub-vocal rehearsal condition (repeating the novel word silently). When trained on novel words with native sounds in produced or heard training conditions, adults showed better recall and recognition for items from the produced condition. However, this effect was reversed when participants were trained on novel words with non-native sounds, i.e., non-English vowels (close central unrounded /ɨ/ and high front rounded vowel /y/) and non-English consonants (retroflex stop /ʈ/ and uvular fricative /χ/). Kaushanskaya & Yoo (2011) frame their findings within the phonological loop in Baddeley & Hitch's (1974) model of working memory, as it is the component of working memory involved in the processing of phonological material and word learning (Baddeley 1986; Baddeley et al. 1998; Gathercole 2006; Gupta & Tisdale 2009; Baddeley 2012). In this model, long-term phonological knowledge supports the learning of native sound structures, whereas non-native sound structures are unlikely to have strong direct correspondences in long-term memory that can support learning. Overt rehearsal of phonologically familiar material enhances the engagement of the long-term phonological memory system, leading to the production effect, however, overt rehearsal of phonologically unfamiliar material leads to a reversal of the production effect or similar performance across different training conditions, suggesting that the phonological loop and long-term memory systems do not operate in the same way with phonologically unfamiliar material as with phonologically familiar material (Gathercole & Conway 1988, Kaushanskaya & Yoo 2011). Producing the words aloud caused learners to direct more attention to the phonological structure of the word. For native sounds, speech production highlights the similarities between the participants' native language and the novel words, allowing participants to rely on and use previously existing language knowledge to support word learning and to create robust, distinctive representations; thus, production facilitates learning. In contrast, for non-native sounds, this previous knowledge is most likely missing. Therefore, there is no information available to aid in learning, making the learning task more difficult. The production effect is explained by the number of levels of processing involved in speech production compared to other actions, which provide additional information about words. One of these additional levels, as mentioned previously, being articulatory processing. For native sounds, learners can use previous knowledge to aid their articulation and processing, whereas for non-native sounds, since this knowledge is missing, it is likely that articulatory processing does not provide as much help or is likely to create distinctive representations compared to that of native sounds, resulting in a disadvantage for produced items when compared to native produced items.

Other studies have also produced results in which saying test items aloud during training appears to disrupt learning. This disruption may also stem from other cognitive effects, as multiple factors can mediate the relationship between perception and production during learning. In cases where there is a high cognitive load, from factors such as the task and/or the stimuli, production has been found to disrupt and attenuate adults' learning. For example, Baese-Berk & Samuel's (2016) study found a learning disruption for a non-native sound contrast (novel fricative contrast) when participants had to produce the items during training, compared to adults trained on listening alone. Adults were trained on a novel fricative contrast using an ABX task. There was a perception-only group (hear-only) and a perception + production group that produced the training tokens aloud. At test, adults in the perception-only group showed successful discrimination of the novel sound contrast, whereas participants in the perception + production group did not, indicating that production during training disrupted learning. Experiment 2 followed the same procedure as the previous experiment, but researchers manipulated the amount of previous exposure to the non-native contrast, including participants who had some prior exposure to the novel contrast, to investigate whether this previous exposure would influence the learning of the contrast. Here both the perception-only and perception + production groups successfully discriminated the novel sound contrast. In Experiment 3, participants who had no experience with the novel sound contrast were asked to name letters rather than repeat the novel sounds. This resulted in a learning disruption, though to a lesser extent than in their first experiment, with performance landing between the perception-only and perception + production groups from Experiment 1. These results suggest that both linguistic and cognitive skills (such as attention and working memory) can affect the learning of a sound contrast. In the perception + production group, participants not only had to perceive and identify the new sound contrast but also had to simultaneously produce the non-native sound during training. These dual tasks could have imposed a heavier cognitive load, causing a bottleneck of processing resources that impacted learning (Ferreira & Pashler 2002; Baese-Berk 2019), resulting in a learning disruption for produced trials.

In the studies above, the effect of production was reversed when participants were tested on materials with non-native sound structures that were unfamiliar to the participants, which have no strong correspondences in long-term memory (Kaushanskaya & Yoo 2011). However, phonological familiarity can be defined not just by native versus non-native sounds but also with respect to the frequency and predictability of sounds and sound patterns within a language (Zamuner & Kharlamov 2016). For example, in English /t/ often occurs at the beginnings of words, whereas in the same position, /v/ is a less frequent sound. Phonotactic probability has been found to impact speech production, for example, lower phonotactic probability segments are more likely to be changed during speech errors than high phonotactic probability segments (Goldrick & Larson 2008). In studies testing memory and recall, advantages have been found for frequently occurring phonological patterns (e.g., Gathercole & Baddeley 1990; Vitevitch & Luce 1998; 1999; Thorn & Frankish 2005).

Non-native sound structures are unlikely to have representations (sound and/or motor representations), unless after a period of exposure. Within their discussions, previous researchers do not make a distinction of representations within the native category; however, extending their rationale, we can make a distinction between frequent and infrequent sound patterns. Frequent and infrequent sound patterns in the native language are all represented, though with differential strength or differential detail. Stimuli composed of less frequent sound patterns, however, have fewer existing correspondences. They are less frequent in the language, and thus, speakers have not only less exposure to these patterns, but also less previously stored lexical representations of words containing these patterns. When learning stimuli with frequent sound patterns, the material to-be-learned will overlap with many previously encountered stimuli at the segmental (e.g., phoneme) and suprasegmental (e.g., biphone, demisyllable, syllable) levels.

Thus, while previous work has shown that the strength and direction of the production effect depend on linguistic characteristics of the stimuli, work has been limited to manipulating native versus non-native stimuli. Moreover, in studies with novel words comprised of native sound patterns, the stimuli tend to have sound patterns that frequently occur in the language. Our goal was to examine the effect of within-language phonotactic frequency by comparing frequent and infrequent English sound patterns, exploring whether the production effect would be attenuated or reversed with attested sounds that occur infrequently in English.

Infrequent sound patterns differ from non-native sound patterns in that the former have established phonological representations in long-term memory, albeit weaker than frequent sound patterns. In contrast, previous to exposure, non-native sound patterns lack phonological representations. In the present study participants were trained on novel words that were either Produced or Heard, and which were comprised of either frequent or infrequent English sound patterns. Differences in learning have been found depending on native sounds' frequency and predictability, which are referred to as the sounds' phonotactic probabilities (see review Zamuner & Kharlamov 2016).

We predicted that items with frequent English sound patterns would show the traditional production advantage. We hypothesized that if the reversal and attenuation of the production effect is only limited to instances where the stimuli contain non-native sounds, triggered by processing difficulty, this would suggest that non-native is a category apart from native. In this case, frequent and infrequent stimuli are both within the native category, thus we would expect stimuli composed of infrequent English sound patterns to show the same production effect as stimuli with frequent English sound patterns, as both categories are comprised of sound patterns found in the English language (native). Furthermore, infrequent sound patterns could show a larger effect than non-words with frequent sound patterns.¹ Being more unfamiliar than frequent sound patterns could make them to be even more distinctive, as the non-words would

¹ We thank anonymous reviewer for this suggestion.

be perceived as different or unusual, resulting in differences in the size of the memory advantage for produced infrequent and frequent sound patterns.

On the other hand, if the contrast between native and non-native sound patterns is more nuanced, and the production effect is reliant on well-established sound patterns in the participants' native language, we might expect novel words with frequent English sound patterns to show the classic production effect, and for novel words with infrequent English sound structures to pattern with previous non-native results. This distinction between native and non-native sound patterns follows models such as PAM-L2 (Best & Tyler 2007), where the difficulty of learning L2 contrast is not mediated by a strict native and non-native category, but by how similar phonetically and phonologically the new contrast is to the L2. For example, L2 contrasts that are more similar to L1 phonemes should be easier to learn than L2 contrasts with more dissimilarities to the L1. Producing stimuli with infrequent English sound patterns during training might disrupt or not benefit learning, similar to producing non-native sounds (Kaushanskaya & Yoo 2011), as infrequent sound patterns are less familiar. This could lead to a reversed production effect (advantage for Heard items) or attenuation of the production effect (similar performance for both Produced and Heard items).

3. Methodology

3.1 Participants

Participants were 65 university students (11 males, 54 females, M age = 19 years, range = 18–24) who received partial course credit for participating. Participants were required to have self-reported normal or corrected-to-normal vision, normal hearing, and no history of language deficits. Participants were Monolingual English speakers and were asked to self-report their lifetime exposure to English (M = 94%, range = 70–100). Participants were randomly assigned to either the frequent English sound pattern group (n = 35, 5 males, 30 females, M English exposure = 94%) or infrequent English sound pattern group (n = 30, 6 males, 24 females, M English exposure = 94%). Fourteen additional participants were tested but their data were not included because of: equipment/experimenter error (5); no video for off-line coding (6), accuracy scores less than 50% correct (7, of whom 2 were tested on frequent English sound patterns and 5 were tested on infrequent English sound patterns). The analyses reported on in the paper were also rerun including the seven excluded participants with less than 50% correct responses, with the same results (see supplementary analyses on OSF at the following link: https://osf.io/qk5jy/?view_only=f03219b9e9224c4fb710ac56af3b3b32).

3.2 Stimuli

Stimuli were 32 novel words, 16 comprised of frequent English sound patterns, and 16 comprised of infrequent English sound patterns. There were 2 sets (Set 1 and Set 2) of novel words for each sound pattern frequency. Each set of 16 novel words comprised 8 rhyming pairs to make

Frequent English sound patterns		Infrequent English sound patterns	
Set 1	Set 2	Set 1	Set 2
hɛs	kɛs	ʃuɡ	wuɡ
ɡɛd	pɛd	θub	zub
nɪs	wɪs	noɪf	roɪf
bom	som	dʒɔf	zɔf
dæs	næs	loɪz	tʃoɪz
tɪb	mɪb	tʃɔb	jɔb
hæn	ɡæn	naub	ʃaub
lot	pot	faug	raug

Table 1: Novel word stimuli with frequent and infrequent English sound patterns.

Set 1 and Set 2 analogous (Table 1, audio files provided on OSF). Participants were trained on either Set 1 or Set 2. At test, the trained set served as ‘old’ items and the untrained set as the ‘new’ items. There were 4 lists for each frequency group, with counterbalancing for whether Set 1 or Set 2 was new or old and for the order of appearance of training conditions. The stimuli were recorded by a female native speaker of English and normalized for amplitude (70 dB). The stimuli were controlled for the frequency of the sound patterns in English, which was calculated using phonotactic probabilities, based on the Hoosier Mental Lexicon (Nusbaum et al. 1984), available online (Storkel & Hoover 2010) and in supplementary materials (Storkel 2013). Stimuli were formed by creating novel words that had frequent and infrequent Consonant + Vowel + Consonant (CVC) sequences in English. The average positional segmental sum (likelihood that an individual segment will appear in a given environment, C + V + C) was significantly different between the frequent and infrequent stimuli based on two-tailed independent t -tests ($t(15) = 17.84, p < .001$, novel words with frequent English sound patterns M positional segmental sum = 0.19, novel words with infrequent English sound patterns M positional segmental sum = 0.06). The stimuli were also controlled to have frequently and infrequently occurring Consonant + Vowel (CV) and Vowel + Consonant (VC) sequences. There was a significant difference in the biphone sum (likelihood that a sequence will appear in a given environment, CV + VC) ($t(15) = 6.88, p < .001$, novel words with frequent English sound patterns M biphone sum = 0.01, novel words with infrequent English sound patterns M biphone sum = 0.0005). The novel words also differed in the number of phonological neighbours ($t(15) = 10.85, p < .001$, novel words with frequent English sound patterns M neighbours = 20.56, novel words with infrequent English sound patterns M neighbours = 4.19).

3.3 Design

Participants were randomly assigned to either the frequent English sound pattern group or infrequent English sound pattern group. The experiment consisted of 16 training trials, followed by 16 test trials. On training trials, half of the eight novel words were Produced (4 novel words, 2 trials each), and half were Heard (4 novel words, 2 trials each). Before the experiment began, participants completed a practice task with real words (*apple, cherry, kiwi, lemon, mango, orange*). The practice task had the same design as the experiment and was used to familiarize participants with the design of the experiment. We defined items as Produced when they were presented by the computer and heard by the participants once before being produced (i.e., repeated back). We defined items as Heard when they were presented twice by the computer and heard by the participants. Listening to the items was chosen instead of reading-silently in order to allow us to keep track of whether the pronunciation of the trained items matched the test items. This is important as mismatches between the pronunciations at training and the auditory stimuli at test trials could affect recognition performance. In a reading-silently condition participants do not make overt productions, which would not allow to keep track of the quality of the trained items.

During training, participants were presented with a recording of a novel word. After 2000 ms, a prompt image depicted the appropriate response (see OSF for images). When participants saw a picture of a finger pointing at them, the appropriate response was to repeat the novel word. When they saw a picture of a woman gesturing “shh” with her finger over her lips, the appropriate response was to remain silent and hear a second recording of the novel word, which came 500 ms after the prompt-image. This controlled the number of novel word presentations in both the Produced and the Heard conditions, following previous protocols in the field (Zamuner et al. 2016; Icht & Mama 2019). Participants did not know the assigned training condition (Produced or Heard) until after they saw the prompt-image appear. The presentation of the conditions was pseudo-randomized, with no more than two consecutively trials from each condition.

At test, participants first completed an old/new recognition task where they were presented with the old/trained items and new/untrained items. At the beginning of each trial, a center fixation point appeared for 1000 ms, followed by an auditory stimulus of a novel word. There were 16 test trials (8 old/trained items: 4 Produced, 4 Heard during training; 8 new/untrained items which were minimal pairs to the Produced and Heard trained items). To illustrate based on **Table 1**, a participant in the frequent English sound pattern group who was trained on Set 1 received Set 1 (old/trained items) and Set 2 (new/untrained items) at test. After the old/new recognition task, there was a free recall task.

3.4 Procedure

The experiment was presented using Experiment Builder software (SR Research, Ottawa). Participants were seated in a sound-attenuated booth. Once training was finished, participants

completed the old/new recognition task in which they were instructed that they would hear a series of words, some of which had been taught and some of which were new. They were to press one key if they thought the word had been taught and was old and another key if they thought the word was new. Response keys on the keyboard were labeled and counterbalanced across participants for whether old or new corresponded to the left or right side of the keyboard. Additional auditory and visual recordings were made using a Zoom Q2HD Handy Video Recorder for off-line coding of training responses. At the end of the study, participants performed a free recall task by answering the question “What words do you recall learning?”.

4. Results

4.1 Coding

The audio-video recordings from training were coded off-line. This was to identify trials on which participants did not provide the appropriate response, i.e., produced an item from the Heard condition ($n = 6$, 2% of data), mispronounced an item from the Produced condition ($n = 36$, 14% of data). The corresponding test trials were removed prior to data analysis ($M = 0.65$ trials per participant). Note though that the pattern of results (accuracy and recall analyses) was the same even when these trials were included in the analyses (see supplementary analyses on OSF). Recordings were also used to code the items recalled during the recall task. Items that were correctly pronounced were coded as accurately recalled.

4.2 Old/New Recognition Task: Accuracy

Accuracy was based on old/trained test items (see **Table 2** for means by condition). Accuracy (correct, incorrect) was the dependent variable for mixed-effect logistic regression models performed in R (R Core Team) using the `glmer()` function from the `lme4` package (version 1.1-26; Bates et al., 2015). In each model, there were two fixed effects: Training Condition (Produced, Heard; deviation coded as $[-0.5, 0.5]$), English sound patterns (Frequent, Infrequent;

Analysis	Training Condition	Frequent English sound patterns	Infrequent English sound patterns
		Mean (SD)	Mean (SD)
Accuracy	Heard	0.90 (0.31)	0.92 (0.27)
	Produced	0.97 (0.19)	0.93 (0.26)
Recall	Heard	0.27 (0.45)	0.19 (0.39)
	Produced	0.46 (0.50)	0.16 (0.36)

Table 2: Condition means for proportion of accurate responses and proportion of accurate recalls, by Training Condition and English sound patterns.

deviation coded as $[-0.5, 0.5]$), and their interaction. We started with the most complex random-effects structure, including random intercepts for subjects and items, and random slopes for Training Condition (across subjects and across items). The random effects structure was reduced incrementally until models converged. Post-hoc comparisons of complex effects were done with the emmeans package, using Kenward-Rogers estimations for degrees of freedom and Bonferroni-corrections (Lenth 2020). Data and detailed code can be found at the OSF repository. There were no significant main effects or interactions (**Table 3**); however, learning of the novel words in both the frequent and infrequent English sound pattern groups was successful, as indicated by the over 90% accuracy responses in all conditions (**Table 2**).

Fixed Effects	Estimate	SE	z value	p-value
Accuracy				
Training Condition	-0.59	0.39	-1.53	0.13
English sound patterns	-0.21	0.41	-0.50	0.62
Train Condition * English sound patterns	1.22	0.77	1.45	0.15

Table 3: Results from the mixed-effects logistic regression model estimating Accuracy of novel words from Training Condition (Heard, Produced) and English sound patterns (Frequent, Infrequent).

Note. The final model had the following syntax specified in the lme4 package: `Accuracy ~ Train_Condition_dev * English_sound_patterns_dev + (1 | Item)`.

Data and code for the accuracy analysis are provided in the supplementary data on OSF. Data and code for an additional Reaction Time analysis, which resulted in no significant main effects or interactions, can also be found on OSF.

4.3 Recall Results

Results from the recall task are shown in **Figure 1**. Recall (yes, no) was the dependent variable for mixed-effect logistic regression models. The fixed and random effects structure and procedure for establishing the model was the same as with the accuracy analyses. There was a significant main effect of English sound patterns, as well as an interaction between Training Condition and English sound patterns. Results from the model are shown in **Table 4**. Post-hoc tests for the interaction showed that adults in the frequent English sound patterns group recalled more novel words that were Produced during training than Heard during training (Estimate = -0.90 , $SE = 0.29$, $z\text{-ratio} = -3.16$, $p = .002$). Adults in the infrequent English sound patterns group had no significant differences in recalled items from the Produced versus Heard conditions (Estimate = 0.34 , $SE = 0.37$, $z\text{-ratio} = 0.91$, $p = .36$). See **Table 2** for condition means.

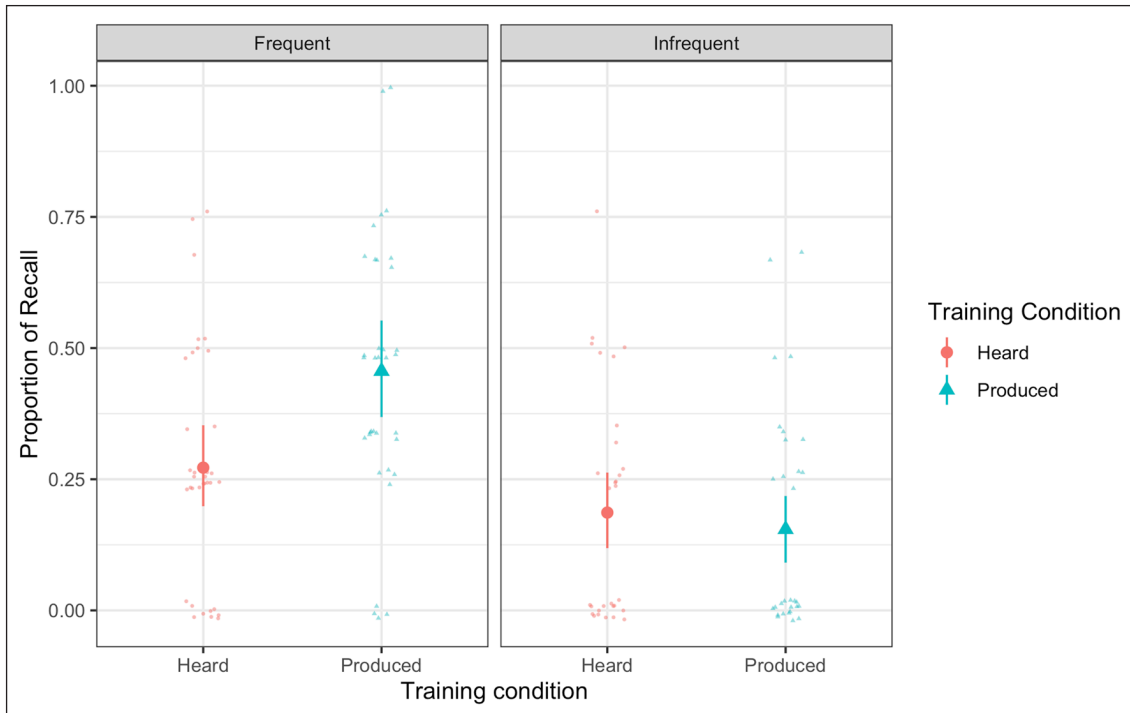


Figure 1: Proportion of recall by Train Condition and frequency of the English sound patterns of the novel words. Points are the condition means with error bars indicating 95% confidence intervals.

Fixed Effects	Estimate	SE	z value	p-value
Recall				
Training Condition	-0.28	0.23	-1.22	0.22
English sound patterns	-1.16	0.35	-3.35	< 0.001
Train Condition * English sound patterns	1.24	0.47	2.65	< 0.01

Table 4: Results from the mixed-effects logistic regression model estimating recall of novel words from Training Condition (Heard, Produced) and English sound patterns (Frequent, Infrequent).

Note. The final model had the following syntax specified in the lme4 package: `Recalled ~ Train_Condition_dev * English_sound_patterns_dev + (1 | Item) + (1 | Subject)`. The proportion of variance accounted for by the final model (pseudo- R^2) was calculated using the `r.squaredGLMM` function: fixed effects (marginal theoretical $R^2_m = 0.10$); fixed and random effects (conditional theoretical $R^2_c = 0.22$).

5. Discussion and conclusion

Adults continue to learn words in their native language across the lifespan. A myriad of factors influence how words are learned, including the manner in which words are learned and the linguistic characteristics of the stimuli. In the current study, we investigated whether there are

differential effects of production on the learning of novel words which varied in the frequency of their sound patterns. For novel words composed of frequent English sound patterns, a recall advantage was found for items that were Produced versus Heard during training (production effect). For novel words composed of infrequent English sound patterns, there was no difference in recall rates (attenuation of production effect). Thus, recall was modulated by training condition only for novel words with frequent sound patterns. Our results pose an important constraint to the distinctiveness account for the production effect. While previous studies had shown that the effect of production depends on the linguistic characteristics of the stimuli, our findings demonstrate that the reversal or attenuation of the production effect does not critically depend on the use of non-native sounds (Kaushanskaya & Yoo 2011; Cho & Feldman 2013, Baese-Berk & Samuel 2016; Cho & Feldman 2016), but also depends on the frequency of the sound patterns within the participants' native language. Our results are also not captured by the initial phonological loop account in Kaushanskaya and Yoo 2011, which was previously used to explain the contrast between native and non-native stimuli but did not make any distinction within native stimuli. There is more complexity in the production effect: completely unfamiliar, non-native sounds result in a reversal of a production effect, while infrequent but native sounds result in an attenuation of the production effect. The familiarity of the sound patterns also mediated overall recall rates, indicated by the main effect of English sound patterns: more items were recalled from the frequent English sound pattern group overall than from the infrequent English sound pattern group. This is also in line with previous results showing that frequent phonological patterns have an advantage in memory and recall (e.g., Gathercole & Baddeley 1990; Vitevitch & Luce 1998; 1999; Thorn & Frankish 2005). Note that although recall rates were relatively low across both groups and training conditions, learning of the novel words in both the frequent and infrequent English sound pattern groups was successful, with over 90% accuracy responses in the recognition task for all conditions (**Table 2**). This suggests that the disadvantage found for infrequent English sound patterns in the recall task did not stem from participants failing to learn the novel words.

While the difference in results between the recognition and recall task might seem unexpected, no production effect in the recognition task for both groups, whereas a production effect for frequent English patterns and an attenuation of the effect for the infrequent group in recall, previous studies have shown similar results. A difference in performance on recall and recognition tasks and the interaction with the production effect is also reported in Cho & Feldman (2016); however, not in the same pattern as the current study. In Cho & Feldman's Experiment 1, they found no interaction in recall rates for heard-then produced versus heard-only items when stimuli were presented in American-accented English versus Chinese-accented English (accent was a within-subjects measure). However, they did find an interaction on recognition rates in their heard-only condition: participants were more accurate at recognizing Chinese-accented

items compared to American-accented items. No difference was found on recognition rates for words presented in different accents from their produced-then-heard condition. It is difficult to draw conclusions across these studies as the tasks were not wholly the same, nor were the conditions. However, these types of differences illustrate that the strength and direction of the production effect is subject not only to the linguistic characteristics of the stimuli, but also based on the frequency to which the experimental material are presented and based on differences in the task.

One source for the difference in performance between recall and recognition in the current study could be the nature of the task. In our recognition task participants heard an auditory token of the stimuli and had to select whether the item had been studied or not but did not include having to overtly produce the token aloud. This is a more passive task compared to our recall task, in which participants had to remember the items without a visual or auditory aid, and then produce them aloud. A similar phenomenon is discussed by MacLeod et al. (2010) and Bodner & Taikh (2012), where the production effect mentioned to only surface in explicit memory tests. While the explicit and implicit memory task compared in the previous studies are not equal to our tasks, this supports the proposal that the characteristics of the testing task, such as the action performed in the task and the underlying mechanisms used during the task could affect the ability to detect a production effect. For example, recognition tasks are better than list-discrimination tasks to test the production effect. In a list-discrimination task participants are asked to indicate whether a word was part of one list or another in a previous study stage. Bodner and Taikh (2012) concluded that list-discrimination tasks are susceptible to influences from knowledge of the composition of each list and bias to attribute items that were not recognized to List 1 (or earlier lists), which can interfere with the goal of a production effect experiment.

It is possible that our recognition task did not require detailed or complex information to be activated, and just the memory of having heard the word may have been sufficient to complete the task, thus the ceiling effects in the recognition task. In the recall task participants must retrieve the representation of the novel words in order to produce them, requiring accessing specific information for each novel word, such as sound and motor information. For frequent sound patterns, producing the words reinforced this information, possibly creating a more robust representation and a production effect, resulting in produced items becoming distinctive compared to heard only items. For infrequent sound patterns, production did not provide the same benefit.

We did not find significant differences in recalled items from Produced and Heard, thus following the distinctiveness account for the production effect, neither training condition was distinctive compared to the other. It is possible that the additional information from production, that is believed to trigger distinctiveness and thus, the production effect, was not sufficient to create a distinctive representation for produced items. It is important to highlight that in

our study, infrequent patterns showed an attenuation of the production effect (no significant differences in recalled items from the Produced versus Heard conditions) and not a complete reversal, advantage for Heard items, as found with non-native patterns in Kaushanskaya & Yoo (2011). These findings support our prediction that the contrast between native and non-native sound patterns is not categorical, and that the production effect requires well-established sound patterns in the participants' native language. The size of the production effect depends on the familiarity of the tokens: The production effect surfaces with items comprised of frequent, native sound patterns, which is supported by long-term phonological knowledge, creating robust distinctive representations. An attenuation of the effect is found for infrequent, native sound patterns, which have less correspondences and are less familiar than frequent sound patterns, resulting in more difficulty for the production effect to emerge due to less information available to support learning and speech production, and a reversal of the effect (disadvantage for produced items) found for non-native sound patterns, which do not have representations in long-term phonological knowledge.

Unlike in previous studies with non-native stimuli, our observed attenuation of the production effect with infrequent English sound patterns cannot be attributed to disruptions caused by mismatches between the target and participants' non-target-like productions, e.g., *fub* mispronounced as *thub*. This is because if participants mispronounced a novel word during training, that item was removed from the analyses (results were the same when these items were included in the analyses, see supplementary materials). This is also similar to Baese-Berk & Samuel (2016)'s finding that producing letter names also disrupts perceptual learning, although to a lesser degree compared to when participants produced non-native sounds during training. They suggest that while some of the learning disruption for produced items likely stems from the mismatch between the auditory target and participants' non-target-like productions, part of the disruption also stems from a higher cognitive load caused by attention and task-switching effects (Baese-Berk & Samuel 2016). Thus, cognitive load can stem not only from linguistic factors, such as familiarity of the sounds, but can also be influenced by the type of task participants perform at test. Developmental factors can also play a role in the effect's direction. When using the same task with real words across a wide age range, younger children showed a reversed production effect, while older children were more likely to display a production advantage (López Assef et al. 2021).

It is not possible to tease apart the impact between linguistic and cognitive factors in the current experiment. Producing novel words with infrequent sound patterns likely required more processing resources because of differences in how frequent and infrequent sound patterns are represented (Warker & Dell 2006; Coady & Evans 2008). Frequently occurring sound patterns correspond to many lexical, phonological, and articulatory structures. Even though infrequent legal sound patterns also correspond to existing structures, the representations are less robust

compared to frequent legal sound patterns (Pierrehumbert 2003; Sosa & Bybee 2008; Munson et al. 2012; Sosa & Stoel-Gammon 2012). Alternatively, the effect of sound frequency may stem from the quality of the phonological representations maintained in the phonological store, which in turn influences the formation of new phonological and lexical representations (Thorn et al. 2002; Gathercole 2006).

Production was beneficial for recalling novel words comprised of frequent English sound patterns, but this was not found with novel words comprised of infrequent English sound patterns. This extends the previous results with unfamiliar and non-native sounds, to sound patterns within a native language. Furthermore, our results suggest that, for the production effect, the linguistic factors that mediate its strength or direction are not categorical (native and non-native being separate) but that there is a continuum encompassing both native and non-native categories: the more unfamiliar or infrequent items are, the more likely to observe a reversal or attenuation of the production effect. While our results are specific to accounts for the production effect, they are also relevant to other areas such as language processing and learning as they suggest that there are differences in the learning and recall of novel words with frequent and infrequent sound patterns. Even more, these are not only learned differently, but it suggests that they are also represented with different strengths in long-term phonological knowledge depending on how the words were learned. These differences in the phonological loop between frequent and infrequent stimuli are relevant as the phonological loop is involved in processing of phonological material and word learning. This could lead, for example, to different learning trajectories for words with frequent and infrequent sound patterns depending on how the words are learned (produced aloud or not). The current study follows language learning models such as PAM-L2 (Best & Tyler 2007), mentioned previously, in which native and non-native are not strict separate categories, and extends this idea to frequent and infrequent sound patterns, where infrequent or less familiar to L1 patterns are likely to behave like non-native stimuli than native stimuli.

In sum, our results reinforce the idea that the production effect depends on multiple factors, which can alleviate or increase task difficulty effects during learning or memorizing, allowing for phenomena such as production effect to occur or not. Further work examining the effect of different factors on task difficulty, and therefore the presence of the production effect, will be useful to shed light on the mechanisms behind not only the production effect but also how we create and access representations.

Data availability

For full transparency, the data that support the findings of this study are openly available in Open Science Framework at the following link: https://osf.io/qk5jy/?view_only=f03219b9e9224c4fb710ac56af3b3b32. Files include our stimuli, data, and analyses code for the main and supplementary analyses.

Ethics and consent

All requirements of the University's Research Ethics Board were observed.

Acknowledgements

This research was supported by Agencia Nacional de Investigación y Desarrollo (ANID, Chilean Government) under grant 72200366 awarded to Belén López Assef and Social Sciences and Humanities Research Council of Canada (SSHRC) grant awarded to Tania S. Zamuner. The authors thank Emma Arbuckle, Amélie Bernard, Eleanor Campbell, Brianna Kelly, Zeinab Kahin, Katherine Lam and Margarethe McDonald for assistance

Competing interests

The authors have no competing interests to declare.

References

- Baddeley, Alan D. & Hitch, Graham. 1974. Working memory. In Bower, Gordon H. (ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, 47–89. New York: Elsevier. DOI: [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Baddeley, Alan D. 1986. *Working Memory*. New York: Oxford University Press.
- Baddeley, Alan D. 2012. Working memory: Theories, models, and controversies. *Annual Review of Psychology* 63. 1–29. DOI: <https://doi.org/10.1146/annurev-psych-120710-100422>
- Baddeley, Alan D. & Gathercole, Susan & Papagno, Costanza. 1998. The phonological loop as a language learning device. *Psychological Review* 105. 158–173. DOI: <https://doi.org/10.1037/0033-295X.105.1.158>
- Baese-Berk, Melissa Michaud. 2019. Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics* 81(4). 981–1005. DOI: <https://doi.org/10.3758/s13414-019-01725-4>
- Baese-Berk, Melissa Michaud & Samuel, Arthur G. 2016. Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language* 89. 23–36. DOI: <https://doi.org/10.1016/j.jml.2015.10.008>

- Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Best, Catherine T. & Tyler, Michael D. 2007. Non-native and second-language speech perception: Commonalities and complementarities. In Munro, Murray J. & Bohn, Ocke-Schwen (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, 13–34. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/llt.17.07bes>
- Bodner, Glen E. & Taikh, Alexander. 2012. Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38(6). 1711–1719. DOI: <https://doi.org/10.1037/a0028466>
- Cho, Kit W. & Feldman, Laurie B. 2013. Production and accent affect memory. *The Mental Lexicon* 8(3). 295–319. DOI: <https://doi.org/10.1075/ml.8.3.02cho>
- Cho, Kit W. & Feldman, Laurie B. 2016. When repeating aloud enhances episodic memory for spoken words: interactions between production-and perception-derived variability. *Journal of Cognitive Psychology* 28(6). 673–683. DOI: <https://doi.org/10.1080/20445911.2016.1182173>
- Coady, Jeffrey A. & Evans, Julia L. 2008. Uses and interpretations of non-word repetition tasks in children with and without specific language impairment (SLI). *International Journal of Language & Communication Disorders* 43(1). 1–40. DOI: <https://doi.org/10.1080/13682820601116485>
- Craik, Fergus I. M. & Lockhart, Robert S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11. 671–684. DOI: [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Ellis, Nick C. & Sinclair, Susan G. 1996. Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology* 49(1). 234–250. DOI: <https://doi.org/10.1080/027249896392883>
- Ferreira, Victor S. & Pashler, Harold. 2002. Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(6). 1187. <https://psycnet.apa.org/doi/10.1037/0278-7393.28.6.1187>. DOI: <https://doi.org/10.1037/0278-7393.28.6.1187>
- Forrin, Noah D. & MacLeod, Colin M. & Ozubko, Jason D. 2012. Widening the boundaries of the production effect. *Memory and Cognition* 40(7). 1046–1055. DOI: <https://doi.org/10.3758/s13421-012-0210-8>
- Gathercole, Susan E. 2006. Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics* 27(4). 513–543. DOI: <https://doi.org/10.1017/S0142716406060383>
- Gathercole, Susan E. & Baddeley, Alan D. 1990. The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology* 81(4). 439–454. DOI: <https://doi.org/10.1111/j.2044-8295.1990.tb02371.x>
- Gathercole, Susan E. & Conway, Martin A. 1988. Exploring long-term modality effects: Vocalization leads to better retention. *Memory and Cognition* 16(2). 110–119. DOI: <https://doi.org/10.3758/BF03213478>

- Goldrick, Matthew & Larson, Meredith. 2008. Phonotactic probability influences speech production. *Cognition* 107(3). 1155–1164. DOI: <https://doi.org/10.1016/j.cognition.2007.11.009>
- Grohe, Ann-Kathrin & Weber, Andrea. 2018. Memory advantage for produced words and familiar native accents. *Journal of Cognitive Psychology* 30(5–6). 570–587. DOI: <https://doi.org/10.1080/20445911.2018.1499659>
- Gupta, Prahlad & Tisdale, Jamie. 2009. Word learning, phonological short-term memory, phonotactic probability and long-term memory: towards an integrated framework. *Philosophical Transactions of the Royal Society B* 364. 3755–3771. DOI: <https://doi.org/10.1098/rstb.2009.0132>
- Hopkins, Ronald H. & Edwards, Richard E. 1972. Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behaviour* 11(4). 534–537. DOI: [https://doi.org/10.1016/S0022-5371\(72\)80036-7](https://doi.org/10.1016/S0022-5371(72)80036-7)
- Hopman, Elise W. & MacDonald, Maryellen C. 2018. Production practice during language learning improves comprehension. *Psychological science* 29(6). 961–971. DOI: <https://doi.org/10.1177/0956797618754486>
- Icht, Michal & Ben-David, Nophar & Mama, Yaniv. 2020. Using Vocal Production to Improve Long-Term Verbal Memory in Adults with Intellectual Disability. *Behavior Modification* 45(5). 715–739. DOI: <https://doi.org/10.1177/0145445520906583>
- Icht, Michal & Mama, Yaniv. 2019. The effect of vocal production on vocabulary learning in a second language. *Language Teaching Research* 26(1). 79–98. DOI: <https://doi.org/10.1177/1362168819883894>
- Kaushanskaya, Margarita & Yoo, Jeewon. 2011. Rehearsal effects in adult word learning. *Language and Cognitive Processes* 26(1). 121–148. DOI: <https://doi.org/10.1080/01690965.2010.486579>
- Krishnan, Saloni & Watkins, Kate E. & Bishop, Dorothy V. M. 2017. The effect of recall, reproduction, and restudy on word learning: a pre-registered study. *BMC Psychology* 5(1). 1–14. DOI: <https://doi.org/10.1186/s40359-017-0198-8>
- Lenth, Russell. 2020. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.4. <https://CRAN.R-project.org/package=emmeans>.
- López Assef, Belén & Desmeules-Trudel, Félix & Bernard, Amélie & Zamuner, Tania S. 2021. A shift in the direction of the production effect in children aged 2–6 years. *Child Development* 92(6). 2447–2464. DOI: <https://doi.org/10.1111/cdev.13618>
- MacLeod, Colin M. & Gopie, Nigel & Hourihan, Kathleen L. & Neary, Karen R. & Ozubko, Jason D. 2010. The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36(3). 671–685. DOI: <https://doi.org/10.1037/a0018785>
- Mama, Yaniv & Icht, Michal. 2016. Auditioning the distinctiveness account: Expanding the production effect to the auditory modality reveals the superiority of writing over vocalising. *Memory* 24(2). 98–113. DOI: <https://doi.org/10.1080/09658211.2014.986135>
- Munson, Benjamin & Edwards, Jan & Beckman, Mary E. 2012. Phonological representations in language acquisition: Climbing the ladder of abstraction. In Cohn, Abigail C. & Fougerson, Cécile & Huffman, Marie K. (eds.), *The Oxford handbook of laboratory phonology*, 288–309. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199575039.013.0012>

- Nusbaum, Howard C. & Pisoni, David B. & Davis, Christopher K. 1984. Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report* 10. 357–376.
- Ozubko, Jason D. & MacLeod, Colin M. 2010. The production effect in memory: evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36(6). 1543–1547. DOI: <https://doi.org/10.1037/a0020604>
- Pierrehumbert, Janet B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46(2–3). 115–154. DOI: <https://doi.org/10.1177/00238309030460020501>
- Sosa, Anna Vogel & Bybee, Joan L. 2008. A cognitive approach to clinical phonology. In Ball, Martin J. & Perkins, Michael R. & Müller, Nicole & Howard, Sara (eds.), *The Handbook of Clinical Linguistics*, 480–490. Oxford UK: Blackwell Publishing Ltd. DOI: <https://doi.org/10.1002/9781444301007.ch30>
- Sosa, Anna Vogel & Stoel-Gammon, Carol. 2012. Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research* 55(2). 596–608. DOI: [https://doi.org/10.1044/1092-4388\(2011/10-0113\)](https://doi.org/10.1044/1092-4388(2011/10-0113))
- Storkel, Holly L. 2013. A corpus of consonant-vowel-consonant real words and nonwords: Comparison of phonotactic probability, neighborhood density, and consonant age of acquisition. *Behavior Research Methods* 45. 1159–1167. DOI: <https://doi.org/10.3758/s13428-012-0309-7>
- Storkel, Holly L. & Hoover, Jill R. 2010. An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods* 42(2). 497–506. DOI: <https://doi.org/10.3758/BRM.42.2.497>
- Thorin, Jana & Sadakata, Makiko & Desain, Peter & McQueen, James M. 2018. Perception and production in interaction during non-native speech category learning. *The Journal of the Acoustical Society of America* 144(1). 92–103. DOI: <https://doi.org/10.1121/1.5044415>
- Thorn, Annabel S. C. & Frankish, Clive R. 2005. Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(4). 729–735. DOI: <https://doi.org/10.1037/0278-7393.31.4.729>
- Thorn, Anabel S. C. & Gathercole, Susan E. & Frankish, Clive R. 2002. Language familiarity effects in short-term memory: The role of output delay and long-term knowledge. *The Quarterly Journal of Experimental Psychology: Section A* 55(4). 1363–1383. DOI: <https://doi.org/10.1080/02724980244000198>
- Vitevitch, Michael S. & Luce, Paul A. 1998. When words compete: Levels of processing in perception of spoken words. *Psychological Science* 9(4). 325–329. DOI: <https://doi.org/10.1111/1467-9280.00064>
- Vitevitch, Michael S. & Luce, Paul A. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40(3). 374–408. DOI: <https://doi.org/10.1006/jmla.1998.2618>
- Warker, Jill A. & Dell, Gary S. 2006. Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32(2). 387–398. DOI: <https://doi.org/10.1037/0278-7393.32.2.387>

Zamuner, Tania S. & Kharlamov, Viktor. 2016. Phonotactics and Syllable Structure. Lidz, Jeffrey L. & Snyder, William & Pater, Joe (eds.), *Oxford Handbook of Developmental Linguistics*, 27–42. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199601264.013.3>

Zamuner, Tania S. & Morin-Lessard, Elizabeth & Strahm, Stephanie & Page, Michael P. A. 2016. Spoken word recognition of novel words, either produced or only heard during training. *Journal of Memory and Language* 89. 55–67. DOI: <https://doi.org/10.1016/j.jml.2015.10.003>

Zamuner, Tania S. & Strahm, Stephanie & Morin-Lessard, Elizabeth & Page, Michael P. A. 2018. Reverse production effect: Children recognize novel words better when they are heard rather than produced. *Developmental Science* 21. 1–13. DOI: <https://doi.org/10.1111/desc.12636>

