

RESEARCH

# Morpho-phonological regularities influence the dynamics of real-time word recognition: Evidence from artificial language learning

Ashley Farris-Trimble<sup>1</sup> and Bob McMurray<sup>2,3,4</sup>

<sup>1</sup> Dept. of Linguistics, Simon Fraser University, Burnaby, BC, CA

<sup>2</sup> Dept. of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, US

<sup>3</sup> Dept. of Communication Sciences and Disorders, University of Iowa, Iowa City, IA

<sup>4</sup> Dept. of Linguistics, and DeLTA Center, University of Iowa, Iowa City, IA

Corresponding author: Ashley Farris-Trimble ([afarrist@sfu.ca](mailto:afarrist@sfu.ca))

---

Phonological alternations are attested in many of the world's languages. In production, these robustly generalize to new words and contexts, suggesting that talkers and listeners of a language have internalized them in some form. However, it is unclear whether listeners' knowledge of phonological alternations is used during real-time spoken word recognition. The present study asks whether listeners use knowledge of phonological alternations to modulate activation of competitor forms during real-time word recognition. In two experiments, listeners learned an artificial language with phonological alternations. We then used eye-tracking in the visual world paradigm to assess real-time spoken word recognition. We examined fixations to competitors that would be a match to the input because of the learned phonological alternation. Results showed that listeners do use phonological alternations in real time. Given a [t] ~ [d] alternation and an auditory stimulus with a surface [d], listeners fixated the [t]-competitor more than one that could not alternate with [d]. They were even able to generalize this to words that had not been learned in their alternated form. However, not all alternations showed the same pattern; listeners did not use a [d] ~ [z] alternation in the same way. Implications for various models of word recognition are discussed.

---

**Keywords:** phonological processing; lexical access; phonological alternations; word recognition; eye-tracking

---

## 1. Introduction

Phonological representations are undeniably intertwined with the process of spoken word recognition. A primary goal of research in phonology is to explain the patterns and regularities in a language's sound structure; conversely work on spoken word recognition seeks to characterize the processes that exploit those regularities to efficiently identify words.

These fields intersect in important ways when we consider phonological alternations—common phenomena whereby the phonetic realization of a word or morpheme can take multiple forms depending on context. Psycholinguistic research demonstrates that when recognizing a word like *bean*, words that overlap at onset like *beach* and *beak* compete for recognition (e.g., Allopenna et al., 1998; Marslen-Wilson, 1987; Marslen-Wilson & Zwitserlood, 1989). This derives in part from the fact that the auditory signal is ambiguous until several phonemes into the word (though coarticulation may assist in earlier disambiguation; Salverda et al., 2014). Phonological alternations, especially those that

neutralize phonological contrast, can potentially create lexical ambiguity and additional competition among candidates (e.g., Gaskell & Marslen-Wilson, 2001; Lahiri & Marslen-Wilson, 1991). Yet if listeners' real-time lexical processing dynamics were sensitive to the regularities governing these alternations (helping 'undo' them), this ambiguity may be reduced. The question of processing, however, cannot be entirely divorced from the issue of representation. The way in which listeners represent these alternations may influence the conditions under which this knowledge is deployed to help listeners process speech in the moment.

A variety of models describe whether and how knowledge of phonological alternations could be used during spoken word recognition (e.g., Gaskell & Marslen-Wilson, 2001; Gow, 2001; Lahiri & Marslen-Wilson, 1991; Pallier et al., 2001). It is also well documented empirically that listeners can use contextual patterns of assimilation (Gaskell & Marslen-Wilson, 1996) and phonological reductions (Brouwer et al., 2012), as well as lexically specific allophonic variation (Ranbom & Connine, 2007) in service of word recognition. However, much of this research has examined gradient or optional phonological processes and thus may not speak to the core issue of rule-based phonological alternations. For example, Gow (2003) showed that in cases of assimilation, the intended sound is often recoverable from a partially assimilated token. Because full-fledged morpho-phonological alternations are not optional and may be different at a phonological level (see Ernestus, 2011, for a review of this assumption), we might expect that such alternations have a different profile during word recognition.

Our main goal is to understand how phonological alternations affect word recognition. The experiments reported here highlight the joint roles of processing and representation during word recognition. Below we provide some background on theories of word recognition and each of these questions.

### **1.1. Word recognition and phonological alternations**

It is generally accepted that spoken word recognition is accomplished by a competition process that is immediate, parallel, and graded. As soon as listeners begin to receive an auditory stimulus, they immediately activate, in parallel, a set of candidate words that match the stimulus heard to that point. Over time, the candidate set is revised and updated as more auditory information accrues (Alloppenna et al., 1998; Dahan & Gaskell, 2007; Marslen-Wilson, 1987; Marslen-Wilson & Zwitserlood, 1989; Spivey et al., 2005). While classic psycholinguistic models (Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland & Elman, 1986) assume a match between surface and underlying forms (i.e., no alternations), a number of more recent models attempt to capture phonological alternations and reductions (Gaskell & Marslen-Wilson, 1998; Hume & Johnson, 2001; Lahiri & Marslen-Wilson, 1991). For our purposes, these phonological processes are most interesting when they create lexical ambiguity.

Lexical ambiguity occurs during word recognition when more than one representation matches the auditory signal. Classically this is considered to be largely the result of the temporary ambiguity as the onsets of a word match many candidates. However, lexical ambiguity can also arise from phonological neutralization. Consider a language with progressive voicing assimilation, such that post-nasal obstruents are realized as voiced (as in Kikuyu, Wembawemba, Japanese, and Greek: Hercus, 1986; Itô & Mester, 1986; Newton, 1972; Peng, 2003, 2008; see also Pater, 2004). We instantiate such a process in toy Language A in (1). This language has three words each with two prefixes: A vowel-final prefix [o] and a nasal-final prefix [an]. While the initial consonants of each word are realized faithfully following the prefix [o], the nasal-final prefix triggers an alternation in subsequent voiceless consonant. A voiceless stop like /t/ is realized as voiced [d] following

a nasal prefix, as in (1a) where the initial consonant of the stem alternates in [otib] and [andib]. Underlyingly voiced stops do not alternate, as is seen in (1b), nor do other voiced segments like fricatives (1c).

(1) Language A: Progressive voicing assimilation

	Underlying form	Intervocalic context	Post-nasal context
a.	/tib/ <sup>1</sup> 'dog'	[otib] 'one dog'	[andib] 'dogs'
b.	/diʃ/ 'coat'	[odiʃ] 'one coat'	[andiʃ] 'coats'
c.	/zif/ 'pickle'	[ozif] 'one pickle'	[anzif] 'pickles'

Though /tib/ and /diʃ/ have only moderate phonological overlap after the prefix [o], the assimilation process results in greater overlap after the nasal prefix: [andib] and [andiʃ] are now cohorts (candidates that are identical at word onset but disambiguated later). Thus the derived [d] in the alternated form creates lexical ambiguity in the surface form.

Models of word recognition differ in how they deal with phonological alternations and in the types of lexical representations they posit as the basis of word recognition. At the most abstract end of the spectrum, *lexical underspecification* models (e.g., Lahiri & Marslen-Wilson, 1991) posit that listeners store abstract lexical representations whose content is underspecified for predictable or default information (e.g., Archangeli, 1988; Archangeli & Pulleyblank, 1989; Dinnsen, 1996; Pulleyblank, 1986, 1988a, 1988b; Ringen, 1988; Stemberger, 1991; Stemberger & Stoel-Gammon, 1991). During word recognition, listeners determine whether features extracted from a surface form match the stored features; when features are underspecified, they neither support activation of the target word nor elicit a mismatch (Eulitz & Lahiri, 2004; Lahiri & Marslen-Wilson, 1991; Pallier et al., 2001; Scharinger et al., 2012; Wheeldon & Waksler, 2004). As a result, a listener can accept surface *greem* as a possible match for underlying *green*, because the coronal sound in /grin/ is underspecified for place and does not therefore mismatch with the [labial] feature of the surface form. Such models do not make use of the phonological context that triggers the assimilation, predicting that *greem* will not mismatch /grin/, regardless of whether the assimilation is licensed by a following labial (Wheeldon & Waksler, 2004).

In contrast, *phonological inference* models (Darcy et al., 2009; Gaskell & Marslen-Wilson, 1996, 1998; Lee & Pater, 2008; Mitterer et al., 2013) also posit abstract representations, but argue that listeners use phonological context to recover an abstract surface form, working backward from the phonetic form. For instance, upon hearing an assimilated form like *leam bacon*, the listener recognizes that the [b] in *bacon* may license assimilation of a preceding coronal nasal, and so the listener can identify the intended *lean*. Crucially, when the phonological context does not license assimilation (like in *leam gammon*), listeners do not 'undo' the phonological process (e.g., Coenen et al., 2001).

Finally, *exemplar lexicons* argue for a single-level lexicon in which all possible forms of a lexical item are stored without recourse to rules or processes to transform underlying forms to surface forms. Substantial research supports the idea that listeners store multiple exemplars of individual tokens (Goldinger et al., 1991; Nygaard et al., 1994; Palmeri et al., 1993; Pierrehumbert, 2001), and that this varied and detailed lexicon aids speech

<sup>1</sup> For clarity, we will list and refer to 'underlying forms' throughout, and those forms will contain the contrastive segments that are neutralized by rule. We recognize that in some theories presented here, such abstracted forms are not predicted to occur in the mental lexicon. Our use of them here is purely expository.

processing and word recognition (Goldinger, 1996, 1998; Johnson, 1997). Words are stored with fine-grained acoustic detail, and even sub-phonetic variation can impact lexical activation (Andruski et al., 1994; McMurray et al., 2002). As these are unanalyzed exemplars in memory, they may also contain elements of context that are associated with alternations. In this scenario, there is no need to reference abstract underlying forms during word recognition—the acoustic input can simply be mapped onto the closest phonetic representation (potentially including the alternation-triggering context as part of the similarity mappings).

These three types of models are not mutually exclusive. Models like Lexical Access from Spectra (LAFS) (Klatt, 1979) use stored spectra as the base unit of analysis (like exemplar models), but also incorporate phonological rules. Some models posit mechanisms for generalization over an exemplar lexicon, such that phonotactic patterns can be extracted (e.g., Frisch et al., 2001; Pierrehumbert, 2003). Others have pointed out that more general auditory or speech perception processes may do some of the work of phonological inference (Gow, 2003; Mitterer et al., 2006), though it is not clear how such approaches would handle complete rule-based alternations. Phonological inference may even work alongside more general auditory perception mechanisms (Clayards et al., 2015).

While the present study is not a definitive test of these models, this discussion makes it clear that any model of word recognition must account for listeners' abilities to recognize words that have undergone phonological alternations, especially complete phonological neutralization. However, as we describe below, no current study characterizes whether and how listeners recover from a complete neutralization induced by alternation.

## 1.2. Unanswered questions and the current study

The current study asks three questions:

- 1) Does listeners' knowledge of phonological alternations affect which lexical competitors are active during word recognition?
- 2) Do listeners generalize that process to new words?
- 3) Do the particular sounds on which the alternation is instantiated matter?

Each of these questions has implications for the process of word recognition and the content and structure of stored lexical representations.

First, we ask whether listeners use phonological alternations during lexical access. In the example Language A above, would listeners—who know that [t] and [d] alternate—activate a /t/-initial word upon hearing a [d] form? Studies of regressive inference suggest listeners can 'undo' a phonological process to arrive at an intended utterance (Gaskell & Marslen-Wilson, 1996, 1998, 2001). However, when the phonological process creates lexical ambiguity (as in the sentence *a quick ru[m] picks you up*, where *rum* could refer to a drink or could be an assimilated form of *run*), listeners activate only lexical items that match the surface form (*rum*, not *run*). Only with a biasing sentential context did listeners activate *run* (e.g., *It's best to start the day with a burst of activity. I think a quick ru[m] picks you up*). Without that context, listeners interpreted the form as underlyingly labial. However, it is not known whether *run* would have been more active than *rung*—which also mismatches the input, but is not licensed by either the surface form or the alternated form.

Likewise, underspecification models suggest that upon hearing a form that may be either the result of a phonological process or part of an underlying representation, listeners interpret the form as underlying, ruling out competitors whose underlying form does not

match the surface input (but could via an alternation). For example, in Bengali, nasal and oral vowels contrast, but that contrast is neutralized before a nasal consonant. In Lahiri and Marslen-Wilson's (1991) underspecification model, a vowel in a pre-nasal context is underspecified for nasality, as surface nasality is predictable. In a gating study, Lahiri and Marslen-Wilson found that upon hearing a surface nasal vowel, Bengali speakers interpreted it as underlyingly nasal (from a word like /kãp/), almost never identifying it as a surface nasal vowel derived by a phonological alternation (as in /kam/). They concluded that listeners interpret a marked surface structure as a specified underlying form, rather than as a form derived by applying a phonological rule.

Both models thus suggest that lexical ambiguity created by phonological neutralization might result in more activation for the non-derived interpretation of a [d] form than for a lexical item that must be derived to achieve the surface form (or even no activation for the derived word). However, these studies simply show that listeners show more activation for targets that underlyingly match the input, rather than having been created by an alternation. The design of the current study allows us to examine whether listeners *also* activate the alternated form, even if it is not the preferred interpretation.

Second, if listeners do use phonological alternations, we ask if this generalizes to new words. Models that include abstract representations (lexical underspecification and phonological inference) use non-lexical phonological mappings to link surface forms with the underlying structures from which they may have been derived. Such models predict listeners should generalize these mappings even to novel words (e.g., Snoeren et al., 2009). For example, in underspecification models, all coronal sounds are underspecified for place, and new forms would be expected to follow the same pattern. In a regressive inference model, listeners would expect that any coronal that precedes a labial might assimilate to that labial, even in new forms. In contrast, in exemplar models, whether listeners generalize (or by what mechanism they do so) is less clear. While speakers clearly can generalize phonological patterns (e.g., Finley & Badecker, 2009), it is unclear whether this generalization occurs during word recognition, and whether it shows the same time course as words that have been experienced with the alternation.

Finally, we ask if the phonological content of an alternation (i.e., the features that alternate) affects how listeners learn and use the alternation during word recognition: Are some types of alternations more likely to affect processing than others? We identified two primary factors that might play a role. The first factor is how similar the alternation is to other native language phonological processes. Several studies have shown that a listener's native language phonology influences lexical processing (Darcy et al., 2007; Darcy et al., 2009; Mitterer et al., 2013; Pallier et al., 2001; but c.f. Gow & Im, 2004; Mitterer et al., 2006). Listeners can also compensate for native language phonological regularities even when processing novel lexical items (Snoeren et al., 2009). What we do not know, however, is whether listeners use newly-learned phonological regularities during lexical access, and whether the phonological content of those regularities plays a role. A second factor is the phonological similarity of alternating sounds. It is possible that listeners are more easily able to compensate for alternations with similar alternating sounds, using the phonetic or articulatory similarity as scaffolding in building associations.

The present study addressed each of these questions using a combination of techniques. Participants were trained on an artificial language (like Language A above) instantiating a complete, obligatory phonological alternation. This allowed us to 1) use completely categorical alternations that participants would have had no history with; and 2) assess generalization by only exposing the participants to the alternation in the context of some words and testing on new words. We then evaluated lexical competition dynamics using



the visual word paradigm (VWP) (Allopenna et al., 1998; see Clayards et al., 2015 or Magnuson et al., 2003, for versions assessing word recognition in an artificial lexicon). Experiment 1 addressed our primary question of whether listeners use phonological alternations to modulate lexical access by comparing fixations to candidates that matched the auditory stimulus underlyingly to those that matched on the surface via an alternation. Experiment 2 addressed our second question, testing whether participants could generalize to new lexical items, and augmented the training regime to understand whether richer training might lead to different patterns of utilization of the rule. Both sets of experiments tested both voicing and manner alternations, to determine whether the specific alternation played a role (our third question).

Our experimental design allowed a number of predictions. When a speaker of Language A hears a word beginning with [an + d...], the signal is temporarily ambiguous. Experiment 1a asked whether listeners use their knowledge of the voicing alternation to activate forms like /antib/, in which the surface [d] derives from a phonological alternation. We presented listeners with a target like [andiʃ] and compared activation for /antib/ (surface [andib]) to the non-alternating /anzif/. Note that neither form completely matches the stimulus underlyingly, but one form, /antib/ (surface [andib]), matches on the surface until the final phoneme, whereas the other (/anzif/) mismatches early. If listeners are sensitive to the process which results in postnasal voicing, /antib/ should receive more consideration than /anzif/; if listeners are not sensitive to the phonological process, then both forms are equally poor matches to the stimulus. Experiment 1b repeated Experiment 1a with a different alternation, one of manner rather than voicing. Experiment 2 trained listeners on a subset of the words in alternating form; we again tested activation of competitors, but this time compared competitor activation in trained words vs. untrained words. If listeners show similar patterns of fixations in both trained and untrained words, generalization has taken place. Finally, Experiment 2 also tested a manner alternation. If the specific phonological content of the alternation is not relevant, then the results of the manner alternations are predicted to be a mirror image of those from the voicing alternations.

Much of the previous research on phonological patterns and word recognition has used tasks that do not allow comparison of activation of multiple competitors. Tasks like phoneme monitoring (Gaskell & Marslen-Wilson, 1998), and ERPs (Mitterer et al., 2006) allow for an examination of target word activation, but it is more difficult to assess activation dynamics for different classes of competitors. This can be done with gating (Lahiri & Marslen-Wilson, 1991) and cross-modal priming (Gaskell & Marslen-Wilson, 1996), though these methods are not well suited to our artificial lexicon paradigm. Gating, for example, requires multiple trials with different gates, and the small set of items would rapidly become highly salient. Perhaps more importantly, gating has been shown to be insensitive to activation for word forms that mismatch at onset (e.g., *candle* given the stimulus *handle*; Allopenna et al., 1998), an essential aspect of our design (e.g., activation for /tib/ after hearing /diʃ/). While cross-modal priming with various inter-stimulus intervals may be more effective, it was unclear whether priming could be established in an artificial lexicon, and to achieve enough power would have required a large number of repetitions of a small number of items which could introduce other effects like cumulative semantic interference.

Thus, we measured lexical activation using eye tracking in the VWP (Allopenna et al., 1998), which permits a more direct comparison of competitor activation over time. In this paradigm, listeners hear a word, phrase, or sentence and select an object from a visual array (typically on a computer screen). The items in the array are manipulated such that their names have phonological relationships. As the participant hears a word and selects

an item by clicking on it, his or her eye movements are monitored and recorded. These eye movements reflect the listener's interpretation of the auditory signal on a very fine-grained timescale. Averaged across time and over many trials, the fixation patterns give a picture of *when* and *how much* each item is fixated over time. The resulting fixation patterns have been shown to map closely on to models of lexical activation, like TRACE (Allopenna et al., 1998; McClelland & Elman, 1986). The VWP has the advantage of being a natural task with no meta-linguistic judgment necessary. We used it in conjunction with an artificial language learning paradigm to allow for control over process and form. For the current study, we focus on competitor fixations as a measure of competitor activation. We make the assumption that activation of newly-learned artificial words parallels that of natural words in the lexicon, but we come back to this issue in the discussion.

## 2. Experiment 1

Experiment 1 examined listeners' fixations upon hearing alternated forms that resulted from two different phonological alternations (a voicing alternation in Experiment 1a and a manner alternation in Experiment 1b). Each alternation was designed to be compared with a control condition with no alternation. Because the two experiments shared their design and methods, they are described here together, with relevant differences noted. For simplicity, we chose to analyze the results of each experiment separately, and the results section will discuss each experiment independently.

We compare voicing ([d] ~ [t]) and manner ([d] ~ [z]) alternations to address our third research question, whether the specific content of the alternation played a role. A similar voicing alternation occurs in North American English, the native language spoken by all of the participants, albeit in a different phonological context, in pairs of words like *bi[t]e* ~ *bi[r]ing*, where [r] is a voiced alveolar flap. The process generally applies intervocally preceding an unstressed vowel, and it is pervasive in English because of the large number of unstressed vowel-initial suffixes. While the flap is distinct from a [d], voicing is one of its primary cues (de Jong, 1998). On the other hand, the [d] ~ [z] alternation is not commonly found in North American English. These two alternations thus allowed us to test whether similarity to a native-language process played a role.

We also considered the articulatory and phonetic similarity of the two sets of alternating sounds. The sounds [t] and [d] are distinguished only by about 40–50 ms of VOT; thus, it may be fairly easy for the system to capitalize on the phonetic similarity to learn the alternation rule. On the other hand, [d] and [z] vary in a large number of cues: Overall duration, aperiodic energy, etc. Similarity is difficult to measure; studies differ on how confusable [d] and [t] are compared with [d] and [z]. In Miller and Nicely's (1955) study on consonant confusions, the confusability of [d] and [t] vs. [d] and [z] is highly dependent on signal-to-noise ratio (SNR) and frequency bands; in some conditions [d] is more often misheard as [t], in other conditions as [z]. In contrast, Wang and Bilger (1973) found that, across CV syllables in all SNR conditions, [t] and [d] were confused 123 times and [z] and [d] were confused 120 times—virtually identical results. While there is clearly uncertainty in the perceptual literature, this nonetheless raises the possibility that learning and use of alternations may differ across instantiations.

### 2.1. Methods

#### 2.1.1. Design

Three artificial languages were created: Two implemented phonological alternations (postnasal voicing: [t] alternated with [d], as in Language A above; or postnasal stopping: [z] alternated with [d]) and a third served as a control with no phonological alternations.

Each alternating language was compared independently with the control language which did not exhibit alternation.

Each language contained the same six triplets of lexical items, for 18 words. In each triplet, the onsets of the three words were /t/, /d/, and /z/. The triplets were formed such that other than the first segment, the only disambiguating information was the final consonant. Each triplet was combined with two prefixes, /o-/ and /an-/, one of which had the meaning ‘singular’ (it referred to a picture of a single object) and the other had the meaning ‘plural’ (referring to pictures containing more than one of the same object). The mapping between the specific prefix and the plurality was counterbalanced across participants (the o-singular or o-plural condition). Each participant thus learned 18 words, as a base form and with the prefixes [o] and [an].

Words were paired with pictures of everyday items (e.g., dog, coat, pencil). Each participant received the same word/picture pairings. We chose to pair words with real objects because we wanted the participants to imagine that they were learning a foreign language so they would be tolerant of a novel alternation not attested in their native language. Had the words been assigned to novel objects, participants may have assumed they were new English labels and might resist adding a new alternation rule.

In developing the artificial lexicons, realistic phonological alternations (i.e., alternations attested in real languages) were necessary. In many languages, a nasal followed by a voiceless consonant is disallowed (Pater, 2004). Likewise, nasal + fricative sequences are dispreferred in many of the same languages and may be repaired by epenthesis, deletion, or a change in manner (e.g., Padgett, 1994). The novel words in all four of the experiments reported in this paper were constructed to resemble Kikuyu (Peng, 2003, 2008), in which post-nasal voiceless consonants undergo voicing and post-nasal voiced continuants undergo stopping. The prefix [an] triggered the alternations. Participants were randomly assigned to the voicing alternation, stop alternation, or control group. Participants in all three language groups learned the same words in their base forms and o-prefixed forms. The groups differed, however, in the surface forms of the /an/-prefixed words. The voicing group learned that a surface [d] derived from either /an + t/ or /an + d/. As a result, within a triplet, the item with an underlying /d/ and the item with an alternating surface [d] were surface minimal pairs when prefixed with [an], differing only in their final consonant. The voicing group also learned a non-alternating /z/-initial form for each triplet, to highlight the fact that not all segments were merged in the post-nasal context. The manner group learned that a surface [d] derived from either /an + t/ or /an + z/, and they also learned a non-alternating /t/-initial form. The control group learned forms that never alternated. **Table 1** shows a sample of one triplet for each condition. The forms that surface as minimal pairs when prefixed with [an] are highlighted in gray.

### 2.1.2. Participants

Fifty-seven monolingual English-speakers between the ages of 18 and 35 participated in Experiment 1. Two were excluded from analysis because they did not complete the full set of test trials, and 2 because they did not make sufficient eye movements (i.e., the participants primarily fixated the center dot). The data from the remaining 53 participants were analyzed (18 in the voicing group for Experiment 1a, 19 in the manner group for Experiment 1b, and 16 in the control group). Participants received course credit or monetary compensation. All reported normal hearing and normal or corrected-to-normal vision.



**Table 1:** Sample triplet (surface forms), with and without alternations. Items that are temporarily ambiguous due to the phonological alternation are highlighted in gray.

Group	Rule description	Inputs (underlying form)		
		/dɪf/	/tɪb/	/zɪf/
voicing (Exp. 1a)	voicing of post-nasal /t/	an-dɪf	an-dɪb	an-zɪf
manner (Exp. 1b)	stopping of post-nasal /z/	an-dɪf	an-tɪb	an-dɪf
control	None	an-dɪf	an-tɪb	an-zɪf

### 2.1.3. Auditory stimuli

The 18 novel words (see Appendix A.1) were composed of English phonemes and complied with English phonotactic restrictions. These 18 words were constructed from six sets of three novel words. Within an item-set, the initial vowel was always identical, and there was one /d/-, /t/-, and /z/-initial novel word. Across item-sets, the initial vowel was unique. Half of the item-sets used monosyllabic words and the other half used disyllabic words.

Stimuli were recorded by a female native monolingual English speaker in the frame “He said \_\_\_.” Stress was placed on the first (or only) syllable of the base word, and not on the prefix. All items were recorded in full (each base word, prefixed word, and alternating word, if applicable, was recorded). We chose to do this, rather than recording a single prefix and splicing it on to each base word, because we wanted the phonological alternations to be as natural as possible. Words were recorded directly to digital at 44,100 Hz in a sound-attenuated booth with a Kay Elemetrics CSL 4300B and a head-mounted XLR microphone. The best exemplar of each word was selected and excised, and 50 milliseconds of silence were added to the onset.

### 2.1.4. Visual stimuli

Clipart pictures of everyday items were assigned to the eighteen words (see Appendix A.2). There were two pictures of each item, one picturing it in the singular (e.g., *one dog*) and one in plural (e.g., *two dogs*). Pictures were chosen to be iconic exemplars, and all were roughly equivalent in visual saliency. Pictures were vetted by a group of lab members with extensive VWP experience, and were edited to ensure roughly equal visual salience.

### 2.1.5. Procedure

Experiment 1 took place in two 1–1½ hour sessions spaced one or two days apart, following Magnuson et al. (2003). On the first day, there were two phases. In the first, participants learned the novel words in isolation, to teach them the underlying form to a reasonable degree of accuracy before introducing the prefixes. On each trial, four pictures appeared on the computer screen and one word was played. The participant clicked on the picture that they thought matched the word and received feedback contingent on their response: A green box appeared around the correct answer after the participant’s selection, and participants heard a short buzz when they made the wrong selection. No other explicit instruction was provided. During this phase, all pictures were singular. Each word was presented as the auditory target 11 times, for a total of 198 trials.

In the second phase, participants learned the prefixed forms of the words (and the alternation rule if present). The four pictures on the screen were composed of two pairs (two target items each in a singular and plural form; e.g., one /dɪf/, two /dɪf/’s, one /zæf/, and

two /zæf/'s). Each word was presented six times with the [o] prefix and six times with the [an] prefix for a total of 216 trials.

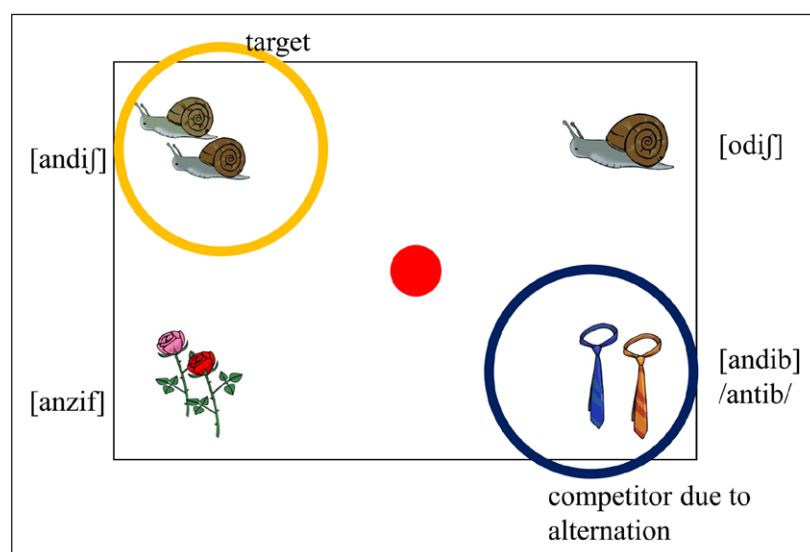
In the second day of the experiment, participants performed another training task followed by a testing task. The training task was a shortened version of the second phase of day one training: Each word presented three times in each of the two prefixed forms for a total of 108 trials (18 items  $\times$  2 prefixes  $\times$  3 repetitions).

Testing occurred immediately following this shortened training. It was during this task that participants' eye movements were recorded. The task in the testing portion was the same as the training task but with no feedback. This time, the three pictures from a triplet always appeared together in the testing phase and always matched in number (singular or plural). The fourth picture was the opposite-number form of one member of the triplet. A sample testing screen is shown in **Figure 1**. Each trial started with the presentation of a small red dot at the center of the screen, and the four pictures. After 500 ms, the dot turned blue and participants clicked it to play the auditory stimulus. The dot disappeared and participants then clicked on the picture that matched the word they heard. The short pre-scanning period was meant to minimize the likelihood that eye-movements were driven by visual search (since the participants were more likely to know what was on the screen and where it was located). Clicking on the dot to start each trial ensured that the mouse (and likely the gaze) was not located at any of the objects at the trial's beginning.

Each of the 18 items occurred as the auditory stimulus 16 times with each of the two prefixes, yielding 576 trials (18 items  $\times$  2 prefixes  $\times$  16 repetitions). These 576 trials were divided into critical trials (75%) and filler trials (25%). That is, of the 16 repetitions of a given word, 12 were critical trials (in which the target was a member of a matching triplet which all shared the same number prefix), and four were filler trials, in which the item had no same-number competitors on the screen (and thus the prefix alone was disambiguating). The inclusion of filler trials ensured that participants had to attend to the prefix as well as the base word.

#### 2.1.6. Eye tracking

Eye movements were recorded with an Eyelink 1000 eye tracker (SR Research). Participants were seated in front of the screen, and their head was supported by a fixed chin and



**Figure 1:** Annotated sample testing screen (participants saw no text).

forehead rest. The standard nine-point calibration occurred immediately before the testing phase began. Eye position was sampled every 4 milliseconds. A drift correction procedure was run every 31 trials to correct for small drifts in the calibration; the participants were recalibrated at this point if necessary. The Eyelink 1000 automatically divides the recording into fixations, saccades, and blinks. In post-processing, fixations were combined with the immediately preceding saccade to form a ‘look.’ Any look that fell within a 300x300 pixel image or a surrounding 50-pixel buffer (to allow for drift or slight error in the eye-track) was considered a look to that item. This additional 50 pixels did not result in any overlap among the regions of interest.

## 2.2. Results: Experiment 1

We begin by examining accuracy at test to document that people learned the words and prefixes. Next we examine fixations to the competitor when the targets were /d/-items in an alternating context (after the an-prefix) and a non-alternating context (after the o-prefix) to address our primary questions about how experience with a phonological alternation affects real-time spoken word recognition.

### 2.2.1. Accuracy

**Table 2** shows accuracy broken down by language group for each type of stimulus. Overall performance was quite strong, averaging around 94% correct. Data in each of the alternation conditions were analyzed separately, and complete results are reported in Appendix B.1. These generally showed that performance was lower for the words that began with alternating phonemes in the corresponding alternation group (i.e., /t/ in the voicing group, /z/ in the manner group).

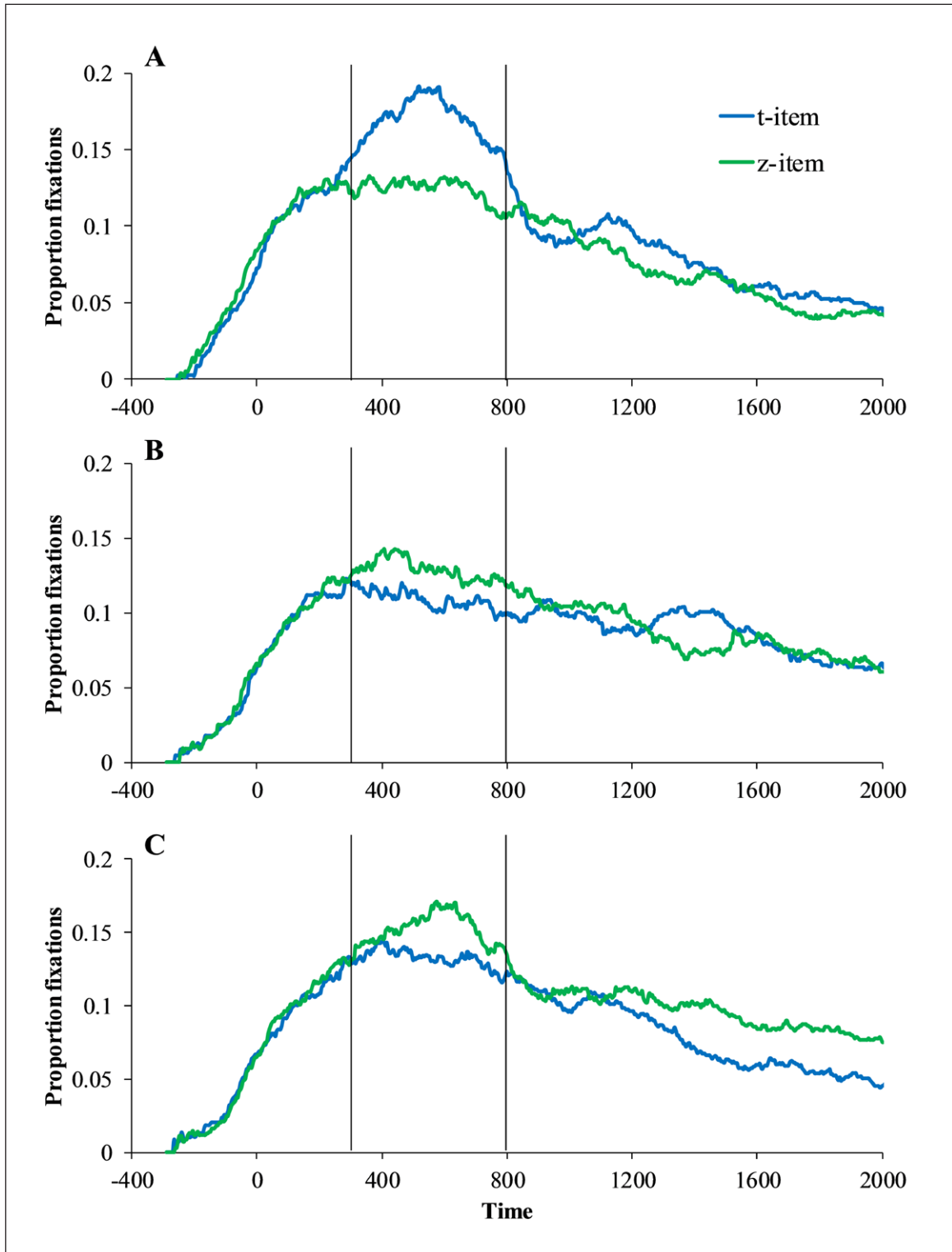
### 2.2.2. Experiment 1a: Fixations

We next turn to our primary analysis: The effect of alternation on real-time lexical competition dynamics. To assess this, we examined fixations during experimental trials in which the auditory stimulus was prefix + /d.../ (i.e., a /d/-item with the /an/ prefix or the /o/ prefix). We focused on those trials because they were the only trials that allowed us to compare fixations to the two *competitors* (/t/-item and /z/-item) across groups within a given trial. Here, if listeners had learned the alternation, we would expect to see heightened fixations to the /t/- item in the voicing group for the /an/ prefix (since after training, the /t/-item was a possible source of an [an + d...] surface form), but not in the control group and not in the case of the /o/ prefix for either group.

We started by computing the proportion of trials on which participants were fixating each competitor at each 4 ms time slice, divided by group (**Figure 2**). In this figure and all subsequent analyses, time was shifted for each item such that 0 ms is the offset of

**Table 2:** Average accuracy (standard deviation) by group for Experiment 1.

	/an/					/o/				
	d	t	Z	filler	total	d	t	z	filler	total
<b>control</b>	.946	.943	.954	.957	.952	.944	.954	.946	.954	.951
<b>group</b>	(.039)	(.088)	(.052)	(.031)	(.054)	(.051)	(.063)	(.065)	(.037)	(.058)
<b>voicing</b>	.933	.867	.955	.938	.928	.939	.951	.960	.934	.942
<b>group</b>	(.074)	(.116)	(.047)	(.054)	(.078)	(.066)	(.052)	(.035)	(.058)	(.060)
<b>manner</b>	.928	.948	.817	.937	.918	.937	.944	.974	.950	.951
<b>group</b>	(.073)	(.065)	(.121)	(.068)	(.096)	(.048)	(.071)	(.033)	(.039)	(.056)



**Figure 2:** Proportion of fixations to /t/- and /z/- items when target is an +/d/- item in Experiment 1. **Panel A:** voicing group; **panel B:** control group; **panel C:** manner group. Black lines indicate analysis time window.

the prefix/onset of the lexical item. About 200 milliseconds after this point, fixations to the target (not shown) begin to diverge from fixations to the competitors. This point represents the earliest point at which eye movements could be driven by the post-prefix signal, as it takes about 200 ms to plan and launch an eye movement (Viviani, 1990). Fixations to the competitors peak at around 400–500 ms, and then begin to decrease

as the candidate set is narrowed and the target is selected. Panels A and B of **Figure 2** show competitor fixations in the voicing and control groups, respectively. In the analyses below, the dependent measure is the average the proportion of looks to each item between 300 and 800 ms (i.e., area-under-the-curve). We selected this window (marked with the vertical black lines in **Figure 2**) after visual inspection of the data because it encompasses the period in which participants made the most fixations to competitor items. It was not possible to set the window in advance of data analysis; eye-tracking on newly-learned novel forms with phonological alternations is not common in the literature, and we were not sure what time course to expect with regards to activation of such forms.

These area-under-the-curve estimates were used as the dependent variable in a linear mixed effects model using the LME4 package (version 1.1-7) of RStudio (version 0.99.484). Because the fixation data were proportions (bounded by 0 and 1), we employed an empirical logit transformation on all fixation data. As some of the data were zero (indicating no fixations to a competitor during the time window), we added the equivalent of ‘half a look’ (i.e., half the duration of the average look over the course of the experiment) to each data point to avoid problems resulting from zero values. The model included itemtype (/t/-item = 0.5, /z/-item = -0.5), prefix (/an/ = 0.5, /o/ = -0.5), and languagegroup (voicing = 0.5, control = -0.5) as fixed effects. We did not include the effect of o-condition (o-singular vs. o-plural) in the final model, as it did not offer a significantly better fit. Critically, if the voicing group used the alternation to activate competitors, we expected to see a significant three-way interaction between itemtype, prefix, and group, such that the voicing group would make more fixations to the /t/-item than the /z/-item only when it was in an alternating context (i.e., preceded by the /an/ prefix), and the control group would not show any differences in fixations to either item with either prefix.

Prior to examining fixed effects, we examined models with various random effects structure to determine the best random effects structure for the data. We followed Matuschek et al. (2017) by starting with the simplest model and including random slopes only if they significantly improved model fit (as determined by a chi-square test of model comparison). We did not consider random effects of item as there were only six items. However, as we were concerned that different item-sets may have elicited different amounts of fixations overall, data were aggregated within an item-set, and the average fixations to the unrelated items (during that same time period) were centered and used as a covariate (as in McMurray et al., 2014). This examines fixations as a function of experimental condition *over and above* fixations in general, and can be particularly useful for controlling for differences in the likelihood of fixating anything (which can differ in special populations, or in cases of heightened uncertainty). The final model (2) only included random intercepts of subject, as adding a random slope of itemtype did not improve model fit.

$$(2) \quad \text{fixations}_{\text{competitor}} \sim \text{itemtype} * \text{languagegroup} * \text{prefix} + \text{fixations}_{\text{unrelated}} + (1|\text{subject})$$

P-values on fixed effects were computed by using the Satterthwaite approximation for df, as implemented in the lmerTest package (ver. 2.0-29) of RStudio.

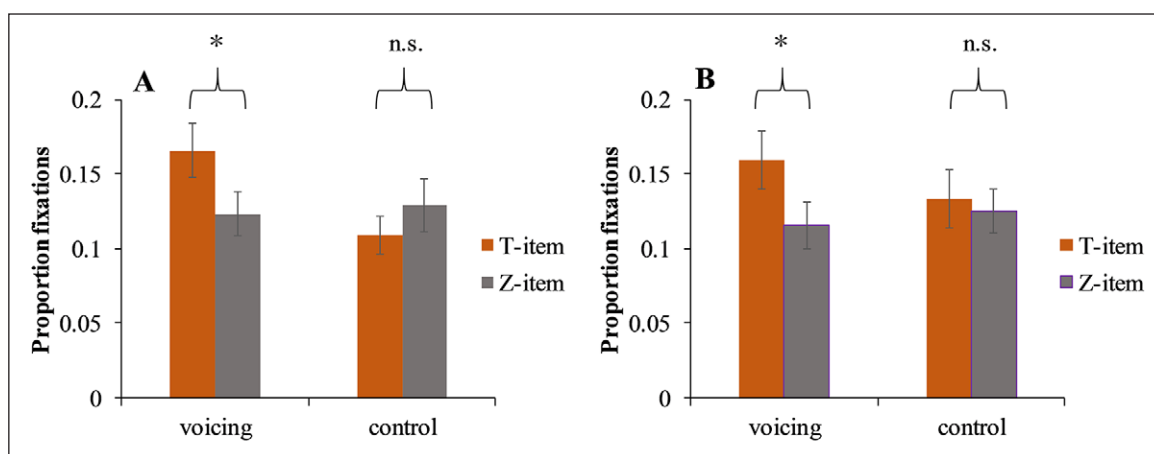
Results of the model are shown in **Table 3** and **Figure 3**. There was a main effect of itemtype such that overall, participants fixated the /t/-item more than the /z/-item, and an interaction between itemtype and group ( $p = 0.00193$ ). There was no main effect of prefix, nor did it interact with the other variables.

These results are reflected in **Figure 3**, which shows a difference in the predicted direction for the voicing group but no difference for the control group. Contrary to our



**Table 3:** Mixed effects model examining looks to competitor objects.

Effect	B	SE	T	df	p
Item type	0.122	0.054	2.28	773.9	0.02311*
Language group	0.094	0.147	0.64	31.2	0.53
Prefix	-0.038	0.065	0.59	217.8	0.56
Unrelated looks (covariate)	-0.164	0.034	-4.77	802.2	0.000002*
Item type × group	0.277	0.089	3.11	773.9	0.00193*
Item type × prefix	0.153	0.107	-1.43	773.9	0.15
Group × prefix	0.040	0.105	0.38	268.2	0.70
Item type × group × prefix	.008	0.178	0.05	773.9	0.96

**Figure 3:** Fixations to /t/-item and /z/-item as a function of languagegroup (averaged across o-conditions) when target is a [d]-item in an alternating context (an-prefix) (**Panel A**) or a non-alternating context (o-prefix) (**Panel B**) in Experiment 1a.

predictions, however, this pattern held for both prefixes (the interaction with prefix was not significant). To evaluate this further, we conducted post-hoc tests which examined the fixations to each item type separately as a function of group (but used an otherwise identical model structure). These indicated that the voicing group fixated the /t/-item more than the /z/-item,  $t(411.4) = 4.0, p = .00007$ , but the control group fixated the two items equally,  $t(365.6) = -.80, p = .43$ . Moreover, the voicing group fixated the /t/-item more than the /z/-item in both the alternating context,  $t(195.3) = 2.8, p = 0.00564$ , and the non-alternating context,  $t(195.2) = 2.9, p = 0.00474$ . This provides clear evidence that when hearing [d], listeners in the voicing group biased lexical competition toward /t/-initial items (which could have undergone alternation) over /z/-initial items, indicating some sensitivity to the training language. However, they did this equally for both contexts, suggesting a fairly coarse solution to this problem.

### 2.2.3. Experiment 1b: Fixations

We analyzed fixations as above. Panel C of **Figure 2** shows competitor fixations in the manner group. The relevant comparison is still looks to the /t/-item vs. the /z/-item by context; the only difference in Experiment 1b is that the manner group should make more fixations to the /z/-item (unlike Experiment 1a where the voicing rule led to more fixations to /t/). We used the same approach to model selection as Experiment 1a and arrived at an identical model (fixed effects of item type, language group, and prefix; random intercepts on participant; and unrelated items serving as a covariate).

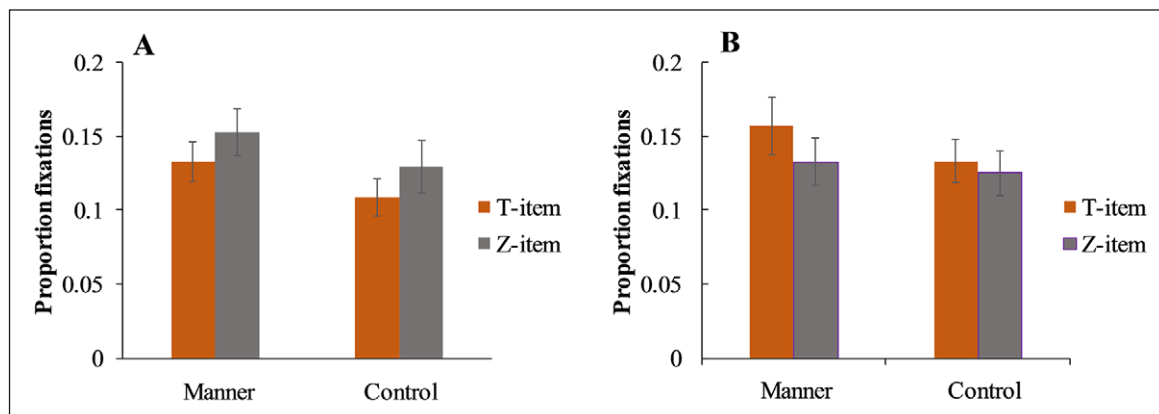
Results of the model are shown in **Table 4** and **Figure 4**. Unlike in Experiment 1a, there was no significant interaction between item type and language group ( $p = 0.37$ ), though there was an interaction between item type and prefix ( $p = .0411$ ). This interaction appears to be the result of the fact that both groups looked slightly more to the /z/-item when it was preceded by the /an/-prefix and slightly more to the /t/-item when it was preceded by the /o/-prefix (**Figure 4**), though there was no significant difference in either case (/an/-prefix:  $t(382.5) = 1.5, p = .12$ ; /o/-prefix:  $t(382.1) = 1.1, p = .25$ ).

### 2.3. Discussion

Experiment 1 indicated that even with a relatively short training time, learners could acquire novel words and new phonological alternations effectively, though both alternation groups showed lower performance for the alternating forms, suggesting that the ambiguity involved with alternations slows learning. More importantly, the pattern of fixations in Experiment 1a revealed that upon hearing [d] in an alternating context ([an + d...]), participants who had learned a voicing rule were more likely to activate an item in which [d] alternated with [t] than participants who had not learned a voicing rule. However, those participants were also more likely to activate an underlying /t/ item upon hearing [d] in a non-alternating context ([o + d...]). Taken together, these results suggest that listeners have learned the [t] ~ [d] alternation and are willing to activate /t/ forms upon hearing the surface [d]. However, they have adopted a fairly coarse solution to this, and are willing to entertain both items even outside of the context for alternation.

**Table 4:** Mixed effects model examining looks to competitor objects in Experiment 1b.

Effect	B	SE	T	df	p
Item type	-0.007	0.054	0.14	797.2	0.89
Language group	0.146	0.127	1.15	32.6	0.26
Prefix	-0.096	0.063	1.53	295.0	0.13
Unrelated looks (covariate)	-0.170	0.035	4.84	823.1	0.000002*
Item type × group	0.080	0.090	0.89	797.2	0.37
Item type × prefix	-0.220	0.108	2.05	797.2	0.0411*
Group × prefix	-0.046	0.102	0.45	379.2	0.65
Item type × group × prefix	-0.091	0.179	0.51	797.2	0.61



**Figure 4:** Fixations to /t/-item and /z/-item by rule group (averaged across o-conditions) when target is a /d/-item in an alternating context (an-prefix) (**Panel A**) or a non-alternating context (o-prefix) (**Panel B**) in Experiment 1b. All individual comparisons were non-significant.

Conversely, Experiment 1b suggests that listeners were not able to use the manner alternation during lexical access. While listeners did fixate the /z/- more than /t/-items numerically, this was not significant, and was not larger than the bias toward /z/ in the control group. These results do not support an analysis in which all listeners simply accept [d] as a potential alternate for /t/, but not for /z/. If that were the case, then we also expect all groups (including the manner and the control groups) to fixate the /t/-item more than the /z/-item, which is not what we found. We will return to this in the general discussion.

In Experiment 1a, though, participants' behavior could be explained primarily by surface forms, without explicit reference to alternations. That is, listeners learned that some meanings were associated with both surface [t] and surface [d], and upon hearing a surface [d], they activated any competitor that could be realized with a [d], regardless of context and of underlying form. This would not require explicit use of a phonological rule during processing—the effects would emerge solely during learning as these phoneme-to-word mappings are formed. While this indicates that listeners do activate words that are competitors as the result of (at least some) phonological alternations, it does not tell us the mechanism by which this occurs, or whether knowledge of the alternation affects lexical access. In Experiment 2, we turn to the issue of generalization to begin to answer this question.

### 3. Experiment 2

The primary goal of Experiment 2 was to examine generalization of the phonological rule to new items. Thus, Experiment 2 held out a subset of the novel words for use in later tests of generalization. Participants initially learned the full set of words, but only some of those were heard in their alternated (post-nasal) form during training. After training, we again compared fixations to the /t/-item and /z/-item, and asked whether this differed in words that were trained with the alternation (the trained items) or only without it (generalization items).

Additionally, given that Experiment 1 did not appear to show evidence for context sensitivity, we made a second change to see if we could encourage abstraction of a more flexible (context dependent) rule. While Experiment 1 used a fairly small set of words (six triplets), Experiment 2 increased the total number of items presented during training. This meant that listeners would have fewer repetitions of each item, but these additional items might offer robust evidence for the alternation rule (illustrating that it applied to more words and was thus more likely to be a rule than an item-specific effect).

We did not include a control group in this study. Because Experiment 1 found no evidence of learning for manner alternations, we used the manner alternation group as a comparison for the voicing alternation. Here, evidence for learning would appear as more fixations to /t/ than /z/ (in the voicing alternation group) and either equal fixations or the reverse in the other group (an interaction of alternation group and item). The degree to which this interaction is moderated by the prefix context would suggest contextual dependency (which was not observed in Experiment 1), and most critically, if this interaction is weakened in generalization words, this would suggest an inability to generalize the rule to new words.

#### 3.1. Methods

##### 3.1.1. Design

Experiment 2 used thirteen t/d/z triplets: The six from Experiments 1, plus seven new ones<sup>2</sup> (see Appendix A.1). Not all of the items were trained with prefixes. Consequently for

<sup>2</sup> We piloted Experiment 2 with different numbers of item-sets; we found no generalization when we used only eight sets. We settled on thirteen sets as a number that participants could learn and that allowed for generalization.

10 item-sets (henceforth the *trained* items), participants learned the unprefix form, and forms with both the [o] and [an] prefixes; for the remaining three item-sets (*generalization* items), participants learned the unprefix form, but neither prefixed form.

We were concerned that adding so many more items would result in lower accuracy, complicating our ability to use the fixation measure (which relies on correct trials). Thus, for each participant, three of the trained sets were presented more frequently than the others. These sets were used at test (along with the three generalization sets). The additional seven sets were trained at a lower frequency (fewer repetitions), and did not appear at test. Generalization sets were trained at the same frequency as the trained set, but only in their unprefix forms.

We made two other minor changes to the design of Experiment 2. First, based on anecdotal feedback from participants in Experiment 1, we worried that we confused participants by teaching them a base form paired with a singular picture and then later a prefixed form paired with the same singular picture. In Experiment 2, the prefixes (o- and an-) indicated specific numbers—two of something and three of something, with the unprefix form indicating singular. Participants were divided into two groups for the meaning of the prefixes /o/ and /an/ (whether /o/ meant ‘two of something’ and /an/ meant ‘three,’ or the reverse). Second, in Experiment 2, the pairing of words with pictures was determined randomly for each participant as an additional form of counterbalancing. Similarly, the assignment of sets as trained or generalization was randomized for each subject with the constraint that the three generalization sets always contained at least one one-syllable set and at least one two-syllable set (with the third set randomized for syllable structure).

### 3.1.2. Participants

Forty-six adults between 18 and 35 years of age participated. Two were removed from the analysis, one because of insufficient fixations and the second because his overall accuracy (0.44) was more than two SDs lower than the average accuracy in his alternation condition (0.87). The data from the remaining 44 participants (voicing-alternation:  $N = 22$ ; manner-alternation:  $N = 22$ ) were analyzed.

### 3.1.3. Auditory stimuli

Thirteen triplets were used (Appendix A.1): The six from Experiment 1, plus seven new triplets. Words were recorded and processed using the methods described in Experiment 1 by a male native monolingual speaker of standard American English.

### 3.1.4. Visual stimuli

Pictures from Experiment 1 were used, with the addition of seven sets (Appendix A.2) that went through the same development process. Additionally, pictures depicting three of each of the items (e.g., three dogs) were created. Pictures were randomly assigned to auditory stimuli for each participant.

### 3.1.5. Procedure

As in Experiment 1, Experiment 2 took place in two sessions scheduled one to two days apart, and the basic training/testing structure was the same as that of Experiment 1. On the first day, researchers obtained informed consent and then participants began training, during which no eye-movements were tracked.

Day 1 consisted of a total of 633 training trials in two phases. Phase one contained 171 trials presenting the 39 items (13 sets of three items) in their singular forms only. Each of the six high-frequency items (three trained and three generalization triplets) was presented six times (108 trials), and each of the seven low-frequency items (all trained

items) was presented three times (63 trials). As in Experiment 1, participants selected the referent from a four-referent display and received feedback.

In the second phase (462 trials), participants learned the trained items in their prefixed forms, but continued to see generalization items in unprefixed forms only. This portion contained two types of trials. Fifty-seven trials containing all singular items like those from Phase 1 were presented (one per low-frequency item, two per high-frequency trained and generalization items; these included all 13 item-sets). Additionally, 405 trials mixed both types of prefixed items with unprefixed trained items. In the latter, each of the 21 low-frequency words was a target three times in each prefixed or unprefixed form ( $7 \text{ sets} \times 3 \text{ words/set} \times 3 \text{ forms} \times 3 \text{ repetitions} = 189 \text{ trials}$ ), while each high-frequency trained word was a target eight times in prefixed or unprefixed form ( $3 \text{ sets} \times 3 \text{ words/set} \times 3 \text{ forms} \times 8 \text{ repetitions} = 216 \text{ trials}$ ). The generalization words (3 sets) occurred in the singular-only trials but never in the mixed trials, nor were they foils in any mixed trial.

Day 2 included a brief training session followed by a test session. The training session was 183 trials; of these, 39 trials were singular (each word presented once), 63 were low-frequency words presented in all three forms ( $7 \text{ sets} \times 3 \text{ words/set} \times 3 \text{ forms}$  [singular and both prefixes]), and 81 were high-frequency trained words presented in all three forms ( $3 \text{ sets} \times 3 \text{ words/set} \times 3 \text{ forms} \times 3 \text{ repetitions}$ ). Again, the generalization words were never presented in mixed trials as targets or foils.

The eye-tracker was calibrated immediately before the test session using the same procedure as Experiment 1. The test session was 480 trials, in which half the words were high-frequency trained items and the other half were generalization items. During testing, items were presented only in prefixed forms (with no singular pictures/ unprefixed forms). Items were always displayed in sets (one /t/-item, one /d/-item, one /z/-item, and one of those items repeated in the opposite number). Each item was presented with each of two prefixes 10 times for 360 experimental trials ( $6 \text{ item-sets} \times 3 \text{ items/set} \times 2 \text{ prefixes} \times 10 \text{ repetitions}$ ); the remaining 120 trials were fillers, in which the opposite-prefix item was the auditory stimulus (i.e., when the prefix of the auditory stimulus matched only one item on the screen).

### **3.2 Results: Experiment 2**

As above, we report an analysis of accuracy first, followed by fixations. For the latter, given the modified design (with no control group) we start with an analysis combining both groups to establish that training condition modulated competitor fixations. Having established learning, we then go on (where needed) to examine each alternation condition separately to assess the effect of prefix and generalization.

#### 3.2.1. Accuracy

Participants in Experiment 2 were less accurate at test than those in Experiment 1 (Table 5), though they achieved close to 90% average accuracy for the trained items. As in Experiment 1, there was a marked decrement in performance for the alternating item in the given language (i.e., /t/ in the voicing alternating condition, /z/ in the manner). See Appendix B.2 for more detailed results.

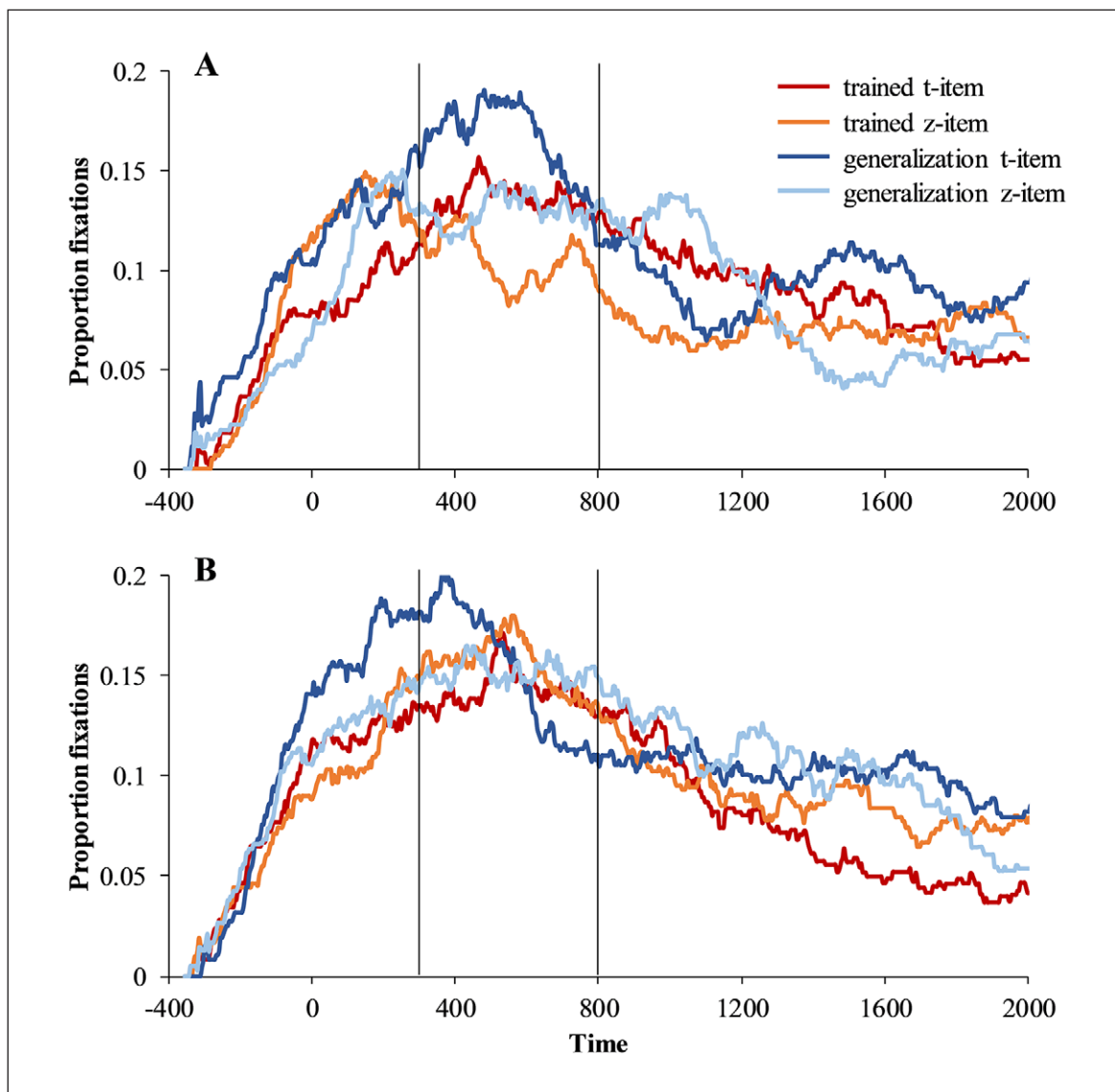
#### 3.2.2. Fixations: Overall effect of learning

The time course of fixations to /t/- and /z/-items in both trained and generalization trials is shown in Figure 5, with the voicing group in Panel A and the manner group in Panel B. Again, competitor fixations peak around 400–500 ms after the offset of the prefix, and we used an analysis window from 300 to 800 ms.



**Table 5:** Average accuracy (standard deviation) by condition for Experiment 2. Note as this was accuracy only on the test trials, it only reflects the high frequency training items.

		/an/					/o/				
		d	t	z	filler	total	d	t	z	filler	total
<b>Voicing</b>	<b>trained</b>	.909 (.084)	.764 (.200)	.961 (.044)	.950 (.074)	.896 (.069)	.908 (.097)	.903 (.112)	.947 (.066)	.946 (.046)	.926 (.062)
	<b>gen.</b>	.865 (.121)	.714 (.190)	.850 (.161)	.912 (.096)	.835 (.114)	.859 (.160)	.824 (.185)	.870 (.136)	.946 (.081)	.875 (.114)
<b>Manner</b>	<b>trained</b>	.915 (.110)	.939 (.070)	.768 (.178)	.922 (.114)	.886 (.094)	.944 (.086)	.952 (.056)	.917 (.113)	.925 (.102)	.934 (.080)
	<b>gen.</b>	.832 (.149)	.850 (.164)	.608 (.233)	.912 (.105)	.800 (.130)	.883 (.131)	.886 (.121)	.861 (.170)	.928 (.101)	.890 (.104)



**Figure 5:** Proportion of fixations to /t/- and /z/-items when target is an+/d/-item in Experiment 2. **Panel A:** Voicing group; **panel B:** Control group. Black lines indicate analysis time window.

Our analysis started by constructing a linear mixed effects model comparing fixations across both training groups. The goal here was to document an effect of learning on the balance of fixations to /t/- and /z/-initial items. For this we ran a model (in [3]) with empirically logit transformed fixations (from 300 to 800 ms) as the dependent variables. The fixed effects included languagegroup (voicing vs. manner, +0.5/−0.5), prefix ([an]- vs. [o]-, +0.5/−0.5), itemtype (/t/-initial vs. /z/-initial, +0.5/−0.5) and trainingtype (trained/generalization, +0.5/−0.5), and unrelated fixations were included as a covariate. As in our prior models, the random effects included only a random intercept by subject.

$$(3) \quad \text{fixations}_{\text{competitor}} \sim \text{itemtype} * \text{languagegroup} * \text{prefix} * \text{trainingtype} + \text{fixations}_{\text{unrelated}} + (1|\text{subject})$$

Here, the bias to fixate /t/- over /z/-items is indicated by a main effect of itemtype, and the degree to which this interacts with languagegroup would indicate an effect of learning.

Supporting this, the model found a marginally significant interaction of languagegroup and itemtype ( $B = .172$ ,  $SE = .094$ ,  $t(990.5) = 1.827$ ,  $p = .068$ ), and a marginally significant three-way interaction of languagegroup, itemtype, and trainingtype ( $B = .364$ ,  $SE = .189$ ,  $t(990.5) = 1.93$ ,  $p = .054$ ). This documents some modulation of the looking to /t/- vs. /z/-items by experience (and perhaps a further moderation by whether the item was trained or generalization). To understand these patterns better, we next conducted follow-up analyses separately for each language group.

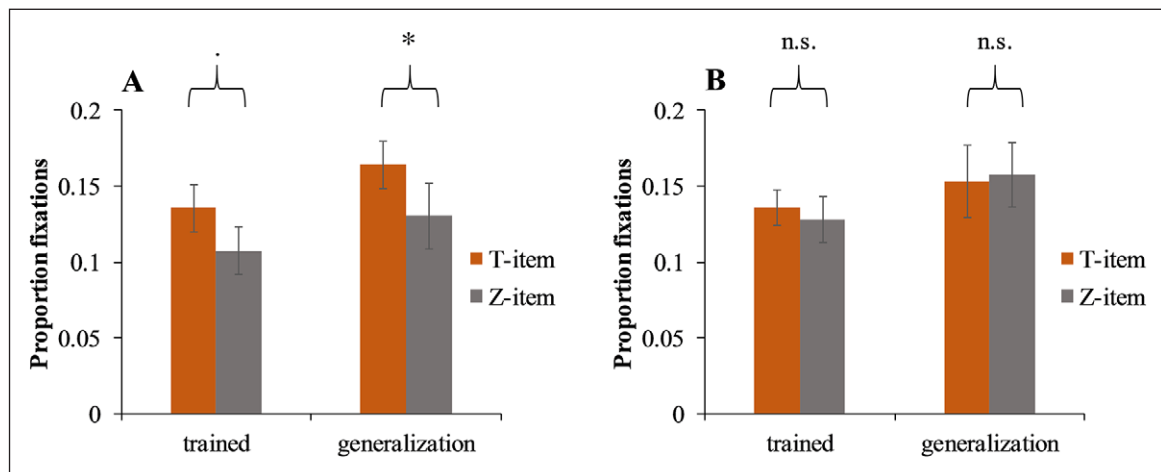
### 3.2.3. Fixations for voicing alternations

We next examined just the participants experiencing a voicing alternation to ask if 1) there was a decrement in the bias to fixate /t/ over /z/ (after hearing [d]) for generalization words; and 2) if the more robust training led to an effect of prefix type. This model was similar to the previous one, including item type (/t/-item vs. /z/-item), prefix (o- vs. an-) and training type (trained vs. generalization) as fixed effects, random intercepts on participant, and looks to the unrelated item as covariate.

Results of the mixed effects model are shown in **Table 6** and **Figure 6**. There were significant effects of itemtype ( $p = 0.0182$ ) and trainingtype ( $p = 0.0023$ ), but no interaction. That is, participants looked more to the /t/-item than the /z/-item in both trained and generalization trials, though they looked more (overall) in the generalization trials than in the trained trials. The lack of an itemtype by trainingtype interaction indicated robust generalization to new items. There was also a marginal

**Table 6:** Mixed effects model examining looks to competitor objects in Experiment 2, voicing condition.

Effect	B	SE	T	df	p
Item type	0.161	0.068	2.37	495.0	0.0182*
Training type	−0.208	0.068	3.07	495.1	0.0023*
Prefix	−0.072	0.068	1.06	495.6	0.29
Unrelated looks (covariate)	0.080	0.054	1.48	482.1	0.14
Item type × training type	0.030	0.136	0.22	495.0	0.83
Item type × prefix	0.249	0.136	1.83	495.0	0.0678.
Training type × prefix	−0.101	0.136	0.74	495.2	0.46
Item type × training type × prefix	−0.098	0.271	0.36	495.0	0.72



**Figure 6:** Fixations to /t/- and /z/-items by training type when the target is /d/-item in an alternating context (an-prefix) (**Panel A**) or a non-alternating context (o-prefix) (**Panel B**) in Experiment 2, voicing condition.

interaction of itemtype and prefix. Posthoc comparisons indicated that in the alternating context (the/an/-prefix), participants fixated the /t/-item more than the /z/-item,  $t(234.5) = 3.07$ ,  $p = 0.0024$ , but in the non-alternating context (the /o/-prefix), participants fixated both items equally,  $t(235.8) = 0.38$ ,  $p = 0.71$ . In both cases, the itemtype by trainingtype interaction was non-significant (alternating context:  $t(234.5) = 0.11$ ,  $p = 0.92$ ; non-alternating context:  $t(235.8) = 0.40$ ,  $p = 0.69$ ). This suggests that the effect of the item type was conditioned by the prefix but not by training type. To examine generalization, we split the alternating context by trained and generalization items and found that participants fixated the /t/-item marginally more than the /z/-item in the trained words,  $t(100.7) = 1.9$ ,  $p = 0.0573$ , and significantly more in the generalization words,  $t(106.6) = 2.3$ ,  $p = 0.0223$ . Thus, this bias is robust for the alternation-cuing prefix in both trained and generalization words.

#### 3.2.4. Fixations for manner alternation

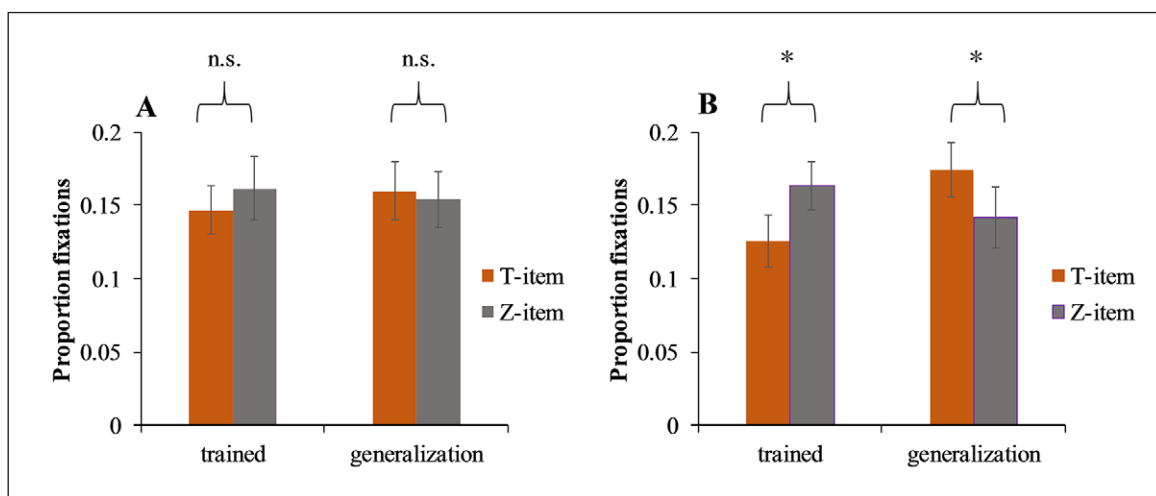
A similar mixed effects model was applied to the manner alternation condition. Results are shown in **Table 7**, with a plot in **Figure 7**. There were no significant effects of item type, training type, or prefix, though there was an interaction between itemtype and trainingtype ( $p = 0.0111$ ). At first glance, this result appears consistent with some form of learning effect. However, post-hoc tests confirm that the interaction is entirely driven by fixations in the /o/-prefix trials. In these trials, participants fixated the /z/-item more than the /t/-item in the trained words ( $t(104.9) = 2.02$ ,  $p = 0.0462$ ), but reverse for the generalization words ( $t(104.5) = 2.0$ ,  $p = 0.0482$ ) (see **Figure 7B**).

### 3.4. Discussion

This experiment again showed significant learning effects for the participants in the voicing alternation language. Listeners fixated the /t/-item more than the /z/-item, but only in the context of the alternation-conditioning prefix. This suggests that the larger number of words may have led to learning that yielded a somewhat more context-sensitive response to the phonological alternation. The lack of an interaction with trainingtype suggests that this held true for both the trained and the generalization items (see **Figure 6**). While participants fixated competitors more overall after hearing the generalization items (perhaps suggesting that they were less certain), the bias to fixate /t/- over /z/-items during the generalization trials looked much like fixations during trained trials. We interpret this to indicate that listeners both acquired the rule and generalized their behavior, activating

**Table 7:** Mixed effects model examining looks to competitor objects in Experiment 2, manner condition.

Effect	B	SE	T	df	P
Item type	-0.012	0.066	0.18	493.5	0.86
Training type	-0.064	0.066	0.97	494.0	0.33
Prefix	0.034	0.066	0.51	494.0	0.61
Unrelated looks (covariate)	-0.012	0.051	0.23	513.8	0.82
Item type × Training type	-0.334	0.131	2.55	493.5	0.0111*
Item type × prefix	-0.037	0.131	0.28	493.5	0.78
Training type × prefix	0.029	0.131	0.22	493.7	0.83
Item type × training type × prefix	0.316	0.262	1.20	493.5	0.23

**Figure 7:** Fixations to /t/-item and /z/-item by training type when target is /d/-item in an alternating context (an-prefix) (Panel A) or a non-alternating context (o-prefix) (Panel B) in Experiment 2, manner condition.

the /t/-item more than the /z/-item in the alternating context, even when they had never heard the particular lexical item in that context.

Perhaps more interestingly, unlike Experiment 1a, this pattern was not observed with the /o/- prefix (which did not cue an alternation). This suggests that with much richer data across which to generalize, listeners' use of the phonological regularity is conditioned on the licensing context. This finding supports a phonological inference model of recognition.

In contrast, the results of the manner alternation mostly paralleled those of Experiment 1b. After the /an-/ prefix, participants who learned a stopping rule were not more likely to fixate an underlying /z/-item than an underlying /t/-item (in either trained or generalization items). This contrasts with the voicing alternation group, who *were* more likely to fixate an item whose representation was a candidate for the voicing rule. The one exception to this was when the [o+d...] prefix was heard (a prefix which did not condition alternation). On those trials, participants did fixate the /z/-item more than the /t/-item in trained trials, but the opposite occurred in generalization trials. It is not clear what to make of this as this was a neutral context. It may represent simply noise in the data (a Type I error). Alternatively, it is possible that the manner alternation was harder to learn and this represents a sort of early (erroneous) hypothesis about the regularity, much as children often go through temporary phases of over- or undergeneralizing phonological and morphological regularities in speech production before arriving at the correct regularity.

## 4. General discussion and conclusion

These experiments examined real-time lexical competition when a recently learned phonological alternation created lexical ambiguity. We started with three research questions. First, do listeners activate stored forms whose underlying form differs from the auditory input, but whose surface forms match the signal as the result of alternation? Second, would listeners generalize this pattern to novel words whose alternated forms had not been encountered in training? And lastly, does the ability to learn a new phonological regularity depend on the content (e.g., the phonemes that are modified), and the learners' prior experience with those phonemes?

In each experiment, participants learned a phonological alternation triggered by a prefix, and we used the VWP to test whether they activated forms that matched due to the alternation compared to forms that did not undergo the same alternation. Our results show clear evidence that learned phonological regularities can influence real-time lexical competition dynamics, though not in every situation or always in the same way. This suggests that how listeners identify and phonologize the source of a phonological alternation is influenced by a variety of factors related to *learning*, including the size of the item-set and the featural content of the alternation itself.

We thus organize our discussion around our three research questions, as well as a new one: Why do learners appear to use phonological alternations differently depending on learning history? We start by focusing on the core issue of whether listeners can use alternations during real-time word recognition, and whether this can generalize to new words. We then turn to the issue of representation and learning to discuss the apparent limits illustrated by the manner alternation, and the unexpected difference in the use of the triggering context between Experiments 1 and 2. Together, these results suggest that whatever system integrates phonological knowledge into online word recognition must be flexible and is profoundly shaped by the learner's history with the regularities.

### 4.1. Real-time processing

#### 4.1.1. Listeners can use phonological alternations in real-time word recognition: Voicing

Experiment 1 showed that listeners who had learned a [t] ~ [d] voicing alternation were more likely to fixate a /t/ competitor than a /z/ competitor. In contrast, no such pattern was found in listeners exposed to a control language with no alternations. This suggests that upon hearing a surface [d], listeners activated novel words with surface forms that contained a [d] as the result of an alternation. Experiment 2 replicated this with a larger set of items and showed that listeners generalize even to items that were learned in their unprefixed form and never encountered in an alternating context. This suggests a clear answer to our primary question—phonological regularities *can* shape how strongly lexical competitors are considered, even those whose underlying form differs from the input.

At the same time, as we discuss shortly, this was not observed uniformly. Listeners exposed to a manner alternation ([z] ~ [d]) in Experiment 1b and in Experiment 2 did not show evidence that the alternation impacted real-time processing. In an artificial language learning paradigm, one has to consider whether this is a failure of processing (e.g., people knew the rule, but could not use it in real time) or learning (they simply did not acquire the regularity). We saw no evidence that participants had learned the rule (for example, it was not the case that the early fixations [our target of analysis] showed no effect of the regularity, but the later ones [perhaps indicating late or offline processing] did show such an effect), and as we discuss in Section 4.2, there are good reasons to consider this a failure to acquire the regularity (as opposed to a situation in which the regularity had been acquired but could not be deployed in real-time). Thus, the lack of an effect with



the manner alternation does not constitute a lack of evidence for the hypothesis that phonological regularities impact real-time processing.

#### 4.1.2. Generalization

In identifying models of real-time processing that can account for these effects, a critical factor is generalization: Are the effects observed due to representations of individual words, or are they due to the application of a rule? If the regularities are fully lexicalized, then most standard models of word recognition can account for these effects by simply ‘expanding’ the phonological template for a given word. In contrast, if effects generalize to words that have not been encountered in their alternating form, this may require additional mechanisms. In this regard, Experiment 2 showed clear evidence for generalization, at least in the voicing alternation condition, as learners showed real-time sensitivity in words for which they had never experienced this alternation.

#### 4.1.3. The role of conditioning context

Perhaps one of the most surprising effects was the way that the conditioning prefix affected performance in the voicing alternation languages. In Experiment 1, the bias to fixate /t/ over /z/ upon hearing a [d] form was observed even in the non-alternating context. This suggests that listeners adopted a relatively coarse, context-independent solution to the problem. In the extreme one could even argue that the effects of the regularity are solely due to learning—participants encode a more expanded lexical template (for alternating words) during training, and need not deploy the phonological rule during training at all. In contrast, in Experiment 2 when listeners were trained on a larger item-set, they exhibited the same bias, but *only when they heard the [d] in its alternating context* (and this appeared in both trained and generalization words). Thus, in this richer training context, listeners appear to adopt a much more context-sensitive solution to the ambiguity created by the alternation.

These two experiments make potentially conflicting claims about the nature of the representations that support phonological regularities, and we return to that in Section 4.2.2. Here, our goal is to examine the consequences of these findings for models of real-time processing. In this light, we would argue that the situation in Experiment 2 is more representative of real-world listeners than that of Experiment 1a—real listeners know far more words and are more likely to have had varying degrees of experience with them. If Experiment 2 is taken as more representative, it suggests that well-learned alternations are likely to be engaged in real-time processing, in a highly context-sensitive way.

#### 4.1.4. Models of real-time processing

As we described in the introduction, prior work on the intersection of phonology and word recognition has often focused on regularities like assimilation that leave subphonemic ‘traces’ in the signal (e.g., Gow, 2002); as a result, there is little evidence that a purely rule-based, categorical alternation in the signal can influence processing. Our study constitutes strong evidence that it can (although it does not always), and it does so with some level of abstraction. This has important implications for models of word recognition.

While standard models like TRACE (McClelland & Elman, 1986) and Shortlist-B (Norris & McQueen, 2008) do not stand theoretically opposed to effects of categorical regularities, they also do not have a clear mechanism by which to implement them. In these models, word recognition is conceived as a largely bottom-up mapping from phonemic representations to words. While TRACE can often handle such regularities via feedback from ‘gangs’ of neighbor words (McClelland, 1991), it is not clear if such effects can cross morpheme boundaries (as in our language). While exemplar models (Goldinger, 1998)

and underspecification accounts (Lahiri & Marslen-Wilson, 1991) may be able to acquire categorical regularities, neither has been examined as a model of real-time processing (we return to those models in the next sections when we talk about representation). Perhaps the most promising approach would be recurrent connectionist networks (e.g., Elman, 1990; Gaskell & Marslen-Wilson, 1997; Gupta & Tisdale, 2009) which use a distributed network to track sequential dependencies among phonemes and may be highly sensitive to such alternations (and see Gaskell, 2003, for such a model applied to assimilation).

#### 4.2. Learning and representation

We found robust evidence that listeners can use phonological regularities to bias activation when encountering phonemes that alternate. However, this was not observed uniformly across studies, most notably not for the manner alternation, but even within the voicing alternation it appeared to take different forms across Experiments 1 and 2. This too has implications for models of word recognition.

##### 4.2.1. Phonetically conditioned phonological rules: Manner alternation

Experiments 1b and 2 presented listeners with a different alternation, one of manner rather than voicing ([d] ~ [z]). This should have paralleled the voicing alternation with listeners fixating the /z/-item more than the /t/-item (particularly after [an-]). Instead, we found that listeners fixated the /z/- and /t/-items equally in both Experiments 1 and 2 (with the exception of the one anomalous result in Experiment 2). As the participant population and design were the same for the voicing and manner alternations, the difference must have come from the specific qualities of the two alternations.

One possibility is that during word recognition, listeners are more attuned to manner differences than voicing differences, as evidenced by mispronunciation tasks in which listeners are less likely to identify a voicing mispronunciation than a manner one (Cole et al., 1978; Martin & Peperkamp, 2015).<sup>3</sup> This may explain why the listeners who learned a voicing alternation were better able to activate the relevant competitor than those who learned a manner alternation—it was simply easier to ignore a mismatch in voicing. It is important to distinguish, however, between activation of mispronounced forms and activation of alternated forms; phonological alternations are regular and predictable, and listeners may use that regularity in a way that they cannot in a mispronunciation detection task. While Experiment 1 may be consistent with this sort of account, the fact that the relative activation of /d/ and /t/ was conditioned on the context in Experiment 2 does not support it.

A second possible explanation is that participants' English phonologies might have played a role. In North American English, a flapping alternation (*bi[t]e* ~ *bi[r]ing*) affects stop voicing, and other processes affect obstruent voicing, like the assimilation that occurs in the plural *cat[s]* vs. *dog[z]*. While fricative ~ stop alternations can occur in English (e.g., Reynolds [1994] and Schilling-Estes [1995] note forms like [wʌdɹt] for 'wasn't' and [kədən] for 'cousin' in Southern States English), they are less pervasive than the voicing alternation that arises from flapping and are not productive, especially in the dialects spoken by our Midwestern participants. It may be the case that listeners are able to scaffold the voicing alternation on a phonological generalization that exists, in a somewhat different context, in English. This may in part be because given their history with English, such alternations sound more phonetically natural (e.g., Pitt, 2009).

<sup>3</sup> Thanks to anonymous reviewers for suggesting this possible explanation and pointing out the relevant references.

Similarly, perhaps listeners place relatively little weight on features that they know can alternate in their language (e.g., as predicted by underspecification). Ernestus and Mak (2004) found that Dutch speakers were less attuned to voicing mispronunciations than manner or place in word-initial fricatives but not in word-initial stops, and they attributed this finding to the fact that Dutch word-initial fricatives alternate in voicing, while word-initial stops do not. Perhaps in our studies, the native English speakers were simply more likely to ignore the featural differences and activate /t/-items for [d]-targets because these speakers have down-weighted the voicing feature for English stops. However, if this is the explanation behind the asymmetries between the voicing group and the manner group, then it still suggests that knowledge of the phonological processes of the language do influence processing—in this case it would simply be that knowledge of the phonological processes of *English* influences processing of newly learned words, perhaps conspiring with knowledge of the regularities in the words to use the new (presumably less robustly represented) rule. This explanation may still support a phonological inference model of recognition; if listeners are heavily biased by the phonological alternations best known to them (that is, those of their native language), then they may have difficulty forming non-native inferences in the amount of training time provided here. This suggests that with additional (perhaps massive) exposure, listeners may in fact be able to learn the manner alternation rule.

A third possibility is that listeners are simply matching sounds in acoustic space and activating items whose forms are most similar to the auditory signal. As noted above, it is hard to quantify phonetic similarity, but intuition may suggest that [d] is more similar to [t] than to [z], leading participants to be more likely to accept a [d] ~ [t] alternation than a [d] ~ [z] one. Because the control participants presented in Experiment 1, who had learned no alternations at all, showed no preference for the /t/-competitor over the /z/-competitor, we can conclude that acoustic similarity alone did not drive fixation patterns. However, it may have played a role in how easily participants learned and abstracted the alternations, somehow making the t/d alternation easier to grasp than the z/d alternation.

#### 4.2.2. Contextually-conditioned effects

The fact that Experiments 1 and 2 resulted in differential use of the conditioning context (in the voicing alternation condition) also has crucial implications for models of how phonological regularities are learned and represented (independent of real-time processing). While both results support the hypothesis that phonological knowledge can shape real-time perception, they differ on whether or not this is context sensitive. Here, the models of word recognition introduced in Section 1.1 may help explain the disparities. The fact that participants in Experiment 1a fixated the /t/ word more than the /z/ word *even in non-alternating contexts* suggests that they extracted a relationship between [t] and [d], but that they did not restrict that relationship to the phonologically viable context. These results are most consistent with something resembling an underspecification account; listeners matched a surface form to its potential lexical matches, regardless of context. In this account, particular features (e.g., those that alternate) are unspecified in lexical representations for words; consequently, the system is tolerant of mismatch in the auditory signal. Here, underlying /t/ is underspecified for voicing and would therefore not mismatch a surface [d]. Consequently, words that differ only in underlying /t/ or /d/ could be similarly active after hearing a /d/, essentially tolerating this alternation. Critically, all this happens purely as a function of the lexical representation—it is independent of context, much like the results of Experiment 1a.

Exemplar accounts also may be able to account for our findings. In these accounts, listeners store all forms of a word and extract the closest match. In this case, listeners will have learned to associate both [tib] and [dib] with the same meaning; consequently upon hearing [dib], listeners might activate that meaning, even when it is not in an alternating context. A critical limit on this claim, however, is that to the extent that the exemplar representations contain neighboring context (e.g., the prefix), they may act context sensitively with a mechanism for compensation or regressive inference. This has not been spelled out explicitly for exemplar models; however, if the exemplar is narrowly defined, these models may be consistent with Experiment 1.

In contrast, the fact that effects in Experiment 2 were context dependent supports something like phonological inference, as listeners relied on phonological context to extract an underlying form that could match the surface because of the alternation. This suggests that with a large enough set of items, listeners adopted a different strategy of word recognition, forming an accurate phonological generalization in only the licensing context. Again, it is conceivable that exemplar models could also account for this effect (if the exemplars included the conditioning context).

It seems unlikely to us that listeners in the two experiments used completely different models of word recognition. So what was the source of their different behaviors? At an empirical level, it is likely that the *type* of training in Experiment 2 played a role. Certainly the fact that there were more words offers a plausible route for establishing a more general abstract rule—variability in irrelevant elements has often been shown to help learners achieve a more context-invariant representation (Apfelbaum et al., 2013; Gómez, 2002; Lively et al., 1993; Rost & McMurray, 2010). There may also be subtler training effects. In Experiment 2, listeners were trained on some forms at a fairly low frequency, which may mean that they had a hard time learning those forms. Perhaps when listeners are less able to learn specific words, they are more able to extract the one constant—the phonological generalization (Gómez, 2002).

However, this does not speak to what kinds of models listeners may deploy. If they had an underspecified phonological system (as supported by the results of Experiment 1), why would differences in the training regime lead them to abandon it in Experiment 2? Or if regressive inference were in place for other regularities, why was it not adapted to Experiment 1? The differences between the experiments seem to speak to the flexibility of learning. These findings argue that the learning system does not appear to be ‘pre-wired’ for any specific way of dealing with phonological alternations. Instead, it appears to adopt a mode of processing depending on particular items being learned (and likely depending on the tasks they will be used for). This ultimately makes it difficult to make conclusive claims that one model (e.g., underspecification or an exemplar model) accounts for listeners’ behavior. Rather, whatever model one adopts must be a developmental or learning model that can account for this flexibility.

In this regard, models like distributed connectionist models may offer the best approach. These models do not start with a predefined set of internal representations and processing steps, but rather acquire the representations that appear to best capture the task at hand (e.g., Elman & Zipser, 1988). This is clearly seen in recent models of reading (Armstrong et al., 2017; Kim et al., 2013) where even a handful of items are sufficient to warp internal representations linking letters and sounds. During training, these models adopt internal representations that are in some ways a hybrid of rule-based mappings (e.g., O makes the /a/ sound) and clusters of exceptions or item-based learning. These can be shifted around by the precise distribution of items in a way that may be optimized for reducing error, but which may not reflect any simple or rational approach.

Such models can also adopt systematically inferior representations early in training (e.g., over-regularizing the past-tense: McClelland & Patterson, 2002). It is possible that the participants in Experiment 1 (for example) or the manner alternation condition of Experiment 2 are in such a state, which may account for their rather unexpected failure to use context (Experiment 1) or the strange use of context with the non-alternating prefix (Experiment 2). These models can be strongly constrained by prior experience, such that experience with analogous regularities in the native language could make future learning of regularities in a new language easier to the extent that they already fit with the schema provided by the first language (McClelland, 2013). This would be consistent with listeners' failure to learn the manner alternation. Finally, such models are also likely to show the kinds of learning effects by which more items and more variability increases the ability to extract regularities.

Gaskell and Marlen-Wilson's (1997) Distributed Cohort model operates using a similar architecture. It has already been shown to account for regressive inference effects in assimilation (Gaskell, 2003), though it is unclear if it can account for the learning effects shown here, or the apparent underspecification-like effect of Experiment 1. It remains to be seen whether this model (or class of models) can acquire these modes of processing from the specific statistics of our training set, but this may be a good next step. However, whatever model one ultimately adopts must indeed be grounded in highly flexible learning mechanisms that can acquire not only the words and rules of the language, but can be retuned to support different ways of using this knowledge.

#### **4.3. Conclusion and future directions**

As is clear from our discussion of learning, the artificial language learning aspect of our paradigm raises a number of complications. Thus it is vital to replicate this experiment with native speakers of a real language. Alternatively, future experiments could incorporate rules that are equally native or non-native. Experiment 2 also indicated that learning a larger variety of words, even when some are encountered infrequently, may help listeners generalize a process to new forms during recognition. The limits of this learning are another key area of study. How much training is necessary to learn to generalize a rule? Are there other manipulations of the data (e.g., multiple talkers) that might improve learning? The development of each of these skills in childhood or in second language learning could reveal a great deal about the mechanisms of processing. Finally, it may be useful to explore other measures (both online measures and offline measures like goodness ratings) to potentially overcome longstanding concerns about strategic effects in the visual world paradigm (though see Pontillo et al., 2015).

These are crucial questions for future work. However, this study offers evidence that sub-lexical, rule-based phonological processing is incorporated into real-time word recognition. Moreover, such effects do not appear to be consistent with any one model of phonological representation, but rather derive from a highly flexible developmental and/or learning system.

#### **Additional Files**

The additional files for this article can be found as follows:

- **Appendix A.** Stimulus sets for both experiments. DOI: <https://doi.org/10.5334/labphon.41.s1>
- **Appendix B.** Complete Analyses of Accuracy. DOI: <https://doi.org/10.5334/labphon.41.s2>



## Acknowledgements

Thanks to Hannah Rigler and the members of the MACLab at the University of Iowa for help with data collection. Thanks also to John Alderete, Paul Tupper, the members of the Vancouver Phonology Group, and attendees at LSA 85 and Mid-Phon 17 for helpful discussion. This work was supported by NIH grants DC011669 to AFT and DC0008089 to BM, and by a research grant from the Social Sciences and Humanities Research Council of Canada (SSHRC).

## Competing Interests

The authors have no competing interests to declare.

## References

- Alloppenna, P., Magnuson, J. S., & Tanenhaus, M. K. 1998. Tracking the time course of spoken word recognition using eye-movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439. DOI: <https://doi.org/10.1006/jmla.1997.2558>
- Andruski, J. E., Blumstein, S. E., & Burton, M. W. 1994. The effect of subphonetic differences on lexical access. *Cognition*, 52, 163–187. DOI: [https://doi.org/10.1016/0010-0277\(94\)90042-6](https://doi.org/10.1016/0010-0277(94)90042-6)
- Apfelbaum, K., Hazeltine, R. E., & McMurray, B. 2013. Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology*, 49(7), 1348–1365. DOI: <https://doi.org/10.1037/a0029839>
- Archangeli, D. 1988. Aspects of underspecification theory. *Phonology*, 5, 183–207. DOI: <https://doi.org/10.1017/S0952675700002268>
- Archangeli, D., & Pulleyblank, D. 1989. Yoruba vowel harmony. *Linguistic Inquiry*, 20, 173–217.
- Armstrong, B. C., Dumay, N., Kim, W., & Pitt, M. A. 2017. Generalization from newly learned words reveals structural properties of the human reading system. *Journal of Experimental Psychology: General*, 146(2), 227–249. DOI: <https://doi.org/10.1037/xge0000257>
- Brouwer, S., Mitterer, H., & Huettig, F. 2012. Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, 27(4), 539–571. DOI: <https://doi.org/10.1080/01690965.2011.555268>
- Clayards, M., Niebuhr, O., & Gaskell, M. G. 2015. The time course of auditory and language-specific mechanisms in compensation for sibilant assimilation. *Attention, Perception, & Psychophysics*, 77, 311–328. DOI: <https://doi.org/10.3758/s13414-014-0750-z>
- Coenen, E., Zwitserlood, P., & Bölte, J. 2001. Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes*, 16, 535–564. DOI: <https://doi.org/10.1080/01690960143000155>
- Cole, R. A., Jakimik, J., & Cooper, W. E. 1978. Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America*, 64(1), 44–56. DOI: <https://doi.org/10.1121/1.381955>
- Dahan, D., & Gaskell, M. G. 2007. The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, 57, 483–501. DOI: <https://doi.org/10.1016/j.jml.2007.01.001>
- Darcy, I., Peperkamp, S., & Dupoux, E. 2007. Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In: Cole, J., & Hualde, J. I. (eds.), *Laboratory Phonology*, 9, 411–442. Berlin: Mouton de Gruyter.

- Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. 2009. Phonological knowledge in compensation for native and non-native assimilation. In: Kügler, F., Féry, C., & Van de Vijver, R. (eds.), *Variation and Gradience in Phonetics and Phonology*, 265–309. Berlin: Mouton DeGruyter.
- de Jong, K. 1998. Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics*, 26, 283–310. DOI: <https://doi.org/10.1006/jpho.1998.0077>
- Dinnsen, D. A. 1996. Context-sensitive underspecification and the acquisition of phonemic contrasts. *Journal of Child Language*, 23, 57–79. DOI: <https://doi.org/10.1017/S0305000900010096>
- Elman, J. L. 1990. Finding structure in time. *Cognitive Science*, 14, 179–211. DOI: [https://doi.org/10.1207/s15516709cog1402\\_1](https://doi.org/10.1207/s15516709cog1402_1)
- Elman, J. L., & Zipser, D. 1988. Learning the hidden structure of speech. *The Journal of the Acoustical Society of America*, 83(4), 1615–1626. DOI: <https://doi.org/10.1121/1.395916>
- Ernestus, M. 2011. Gradience and categorality in phonological theory. In: van Oostendorp, M., Ewen, C., Hume, E., & Rice, K. (eds.), *Blackwell Companion to Phonology*, 2115–2136. Oxford: Blackwell.
- Ernestus, M., & Mak, W. M. 2004. Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain and Language*, 90, 378–392. DOI: [https://doi.org/10.1016/S0093-934X\(03\)00449-8](https://doi.org/10.1016/S0093-934X(03)00449-8)
- Eulitz, C., & Lahiri, A. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, 16, 577–583. DOI: <https://doi.org/10.1162/089892904323057308>
- Finley, S., & Badecker, W. 2009. Artificial language learning and feature-based generalization. *Journal of Memory and Language*, 61(3), 423–437. DOI: <https://doi.org/10.1016/j.jml.2009.05.002>
- Frisch, S., Large, N. R., Zawaydeh, B., & Pisoni, D. B. 2001. Emergent phonotactic generalizations in English and Arabic. In: Bybee, J. L., & Hopper, P. (eds.), *Frequency and the Emergence of Linguistic Structure*, 159–180. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/tsl.45.09fri>
- Gaskell, M. G. 2003. Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31(3), 447–463. DOI: [https://doi.org/10.1016/S0095-4470\(03\)00012-3](https://doi.org/10.1016/S0095-4470(03)00012-3)
- Gaskell, M. G., & Marslen-Wilson, W. D. 1996. Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144–158. DOI: <https://doi.org/10.1037/0096-1523.22.1.144>
- Gaskell, M. G., & Marslen-Wilson, W. D. 1997. Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12(5/6), 613–656. DOI: <https://doi.org/10.1080/016909697386646>
- Gaskell, M. G., & Marslen-Wilson, W. D. 1998. Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 380–396. DOI: <https://doi.org/10.1037/0096-1523.24.2.380>
- Gaskell, M. G., & Marslen-Wilson, W. D. 2001. Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, 44, 325–349. DOI: <https://doi.org/10.1006/jmla.2000.2741>
- Goldinger, S. D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183. DOI: <https://doi.org/10.1037/0278-7393.22.5.1166>
- Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279. DOI: <https://doi.org/10.1037/0033-295X.105.2.251>

- Goldinger, S. D., Pisoni, D. B., & Logan, J. S. 1991. On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, 152–162. DOI: <https://doi.org/10.1037/0278-7393.17.1.152>
- Gómez, R. 2002. Variability and detection of invariant structure. *Psychological Science*, 13, 431–436. DOI: <https://doi.org/10.1111/1467-9280.00476>
- Gow, D. W. 2001. Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45, 133–159. DOI: <https://doi.org/10.1006/jmla.2000.2764>
- Gow, D. W. 2002. Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 163–179. DOI: <https://doi.org/10.1037/0096-1523.28.1.163>
- Gow, D. W. 2003. Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65, 575–590. DOI: <https://doi.org/10.3758/BF03194584>
- Gow, D. W., & Im, A. M. 2004. A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language*, 51, 279–296. DOI: <https://doi.org/10.1016/j.jml.2004.05.004>
- Gupta, P., & Tisdale, J. 2009. Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language*, 61(4), 481–502. DOI: <https://doi.org/10.1016/j.jml.2009.08.001>
- Hercus, L. A. 1986. *Victorian Languages: A late survey*. Pacific Linguistic, B-77 (Vol. The Australian National University): Canberra.
- Hume, E., & Johnson, K. 2001. A model of the interplay of speech perception and phonology. In: Hume, E., & Johnson, K. (eds.), *The Role of Speech Perception in Phonology*, 3–26. New York: Academic Press.
- Itô, J., & Mester, A. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry*, 17, 49–73.
- Johnson, K. 1997. Speech perception without talker normalization. In: Johnson, K., & Mullenix, J. W. (eds.), *Talker Variability in Speech Processing*, 145–165. San Diego: Academic Press.
- Kim, W., Pitt, M. A., & Myung, J. 2013. How do PDP models learn quasiregularity? *Psychological Review*, 120(4), 903–916. DOI: <https://doi.org/10.1037/a0034195>
- Klatt, D. 1979. Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7, 279–312.
- Lahiri, A., & Marslen-Wilson, W. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245–294. DOI: [https://doi.org/10.1016/0010-0277\(91\)90008-R](https://doi.org/10.1016/0010-0277(91)90008-R)
- Lee, S., & Pater, J. 2008. *Phonological inference in word recognition: Evidence from Korean obstruent nasalization*. Korea University and University of Massachusetts, Amherst.
- Lively, S. E., Logan, J. S., & Pisoni, D. B. 1993. Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, 94, 1242–1255. DOI: <https://doi.org/10.1121/1.408177>
- Luce, P. A., & Pisoni, D. B. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1), 1–36. DOI: <https://doi.org/10.1097/00003446-199802000-00001>
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. 2003. The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132(2), 202–227. DOI: <https://doi.org/10.1037/0096-3445.132.2.202>

- Marslen-Wilson, W. 1987. Functional parallelism in spoken word recognition. *Cognition*, 25, 71–102. DOI: [https://doi.org/10.1016/0010-0277\(87\)90005-9](https://doi.org/10.1016/0010-0277(87)90005-9)
- Marslen-Wilson, W., & Zwitserlood, P. 1989. Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576–585. DOI: <https://doi.org/10.1037/0096-1523.15.3.576>
- Martin, A., & Peperkamp, S. 2015. Asymmetries in the exploitation of phonetic features for word recognition. *Journal of the Acoustical Society of America*, 137(4), EL307–EL313. DOI: <https://doi.org/10.1121/1.4916792>
- Matuschek, H., Kliegl, R., Vasisht, S., Baayen, H., & Bates, D. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. DOI: <https://doi.org/10.1016/j.jml.2017.01.001>
- McClelland, J. L. 1991. Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, 23(1), 1–44. DOI: [https://doi.org/10.1016/0010-0285\(91\)90002-6](https://doi.org/10.1016/0010-0285(91)90002-6)
- McClelland, J. L. 2013. Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, 142(4), 1190–1210. DOI: <https://doi.org/10.1037/a0033812>
- McClelland, J. L., & Elman, J. 1986. The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86. DOI: [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- McClelland, J. L., & Patterson, K. 2002. Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, 6(11), 465–472. DOI: [https://doi.org/10.1016/S1364-6613\(02\)01993-9](https://doi.org/10.1016/S1364-6613(02)01993-9)
- McMurray, B., Munson, C., & Tomblin, J. B. 2014. Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye-movements. *Journal of Speech, Language and Hearing Research*, 57, 1344–1362. DOI: [https://doi.org/10.1044/2014\\_JSLHR-L-13-0196](https://doi.org/10.1044/2014_JSLHR-L-13-0196)
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. 2002. Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), B33–B42. DOI: [https://doi.org/10.1016/S0010-0277\(02\)00157-9](https://doi.org/10.1016/S0010-0277(02)00157-9)
- Miller, G. A., & Nicely, P. E. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338–352. DOI: <https://doi.org/10.1121/1.1907526>
- Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. 2006. The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science*, 30, 451–479. DOI: [https://doi.org/10.1207/s15516709cog0000\\_57](https://doi.org/10.1207/s15516709cog0000_57)
- Mitterer, H., Kim, S., & Cho, T. 2013. Compensation for complete assimilation in speech perception: The case of Korean labial-to-velar assimilation. *Journal of Memory and Language*, 69, 59–83. DOI: <https://doi.org/10.1016/j.jml.2013.02.001>
- Newton, B. 1972. *The Generative Interpretation of Dialect. A Study of Modern Greek Phonology*. Cambridge: Cambridge University Press.
- Norris, D., & McQueen, J. M. 2008. Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. DOI: <https://doi.org/10.1037/0033-295X.115.2.357>
- Nygaard, L., Sommers, M., & Pisoni, D. B. 1994. Speech perception as a talker contingent process. *Psychological Science*, 5, 42–46. DOI: <https://doi.org/10.1111/j.1467-9280.1994.tb00612.x>
- Padgett, J. 1994. Stricture and nasal place assimilation. *Natural Language & Linguistic Theory*, 12(3), 465–513. DOI: <https://doi.org/10.1007/BF01118137>



- Pallier, C., Colomé, A., & Sebastián-Gallés, N. 2001. The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 12, 445–449. DOI: <https://doi.org/10.1111/1467-9280.00383>
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 309–328. DOI: <https://doi.org/10.1037/0278-7393.19.2.309>
- Pater, J. 2004. Austronesian nasal substitution and other NC effects. In: McCarthy, J. (ed.), *Optimality Theory in Phonology: A Reader*, 271–289. Oxford and Malden, MA: Blackwell. DOI: <https://doi.org/10.1002/9780470756171.ch14>
- Peng, L. 2003. Local conjunction and Kikuyu consonant mutation. *University of Pennsylvania Working Papers in Linguistics*, 9, 199–211.
- Peng, L. 2008. Kikuyu consonant mutation. In: Uyechi, L., & Wee, L.-H. (eds.), *Reality Exploration and Discovery: Pattern Interaction in Language and Life*. Stanford, CA: CSLI Publications.
- Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J. L., & Hopper, P. (eds.), *Frequency Effects and the Emergence of Linguistic Structure*, 137–157. Amsterdam: John Benjamins.
- Pierrehumbert, J. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46, 115–154. DOI: <https://doi.org/10.1177/00238309030460020501>
- Pitt, M. A. 2009. How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language*, 61, 19–36. DOI: <https://doi.org/10.1016/j.jml.2009.02.005>
- Pontillo, D., Salverda, A. P., & Tanenhaus, M. K. 2015. Flexible use of phonological and visual memory in language-mediated visual search. *Proceedings of the Cognitive Science Society*, 1895–1900.
- Pulleyblank, D. 1986. Underspecification and low vowel harmony in Okpe. *Studies in African Linguistics*, 17, 119–153.
- Pulleyblank, D. 1988a. Underspecification, the feature hierarchy and Tiv vowels. *Phonology*, 5, 299–326. DOI: <https://doi.org/10.1017/S0952675700002323>
- Pulleyblank, D. 1988b. Vocalic underspecification in Yoruba. *Linguistic Inquiry*, 19, 233–270.
- Ranbom, L. J., & Connine, C. M. 2007. Lexical representation of phonological variation. *Journal of Memory and Language*, 57, 273–298. DOI: <https://doi.org/10.1016/j.jml.2007.04.001>
- Reynolds, B. 1994. The /z/ à [d] variable rule in negative auxiliary contractions in Southern States English. *American Speech*, 69, 361–372. DOI: <https://doi.org/10.2307/455855>
- Ringen, C. O. 1988. Transparency in Hungarian vowel harmony. *Phonology*, 5, 327–342. DOI: <https://doi.org/10.1017/S0952675700002335>
- Rost, G. C., & McMurray, B. 2010. Finding the signal by adding noise: The role of non-contrastive phonetic variability in early word learning. *Infancy*, 15(6), 608–635. DOI: <https://doi.org/10.1111/j.1532-7078.2010.00033.x>
- Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. 2014. Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, 71(1), 145–163. DOI: <https://doi.org/10.1016/j.jml.2013.11.002>
- Scharinger, M., Monahan, P. J., & Idsardi, W. J. 2012. Asymmetries in the processing of vowel height. *Journal of Speech, Language and Hearing Research*, 55, 903–918. DOI: [https://doi.org/10.1044/1092-4388\(2011/11-0065\)](https://doi.org/10.1044/1092-4388(2011/11-0065))



- Schilling-Estes, N. 1995. Extending our understanding of the /z/ à [d] rule. *American Speech*, 70, 291–302. DOI: <https://doi.org/10.2307/455901>
- Snoeren, N. D., Gaskell, M. G., & Di Betta, A. M. 2009. The perception of assimilation in newly learned novel words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 542–549. DOI: <https://doi.org/10.1037/a0014509>
- Spivey, M. J., Grosjean, M., & Knoblich, G. 2005. Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10393–10398. DOI: <https://doi.org/10.1073/pnas.0503903102>
- Stemberger, J. P. 1991. Radical underspecification in language production. *Phonology*, 8, 73–112. DOI: <https://doi.org/10.1017/S0952675700001287>
- Stemberger, J. P., & Stoel-Gammon, C. 1991. The underspecification of coronals: Evidence from language acquisition and performance errors. *Phonetics & Phonology*, 2, 181–199. DOI: <https://doi.org/10.1016/B978-0-12-544966-3.50015-4>
- Viviani, P. 1990. Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In E. Kowler (ed.), *Eye Movements and Their Role in Visual Cognitive Processes. Reviews of Oculomotor Research V4*, 353–383. Amsterdam: Elsevier.
- Wang, M. D., & Bilger, R. C. 1973. Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America*, 54(5), 1248–1266. DOI: <https://doi.org/10.1121/1.1914417>
- Wheeldon, L., & Waksler, R. 2004. Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language*, 90, 401–412. DOI: [https://doi.org/10.1016/S0093-934X\(03\)00451-6](https://doi.org/10.1016/S0093-934X(03)00451-6)

**How to cite this article:** Farris-Trimble, A. and McMurray, B. 2018 Morpho-phonological regularities influence the dynamics of real-time word recognition: Evidence from artificial language learning. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 9(1): 2, pp. 1–34, DOI: <https://doi.org/10.5334/labphon.41>

**Submitted:** 06 August 2016    **Accepted:** 21 September 2017    **Published:** 23 January 2018

**Copyright:** © 2018 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/>.



*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 