RESEARCH REPORT

The effect of allophonic processes on word recognition: Eye-tracking evidence from Canadian raising

Ashley Farris-Trimble
Simon Fraser University

Anne-Michelle Tessier
Simon Fraser University and *The University of British Columbia*

Whether lexical representations are stored as abstract forms or exemplar tokens is the focus of much debate in both the phonological and word-recognition literature. This research report examines the recognition of words that have undergone Canadian raising and/or intervocalic flapping. Two eye-tracking experiments suggest that listeners are slower to fixate words that have undergone one or more phonological processes within their own raising dialect, supporting the idea that they must calculate a mapping from surface word forms to more abstract representations. Implications for representational and phonological theories are discussed.*

Keywords: phonology, laboratory phonology, psycholinguistics, word recognition, representations, Canadian raising

1. Introduction.

The interaction between two mostly predictable segmental processes in Canadian English—Canadian raising, which causes some diphthong nuclei to be raised, and intervocalic flapping, which reduces some /t/s and /d/s to [ɾ]—has long been of interest to phonologists, in part because its analysis highlights a core question: How are words that are subject to phonological processes stored in the mind? One way to frame this question is to ask whether the 'distance' between a surface form of a word and its underlying representation, according to a particular phonological analysis, is reflected somewhere in lexical processing. This research report tries to address this version of the question, using new eye-tracking data, with results that support some degree of abstract representations in the minds of listeners. We discuss their implications for both phonological grammar and word recognition.

1.1. The phonological analysis of two interacting processes.

In Canadian raising dialects, the diphthongs typically produced as [aɪ] and [aʊ] surface as the raised variants [ʌɪ] and [ʌʊ] preceding voiceless obstruents (Joos 1942) within the same syllable (e.g. Paradis 1980, Chambers 1989, Pater 2014) or foot (Chambers 1973, Kiparsky 1979). This phenomenon occurs across Canada and also in several portions of the United States, including upstate New York (Vance 1987), Michigan (Dailey-O'Cain 1997), Philadelphia (Fruehwald 2016), northern Indiana (Berkson, Davis, & Strickler 2017), and Chicago (Hualde, Luchkina, & Eager 2017). This Canadian raising process interacts with the general North American process of intervocalic flapping, which neutralizes the contrast between /t/ and /d/ to a surface flap [ɾ] in predictable prosodic environments,[1] most notably after a stressed vowel (e.g. Kahn 1980), as in 1. The famous interaction of these two processes occurs when both of their structural descriptions are

* Many thanks to Danica Reid, Marleis Bowering, Ethan Pincott, and Erica Halabourda, along with the rest of the Phonological Processing Lab at Simon Fraser University, for help with designing and running these experiments. Thanks to audiences at LSA and NoWPhon 2015 and the University of Alberta, as well as referees, for helpful suggestions and comments. This research was funded by an SSHRC Small Grant from Simon Fraser University to the first author.

[1] We note from the onset that flaps derived from /t/ and /d/ are not identical—those derived from /d/ tend to be preceded by slightly longer vowels, though listeners are not able to use that vowel length to distinguish the source of the flap (Herd, Jongman, & Soreno 2010). Since listeners tend to perform at chance when discriminating /t/-flaps and /d/-flaps, we treat them as essentially identical.

met in the same word, such as *writer* (1b), whose surface form contains both the raised diphthong and the flap.[2]

(1) The interaction between Canadian raising and intervocalic flapping

	GLOSS	SURFACE FORM	PROCESSES APPLIED
a.	write	[ɹʌɪt]	raising
b.	writer	[ɹʌɪɾɚ]	raising and flapping
c.	ride	[ɹaɪd]	—
d.	rider	[ɹaɪɾɚ]	flapping

The debate about the content of phonological representations that involve Canadian raising and intervocalic flapping centers on examples like 1b and 1d. A comparison of *rider* and *writer* shows that the apparent contrast between the raised and unraised diphthong emerges only in a single, highly specific environment: preceding a flap. Though the presence of a contrast usually signals phonemic status, the two vowels in this case are in fact predictable from other morphophonological properties of the words—that is, the only way a word can contain the string [ʌɪɾ] (like *writer*) is if it is morphologically related to a word in which that flap surfaces as a [t] (like *write*). How then should this limited contrast be represented in underlying representations?

Two primary camps have emerged. One argues for abstract, noncontrastive underlying representations paired with an opaque ordering of processes (Kiparsky 1971), as is shown in 2a (Harris 1951, Halle 1962, Chomsky 1964, Chomsky & Halle 1968, Chambers 1973). Opacity is, however, problematic for strictly parallel grammars like classic OPTIMALITY THEORY (Prince & Smolensky 2004 [1993]). The alternative account posits underlying contrasts—illustrated in its most extreme form in 2b—either in all raising contexts (Hall 2005) or in restricted or marginal contexts (e.g. Joos 1942, Vance 1987, Mielke, Armstrong, & Hume 2003, Pater 2014, Hualde et al. 2017). This alternative is less parsimonious, in that it may rely on positing contrasts in entirely predictable environments, and it can fail to account for the existence of productive alternations (Idsardi 2006).

(2) a. A rule-ordering account of Canadian raising and intervocalic flapping (adapted from Chambers 1973)

	'write'	'writer'	'ride'	'rider'
underlying	/ɹaɪt/	/ɹaɪtɚ/	/ɹaɪd/	/ɹaɪdɚ/
Canadian raising	ɹʌɪt	ɹʌɪtɚ	—	—
intervocalic flapping	—	ɹʌɪɾɚ	—	ɹaɪɾɚ
surface form	[ɹʌɪt]	[ɹʌɪɾɚ]	[ɹaɪd]	[ɹaɪɾɚ]

b. A full contrast-based account of Canadian raising and intervocalic flapping

	'write'	'writer'	'ride'	'rider'
underlying	/ɹʌɪt/	/ɹʌɪtɚ/	/ɹaɪd/	/ɹaɪdɚ/
surface form	[ɹʌɪt]	[ɹʌɪɾɚ]	[ɹaɪd]	[ɹaɪɾɚ]

As noted above, one difference between 2a and 2b is the distance between their underlying (UR) and surface (SR) representations. In 2b they are all equally close—by being identical—while in 2a they show different degrees of discrepancy, so that the UR and SR in *ride* are identical, whereas the UR differs from the SR in two different ways in *writer*.

[2] Though Canadian raising applies to both the [aɪ] and [aʊ] diphthongs, in this paper we focus solely on [aɪ], as it occurs in a broader set of words.

In this report, we investigate these questions of phonological contrast and representation by examining the interaction of Canadian raising and intervocalic flapping in word recognition. Theories of word recognition must account for how the listener identifies an underlying from a (potentially nonidentical) surface form; by focusing those theories of word recognition on the problem at hand, we hope to gain insight into the content of the representations themselves.

Before going further, we wish to address a potential concern with our overall approach. The mappings in 2 are traditionally understood in the generative phonological tradition to represent production. The derivations in 2a illustrate how raising and flapping processes turn underlying forms in the minds of Canadian English speakers into phonological surface forms, which are then sent off to be produced phonetically. However, this study's data comes not from word production but from word recognition, in which a listener perceives a surface form and must map it onto some lexical representation. Nevertheless, we see the representational DISTANCE between underlying and surface forms emphasized above as relevant to both production and recognition—and it is that relative distance, across different word-types, that our study is designed to identify.

2. WORD RECOGNITION AS A MIRROR FOR PHONOLOGY. The task of word recognition seems daunting on the surface; in real time, listeners must segment words from the speech stream, identify them, and extract meaning from the appropriate lexical entry. There is consensus that lexical activation begins immediately upon receipt of an acoustic signal; it proceeds incrementally and in parallel, with multiple potential matches being considered at the same time and the candidate set updated as the word unfolds; and it involves competition, such that words that are better competitors receive stronger activation and weaker competitors may be suppressed (McClelland & Elman 1986, Gaskell & Marslen-Wilson 1997, Allopenna, Magnuson, & Tanenhaus 1998, Dahan et al. 2001). This task is complicated by the fact that the acoustic signal contains forms that have been affected or altered by the language's phonological or phonetic processes. While in some cases this may aid word recognition (for example, in English, the presence of a nasal vowel helps the listener predict that the upcoming consonant is nasal; Lahiri & Marslen-Wilson 1991), it may also require the listener to uncover a representation that does not match the surface form—either because that underlying form is in some way abstract (see below) or through its connection to other related forms. Listeners can use a variety of strategies to determine the correct representation for a phonologically altered form, including analogy (Ernestus & Baayen 2003), phonological context and lexicality (Mitterer & Ernestus 2006), and lexical frequency (Connine, Ranbom, & Patterson 2008, Pitt, Dilley, & Tat 2011).

In parallel, a growing body of psycholinguistic research on word recognition has resulted in a number of competing word-recognition models. These models can be roughly divided into two categories: those that assume abstract phonological representations in the lexicon (e.g. TRACE (McClelland & Elman 1986); the FEATURALLY UNDERSPECIFIED LEXICON MODEL (Lahiri & Marslen-Wilson 1991); SHORTLIST (Norris 1994); the NEIGHBORHOOD ACTIVATION MODEL (Luce & Pisoni 1998)) and those that assume episodic or exemplar representations (e.g. Johnson 1997; MINERVA 2 (Hintzman 1984, 1986, 1988), and applied to word recognition by Goldinger 1996, 1998; the complementary systems approach (Goldinger 2007)). Viewed broadly, these two approaches are somewhat analogous to the two phonological explanations of raising and flapping introduced earlier: abstract representations acted on by phonological processes vs. contrasts stored in the representations themselves. This paper focuses on two word-recognition mechanisms that we predict could play a role in processing raised/flapped words: PHONOLOGICAL INFERENCE and EXEMPLAR MATCHING.

Phonological inference (Gaskell & Marslen-Wilson 1996, 1998) is a mechanism by which listeners can use phonological context to 'undo' predictable phonological patterns in the speech signal, thus arriving at abstract lexical representations. For instance, English has an optional process of regressive place assimilation, such that a word-final coronal can assimilate in place to a following consonant, resulting in a production like *leam bacon* for 'lean bacon'. In a cross-modal priming experiment, Gaskell and Marslen-Wilson (1996) found that *leam* could prime orthographic 'lean' when it was in a phonologically viable context (that is, preceding a labial-initial word like 'bacon'), but not if it was in a nonviable context (e.g. preceding a velar-initial word like 'gammon'). This suggests that listeners use phonological context to determine if a sound is licensed, and then work backward to activate that word's abstract representation. This type of account also explains findings that the degree of reduction in a word is correlated with how likely it is to be recognized; words whose phonetic form is more distant from the abstract representation are less accurately recognized (Ernestus, Baayen, & Schreuder 2002). Phonological inference is the word-recognition correlate of a rule-ordering phonological account of Canadian raising. Crucially, the assumption is that phonological inference is a process that occurs during word recognition, taking up cognitive resources and thus slowing the overall process of lexical activation as evidenced in behavioral tasks like slower lexical decision (e.g. LoCasto & Connine 2002).

Exemplar matching is an alternative mechanism that accounts for word recognition in episodic or exemplar systems.[3] In these approaches, lexical representations are stored as banks of episodic traces of acoustic forms; phonologically, this means that both [lin] and [lim] can be stored as possible representations for *lean*. The listener needs simply to match the form in the acoustic signal to a representation directly, with no need for recourse to intermediate levels. Connine and colleagues (e.g. Connine 2004, Connine et al. 2008) have argued that phonological variants due to optional phonological processes, like intervocalic flapping or unstressed schwa deletion, are likely stored in the lexical representation, at least in the cases where those variants are highly frequent. For example, in the case of optional intervocalic flapping, listeners were more likely to activate a lexical item when presented with the more frequent flap variant (*pretty* as [pɹɪɾi]) than the less frequent [t] variant ([pɹɪti]), suggesting that the flap variant was stored lexically and accessed directly, rather than linked to the /t/ variant (Connine 2004). Exemplar matching could be used alongside phonological theories of Canadian raising that assume underlying contrast. If each word is stored with its surface diphthong, listeners should be equally fast to recognize either vowel when produced in its expected phonological contexts, all else being equal (see below).[4]

2.1. PREDICTIONS OF DIFFERENT STORAGE ACCOUNTS FOR WORD RECOGNITION. The current study builds on a long history of using word recognition as a lens to phonological representations and vice versa (e.g. Lahiri & Marslen-Wilson 1991, Kolinsky, Morais, & Cluytens 1995, Marslen-Wilson, Nix, & Gaskell 1995, Gaskell & Marslen-

[3] Other mechanisms similar to exemplar matching are direct mapping or similarity mapping (Marslen-Wilson & Warren 1994, Connine et al. 1997, Gaskell & Marslen-Wilson 1997). In each of these cases, the listener matches an acoustic signal directly to a stored representation, without intermediate representations or processing.

[4] There is also support for a hybrid model of inference and exemplar matching (see e.g. Pinker & Prince 1988, Zuraw 2000, 2010, Pinnow & Connine 2014), which would better support transfer and generalization.

Wilson 1998, Gaskell, Spinelli, & Meunier 2002, Luce, McLennan, & Charles-Luce 2003, Mitterer & Blomert 2003, Spinelli, McQueen, & Cutler 2003, Connine 2004, Eulitz & Lahiri 2004, Sumner & Samuel 2005, Mitterer & Ernestus 2006, Connine et al. 2008, Pitt 2009, Pinnow & Connine 2014, Farris-Trimble & McMurray 2018). Most of these previous studies, however, focused on optional phonological processes, many of which occur across word boundaries, like the nasal place assimilation in *leam bacon*. Likewise, previous studies examined phonological processes that occur in isolation, rather than multiple phonological processes that can interact opaquely. Our study also differs in its methodology: we use eye-tracking in the visual-world paradigm (Allopenna et al. 1998) to measure lexical activation during on-line, real-time word recognition. Fixations gleaned from visual-world studies have been shown to reflect lexical activation levels in a model like TRACE (McClelland & Elman 1986, Allopenna et al. 1998). By monitoring participants' fixations to pictures on the screen, we can estimate activation levels of the target as it unfolds. Because eye movements are not easily controlled by the participant and are responsive to an auditory stimulus, they are both more natural and more revealing than some other types of behavioral data like gating.

Phonological inference and exemplar matching make different predictions about how the listener recognizes words in which Canadian raising and/or intervocalic flapping have occurred. Under the phonological-inference account, the listener must check the context and undo the phonological process in real time during word recognition, working backward from [ʌɪ] to /aɪ/ and from [ɾ] to /t/ or /d/. This should result in timing differences because of the additional processing load required for the rule undoing; we spell these out in further detail once we have introduced our set of stimulus conditions.

Under an exemplar-matching account, by contrast, the listener can access surface-like representations directly, meaning they should match [ɹʌɪɾɚ] to 'writer' or [ɹaɪɾɚ] to 'rider' with no other steps. In this case, we would not expect any differences in the timing of activation between the two words, at least without further elaboration of the theory. We note at the outset, however, that this theory might be elaborated in a few ways. First, most exemplar approaches expect processing to be mediated by frequency, such that whichever words or paradigms are more frequent in the listener's experience would be recognized more quickly, independent of their diphthongs or flaps (Bybee 2001, 2002a). An exemplar model might also take into account the number of word-form variants within a paradigm, under the assumption that accessing and/or choosing between larger numbers of exemplars will slow recognition down. One instantiation of this view would be that listeners who are familiar with both raised and unraised dialects will have larger exemplar clouds for words like *writing*, with and without the raised diphthong, compared to words like *bible* (for related proposals see Connine et al. 1997, Clopper & Walker 2017, and discussion below). Another possibility (Bybee 2002b) is that the greater degree of dissimilarity between the [t]-final vs. flapped variants in raised paradigms like *bite* ~ *biting* should slow lexical access compared to raised paradigms without flapping like *bike* ~ *biking* (although this version of exemplar matching becomes harder to distinguish from the phonological-inference account spelled out above).

In this study, we measured fixations during word recognition to words that demonstrate Canadian raising and/or intervocalic flapping in their surface forms and hoped to thereby reveal something about their underlying representations. We compare the timing of activation of forms like *bible* [baɪbəl], *biking* [bʌɪkɪŋ], and *biting* [bʌɪɾɪŋ] to determine the mechanism(s) used during activation and, in turn, the phonological structure of their representations. If listeners are using phonological inference, we would expect the slowing of word recognition, such that words in which phonological processes have applied are recognized more slowly, and more phonological processes result in additional slowing: *bible* should be faster than *biking* should be faster than *biting*. This timing result would suggest that listeners' representations of words with raising and flapping are abstract, that is, more distant from their surface forms than words that show only one or neither process. However, if listeners are storing both raised and unraised vowels (and both flaps and stops) and using exemplar matching to map the signal to those representations, then we should see a different timing pattern, depending on the exemplar model assumed. Under a simple account, we might expect to see no timing differences between *bible*, *biking*, and *biting*, except as influenced by their relative frequencies; or we might expect slowing of all forms with multiple dialect variants, meaning *bible* is recognized faster than *biking* or *biting*; or we might see slowing of recognition in forms with more varied paradigms, such that *bible* and *biking* are recognized faster than *biting*. We return to these more complex exemplar approaches, in light of our data, in §5.2 of the general discussion.

We report the results of two studies. In each case, participants' productions were measured to confirm Canadian raising. Then a visual-world paradigm experiment tested the recognition of words that had undergone Canadian raising, intervocalic flapping, both, or neither. We report measures of speed of target fixations and interpret those results with regard to word-recognition mechanisms like phonological inference and exemplar matching and their consequences for lexical representations. We focus on fixations to the TARGET item (as opposed to competitor items) because we are specifically interested in how phonological processes may affect the speed and efficiency with which the target representations are accessed.

3. Experiment 1.

3.1. Methods.

Design. Twelve sets of four words were selected. Each set contained four specific word-types: an UNRAISED word in which the diphthong preceded a voiced consonant (e.g. *bible* [baɪbəl]), a RAISED word in which the diphthong preceded a voiceless consonant (e.g. *biking* [bʌɪkɪŋ]), an OPAQUE word in which the diphthong preceded a flap derived from a /t/ (e.g. *biting* [bʌɪɾɪŋ]), and a phonologically unrelated word (e.g. *stapler*). Each of the three related words in a set were cohorts. Because of the limited number of English word sets that share the specific properties required here, we had to choose from a variety of parts of speech (nouns, verbs, and adjectives). In each case, we chose a word that was deemed picturable. In ten of the twelve sets, the opaque word was morphologically complex; we tried to use as many morphologically complex words as possible, so that there was clear evidence that the flap was derived from /t/. The full list of items and their frequencies is given in the appendix.

The lexical frequency of the three experimental word-types (unraised, raised, and opaque) was assessed, as frequency is known to affect lexical activation and competition (Marslen-Wilson 1990). The SUBTLEXus corpus of American English subtitles (Brysbaert & New 2009) was used to determine frequency counts; analyses were performed on the log word-frequency measure. One word, *fifer*, did not appear in this database. We assigned that word a frequency of 0.001 (log frequency = log10(1+0.001) = 0.000434). An ANOVA revealed marginal frequency differences across word-types, $F(2,33) = 3.085$, $p = 0.059$. This effect was driven by the fact that the log frequency in the raised words is lower than that of the other two sets (raised vs. opaque, $p = 0.039$; raised vs. unraised, $p = 0.038$). Because of these frequency differences, we controlled for log frequency in the statistical analyses reported below.

STIMULI.[5] Auditory stimuli were recorded at 44.1 kHz on a Blue Yeti microphone in Praat (Boersma & Weenink 2014) by a trained speaker of Canadian English who was also a phonetician. He produced each word ten times in the carrier phrase 'He said __ ', making sure that the vowel was raised in forms like *biking* and *biting*. Additionally, he produced the opaque words (like *biting*) a second time, this time with an unraised vowel ([baɪɾɪŋ]). We refer to this additional set as FLAPPED words, meaning they are variants of the opaque set that contain flaps, but not raising. Though this pronunciation was not native to the speaker's dialect, he felt confident that his phonetic training would allow him to produce unraised tokens naturally. The first author of this paper, who speaks a dialect of English with no raising, vetted the productions and agreed that they sounded natural.

Following Hall (2005), the F1 and F2 of each diphthong were extracted from the recorded stimuli at the point of maximum F1 (the lowest part of the vowel) in Praat. The best five tokens of each stimulus were selected to maximize the height difference between the raised and unraised words in each set. ANOVAs on the 240 selected tokens (5 tokens × 4 words × 12 sets) confirmed significant differences in diphthong height by word-type as measured by F1: $F(3,236) = 270.4$, $p < 0.001$. Follow-up pairwise *t*-tests indicated a height difference in raised vowels compared to unraised vowels ($p < 0.001$), raised vowels compared to flapped vowels ($p < 0.001$), and opaque vowels compared to flapped vowels ($p < 0.001$), but no differences between raised and opaque vowels ($p = 0.297$) or between unraised and flapped vowels ($p = 0.234$).

Finally, we wanted to remove vowel duration as an additional cue to the voicing of the upcoming segment, as vowels are expected to be longer before voiced segments than voiceless ones (Peterson & Lehiste 1960). We thus equated the vowel length for each set of words. To do this, we first calculated the average duration of all the diphthongs in a set (i.e. five tokens each of the unraised, raised, opaque, and flapped words in each set). We then determined the ratio of diphthong length of the tokens for each word to the total average and increased or decreased the overall length of each diphthong by that ratio, using Praat's duration manipulation function. We scaled the entire diphthong, retaining the original ratio of nucleus to offglide and preserving the dynamics of the formant trajectories. For example, in the set mentioned above, the average diphthong duration of the twenty tokens of *biking*, *bible*, and *biting* was 178 ms. The tokens for *biking*, however, had an average of 165 ms. Thus each diphthong token for *biking* was increased by a factor of 1.08. This meant that the five tokens of a word each retained their natural variation in diphthong duration, but overall, the average diphthong duration for a given word was equal to the average duration of every other diphthong in that set. This removed vowel duration as a cue to upcoming voicing in as natural a way as possible. Upon listening to the resulting tokens, the first author and a team of native-Canadian-English-speaking research assistants agreed that the tokens sounded natural and the manipulation was not noticeable. One hundred milliseconds of silence were added before the onset of each target word.

Visual stimuli consisted of a series of clip-art pictures depicting each word in the study. Pictures were selected to be as canonical as possible and were edited for color and stylistic consistency. All were approved by a member of the team with extensive experience with the visual-world paradigm. After editing, each image was sized to 300 × 300 pixels.

For each word/image pairing, a sample sentence was constructed. The sentence contained the target word in final position and did not contain any other raised (or poten-

[5] Stimuli and the measurements discussed below are available from the first author upon request.

tially raised) diphthongs. For example, the sentence for *biking* was *On sunny days, Peter loves to go biking*.

Participants. Participants were twenty-six adults between the ages of seventeen and thirty-five; in the final analysis, three participants were excluded because of experimenter error, leaving twenty-three participants in the analysis. All were native monolingual speakers of Canadian English who had been born and raised in Canada. All reported normal hearing and normal or corrected-to-normal vision. Participants either were paid with a $10 gift card per session, or received course credit for their participation.

Procedure. Each participant completed two tasks across two sessions, scheduled approximately one week apart. During the first session, participants were first recorded reading the sample sentences containing the stimulus words. Participants saw a screen with the image used in the experiment, the word in isolation, and a sentence containing the word. These displays were randomized in advance and the same order was used for each participant; words were not blocked by type. Participants were asked to view each image and read aloud the text below it (both the word in isolation and the word in a sentence); they read through the full set of displays twice in succession. This task served two purposes: first, because of the limited number of picturable triplets, some of words used were not especially common (e.g. *fifer*), and we felt the participants would benefit from a short introduction to the pictures and the words they were meant to portray. Second, the participant recordings allowed us to measure whether the participants were speakers of the Canadian raising dialect and, if so, whether they raised in opaque forms and how strongly they raised. Because of an error that was not detected until later, the sentences for *tiger* and *spiral* were accidentally left off the training list. Because both of these words are fairly recognizable nouns, we felt that this would not skew the data excessively.

The second task was the eye-tracking task. First, each participant was calibrated in the eye-tracker. After calibration and instructions, the experiment began. Each of the forty-eight words was repeated twenty times (four repetitions of each of five tokens) for a total of 960 trials. The opaque words were further split: half of their occurrences (ten repetitions) contained a raised vowel, and the other half an unraised vowel. Presentation order was randomized for each participant, with no effort made to avoid identical back-to-back trials. The 960 total trials were split across the two sessions, with 480 trials occurring each session. The first session thus lasted about an hour (recording and eye-tracking), and the second session lasted about forty-five minutes (eye-tracking only).

Participants received instructions at the beginning of the experiment, both on the screen and from the experimenter, to click on the center circle when it turned blue, and then to click on the picture that matched the word they heard. Each trial began with the four pictures in the corners of the screen, outer edges located 50 pixels from the edge of the screen, and a red circle (60 pixels in diameter) in the center of the screen, equidistant from the four pictures (see Figure 1). Location of the four items for each trial was randomized, so that the position of each item in the set was completely unpredictable, across each of the targets' twenty presentations. Participants had the opportunity to familiarize themselves with the images and their locations. After 500 milliseconds, the circle turned blue and participants clicked on it to play the auditory stimulus (the target word in isolation). This procedure minimized the chances that participants were fixating any of the pictures when the auditory stimulus began.

Recordings. Participants' productions of the target words and sample sentences were recorded with a Blue Yeti USB microphone in Praat. These recordings were chunked

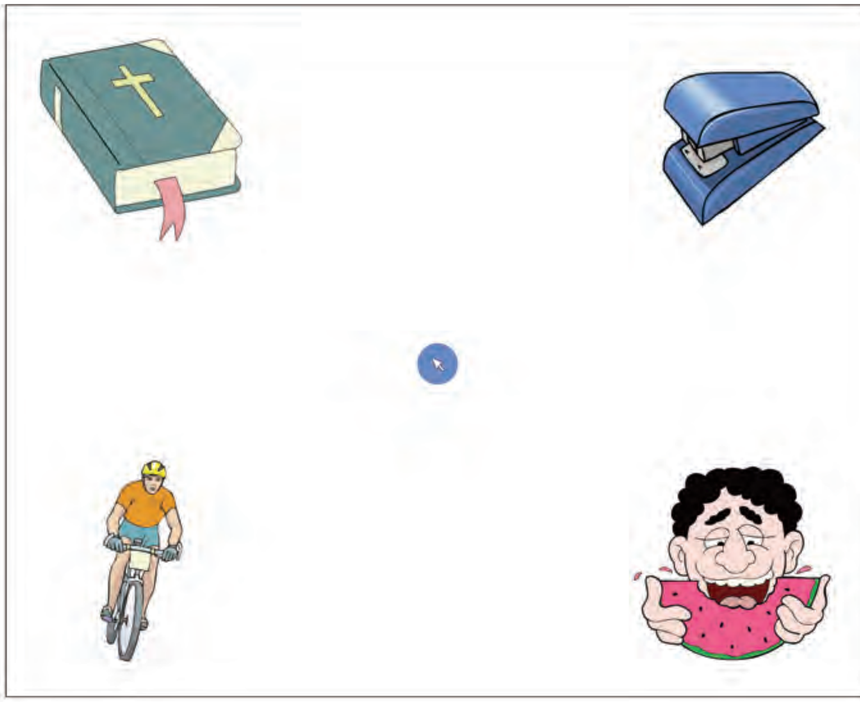

FIGURE 1. Sample trial screen containing images for (clockwise from top left) *bible*, *stapler*, *biting*, and *biking*.

and annotated so that measurements of maximum F1 and F2 at the point of maximum F1 could be automated in Praat.

EYE-TRACKING. Eye movements were monitored and recorded with an Eyelink 1000 with a 250 Hz sampling rate. Participants rested their chins and forehead in a headrest. The eye tracker was calibrated with a standard nine-point calibration procedure. Breaks were offered every thirty-two trials, and a drift correct was performed at the end of each break to ensure that the eye track was still valid. Fixations were recorded from onset of the display screen in each trial until the participant made a mouse-click response.

3.2. RESULTS.

ACCURACY. Participants were highly accurate, suggesting that they did not have trouble recognizing the less familiar stimuli (like *fifer*). On average, participants made four errors in 960 trials (range: 0–40). Only trials in which the participants clicked on the correct target were included in the eye-tracking analysis.

PRODUCTION. To confirm that our participants spoke a raising dialect of Canadian English, we again measured F1 and F2 at maximum F1 in each diphthong from the words produced during the training portion. Figure 2 shows each of the raised and unraised vowels plotted in F1–F2 space, with very little overlap between the raised/opaque vowels and the unraised vowels. The F1 and F2 measures were analyzed in linear mixed-effects models using lme4 in R (R Core Team 2013, Bates et al. 2015) with word-type as a fixed factor, log frequency (centered) as a covariate, and random intercepts and slopes on participant and random intercepts on item (as in 3). Word-type was coded with Helmert contrast codes here and in all analyses in this report. Helmert coding is appropriate for ordered levels (in this case, ordered by the number of phonological processes) and allows us to compare each level to the average of the subsequent levels. Here we compare unraised (+.66) vowels to the average of raised (−.33) and opaque vowels (−.33), and raised vowels (+.5) to opaque vowels (−.5). The numeric values are assigned such that the values for each set of comparisons sum to zero and the different comparisons are orthogonal to one another.

(3) F1/F2 ~ wordtype + logfreq + (1 + wordtype * logfreq | participant) + (1 | item)

Results (Tables 1 and 2) showed that F1 was higher and F2 was lower in unraised words than in raised and opaque words; raised words had marginally lower F1 than opaque words, but F2 values did not differ significantly. Word frequency was not predictive of F1 or F2. Thus, our participants were clear Canadian raisers (following the definition of a difference of at least 60 Hz between the raised and unraised vowels; Labov, Ash, & Boberg 2006), and this holds consistently across participants, as can be seen in Figure 2. The plot shows that there was no overlap in F1/F2 space between the raised and unraised words for any of the participants, while only a single participant produced opaque words whose vowels were similar to unraised words.[6]

WORD-TYPE	F1 MAXIMUM (Bark / Hz)	F2 AT F1 MAXIMUM (Bark / Hz)
$\bar{x}_{raised}$	6.56 / 711	11.89 / 1737
$\bar{x}_{unraised}$	7.98 / 919	11.18 / 1554
$\bar{x}_{opaque}$	6.76 / 738	11.79 / 1709

TABLE 1. Mean F1 and F2 values in Bark and Hz for each word-type in experiment 1.

	EST	*SE*	*df*	*t*	*p*
F1					
(intercept)	7.1	0.1	30.2	60.4	< 0.001
unraised vs. raised/opaque	1.3	0.1	47.9	9.2	< 0.001
raised vs. opaque	−0.3	0.1	35.7	1.97	0.0571
log frequency (covariate)	−0.1	0.1	30.6	1.3	0.1895
F2					
(intercept)	11.6	0.1	37.7	88.0	< 0.001
unraised vs. raised/opaque	−0.6	0.2	34.7	4.0	< 0.001
raised vs. opaque	0.1	0.2	32.8	0.5	0.623
log frequency (covariate)	−0.01	0.1	30.0	0.1	0.888

TABLE 2. Test of fixed effects of word-type on F1 and F2 values in experiment 1.

FIXATION MEASUREMENTS. As our primary measure of lexical activation, participants' fixations to the target were measured. We used two fixation measures to analyze target activation: target-fixation latency and midpoint of the averaged target-fixation function, both described in further detail below. Target-fixation latency is an early measure of lexical activation, while midpoint is a later measure.

TARGET LATENCY. Our first analysis aimed to determine the earliest fixation to the target item for each trial that could be driven by the auditory stimulus, that is, a visual analogue to reaction time (e.g. Dahan et al. 2001, Beddor et al. 2013). We thus measured latency from the onset of the auditory stimulus (excluding the 100 ms of silence), and we excluded trials in which the participant was fixating the target item at 200 ms after onset; since it takes approximately 200 ms to plan and launch an eye movement (Viviani 1990), fixations to the target so early in the trial are unlikely to be driven by the auditory stimulus. This resulted in the exclusion of 1,347 trials (6.1%). In the remaining

[6] Participants did make use of both F1 and F2 to signal raising; some participants distinguished the categories by F1, others by F2, and still others with both. Nevertheless, clear category boundaries for unraised vs. raised/opaque words emerged.

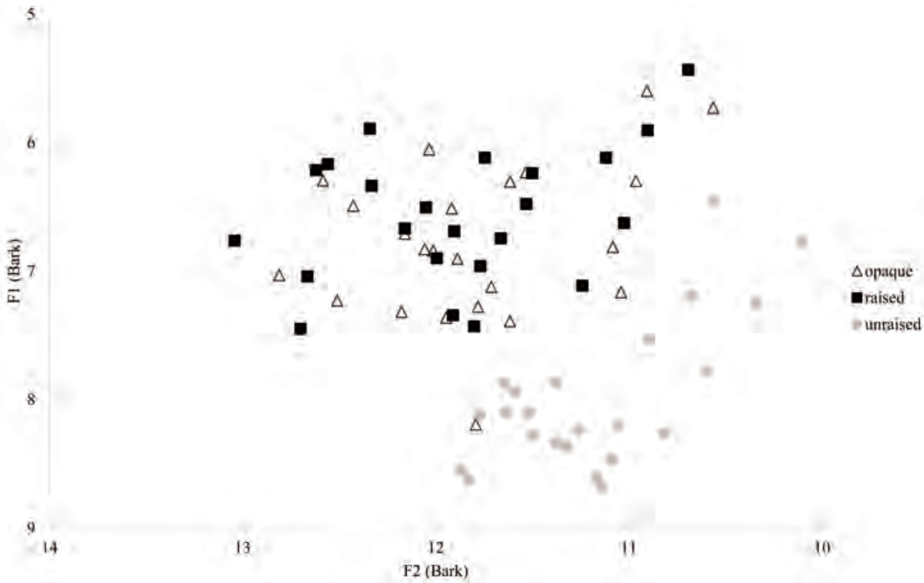

Figure 2. Average vowel productions by participant in experiment 1.

trials, the first fixation to the unraised words occurred the fastest, on average, followed by first fixation to the raised, opaque, and flapped words, in that order (Table 3).

word-type	mean latency in ms (*SD*)
unraised	428 (73)
raised	459 (78)
opaque	496 (87)
flapped	518 (85)

Table 3. Mean target latency in experiment 1.

Trial-by-trial latency results were analyzed with a linear mixed-effects regression model. Model comparison revealed that the model in 4 provided the best fit;[7] it included a fixed effect of word-type and log frequency (centered) as a covariate. Three Helmert contrast codes represented the four word-types: flapped (+.75) vs. the other word-types (−.25), unraised (+.66) vs. raised and opaque (−.33), and raised (+.5) vs. opaque (−.5). Random effects included random intercepts and slopes on participant and random intercepts on item set.

(4) firstFixation ~ wordtype + logfreq + (1 + wordtype + logfreq | participant) + (1 | item-set)

The results in Table 4 reveal that there were significant effects of each word-type comparison. Taking each condition in turn: participants were slowest to make an initial fixation to the flapped words (which were not native to the participants' dialect; we return to this in §3.3); they were faster to make fixations to the unraised words than the raised or opaque words, and they were faster to make fixations to the raised words than the opaque words. Frequency was also a significant predictor of target-fixation latency, such that participants were faster to fixate more frequent words, but not to the exclusion of the other factors.

[7] A model in which word-type and frequency interacted in the random-effects structure failed to converge.

	EST	*SE*	*df*	*t*	*p*
(intercept)	776.7	18.7	31.8	41.5	< 0.001
flapped vs. others	64.2	7.6	24.8	8.5	< 0.001
unraised vs. raised/opaque	−40.8	7.0	23.5	5.8	< 0.001
raised vs. opaque	−56.1	7.4	24.7	7.6	< 0.001
log frequency (covariate)	−26.7	3.8	26.3	7.1	< 0.001

TABLE 4. Test of fixed effects of word-type on target-fixation latency in experiment 1.

These results suggest that lexical activation within our experiment was influenced by at least three factors: lexical frequency, expectation about dialect, and number of phonological processes. For words that were consistent with the participants' dialect, the more phonological processes, the slower the word was to be activated.

TARGET MIDPOINT. Our second analysis examined the midpoints of the target-fixation functions. This is an additional measure of timing that reflects, for a given participant or item, the point at which approximately half of all fixations were to the target item. While target-fixation latency allows us to analyze the immediacy of activation, the midpoint analysis allows us to examine the later time course of activation, taking into account not just the decision to fixate the target, but also the decision to maintain that fixation and not to look away at competitors. The midpoint analysis tells us, in a sense, whether the initial target activation was sustained over the course of the trial.

This analysis utilizes a curve fitted to each participant's or item's averaged proportion of fixations as a function of time. We followed previous research (e.g. McMurray, Aslin, et al. 2008, McMurray, Clayards, et al. 2008, McMurray et al. 2010, Toscano & McMurray 2012, 2014) in dividing the eye-movement record into saccades (rapid eye movements which jump from one point to another) and fixations (continuous eye gaze to a single point). Because the saccade that leads to a fixation is likely driven by a response to the auditory stimulus, each fixation was paired with its preceding saccade. These saccade-fixation combinations were termed *looks*, which were then averaged across items or participants. This allowed us to calculate, at each 4 ms time point, what proportion of looks were to a given item on the screen.

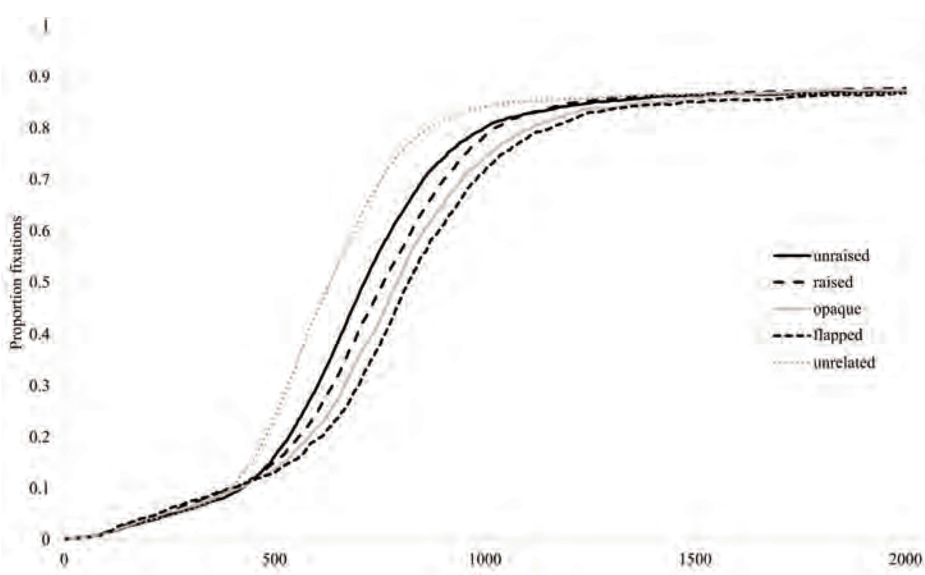

FIGURE 3. Target fixations by word-type in experiment 1.

Figure 3 shows fixations to the target item by item type (raised, unraised, opaque, flapped, and unrelated filler) over time. Visual inspection of the plot reveals that participants were fastest to fixate the unrelated item, which is unsurprising given that it shared no phonological content with the other items on the screen. The figure also suggests differences in timing of fixation to the other items, such that unraised items were fixated faster than raised, which were fixated faster than opaque, and flapped items were fixated slowest. To quantify these differences, we fit a mathematical model to the fixation data for each participant or item in order to determine the curve parameters that correspond to timing and degree of fixations. The target curve equation (following McMurray et al. 2010) is provided in 5. This model returns four parameters for each curve: the minimum and maximum asymptotes (*b* and *p*, respectively, corresponding to minimum and maximum proportion of fixations), the midpoint (*c*, corresponding to timing of fixations), and the slope (*s*, corresponding to the rate of increase in fixations).

$$(5)\ \ P(target) = \frac{p - b}{1 + \exp\left(4 \cdot \dfrac{s}{p - b} \cdot (c - t)\right)} + b$$

To fit a model to a fixation curve, the curve itself needs to be relatively robust, so we must generalize over either participants or items, rather than accounting for effects of item and participant simultaneously. Below we report analyses that generalize over participants and over items.

WORD-TYPE	MIDPOINT MEAN IN MS (*SD*)
unraised (*bible*)	689 (60)
raised (*biking*)	729 (73)
opaque (*b*[ʌɪɾ]*ing*)	761 (79)
flapped (*b*[aɪɾ]*ing*)	791 (74)

TABLE 5. Average midpoint values for each word-type in experiment 1.

	EST	*SE*	*df*	*t*	*p*
Participant					
(intercept)	742.6	14.3	22	52.0	< 0.001
flapped vs. others	64.3	6.0	66	10.7	< 0.001
unraised vs. raised/opaque	−56.5	6.4	66	8.8	< 0.001
raised vs. opaque	−32.2	7.4	66	4.4	< 0.001
Item					
(intercept)	800.5	23.7	41.3	33.8	< 0.001
flapped vs. others	72.3	17.0	32.0	4.3	0.0002
unraised vs. raised/opaque	−47.6	18.2	32.2	2.6	0.0135
raised vs. opaque	−51.2	21.8	32.7	2.3	0.0252
log frequency (covariate)	−27.8	9.8	39.3	2.8	0.0070

TABLE 6. Test of fixed effects of word-type on target midpoint by participant and item in experiment 1.

We compared the parameter most closely related to speed or timing of fixations—target midpoint (Table 5)—across the four target word-types (raised, unraised, opaque, and flapped). We examined the data with two linear mixed-effects models, one each for fixations averaged over participants and items. In each case, we used the same Helmert contrast codes described above. In the first, fixations by participant were analyzed with word-type as a fixed effect and by-participant random intercepts. In the second model, fixations by item set were analyzed with word-type as a fixed effect, log frequency as a covariate, and by-item-set random intercepts. This analysis models fixations by item sets (rather than individual items) to preserve the relationships among items in a set;

that is, *biking* never appears without *bible* and *biting*, and modeling effects of item set rather than individual item captures this. Model syntax is in 6.

(6) a. analysis by participants: crossover ~ wordtype + (1 | participant)
b. analysis by items: crossover ~ wordtype + logfreq + (1 | item-set)

Results from the two models are in Table 6. In each case, the flapped words had a significantly later midpoint than the other words. Unraised words had significantly earlier midpoints than raised words (whether transparent or opaque), and raised transparent words had significantly earlier midpoints than opaque words. In sum, the fewer the phonological processes, the earlier the target midpoint was. However, words that had undergone flapping without raising were an exception—again, these items were shown to be fixated most slowly, although they had undergone only a single phonological process, an anomaly we discuss immediately below.

3.3. Discussion. Experiment 1 showed that not all words were activated equally quickly; with the exception of the flapped-only condition, words with more phonological processes had longer first-fixation latencies and later target midpoints than those with fewer phonological processes.

The exception to our generalization that connects more phonological processes with slower fixations was that words that had undergone flapping without raising were the slowest to be fixated. This is not surprising when the listeners' native dialect is considered. We included flapped forms with unraised vowels, for example, *biting* as [baɪɾɪŋ], with the goal of comparing words that had undergone only raising (*biking*) with words that had undergone only flapping—but of course these are also ungrammatical pronunciations of an input /…aɪtV…/ sequence in our participants' dialect.[8] We hypothesize that participants were slow to activate these targets because they were so unexpected; they required participants to reevaluate their assumptions about forms like *biting*. It is possible, though, that this resulted in participants learning over the course of the experiment that vowel height was not a reliable cue, or perhaps interpreting the speaker as not a member of the Canadian raising dialect. Either of these cases could have caused the slower processing of the raised and opaque forms seen in experiment 1, invalidating our findings. Experiment 2 was designed to replicate experiment 1 without the confound of unraised opaque forms.

4. Experiment 2. Experiment 2 sought to eliminate the potential confound in experiment 1. We wanted to be sure that the delayed fixations we saw in words like *biking* and *biting* were the result of a slowing in overall activation, rather than an uncertainty brought on by the stimuli themselves. Experiment 2 thus replicated experiment 1, with a few small changes detailed below.

4.1. Methods.

Design. Experiment 2 used the same twelve sets of words used in experiment 1, but opaque words like *biting* were presented only with raised vowels ([bʌɪɾɪŋ]). Participants thus never heard words that were mispronounced in their dialect.

Stimuli. The same stimuli used in experiment 1 were used here, excluding the flapped forms. Visual stimuli and target sentences for recording were also identical.

Participants. Participants were twenty-five adults between the ages of seventeen and thirty-five; in the final analysis, one participant was excluded because it later came

[8] Note that in this example, the form [baɪɾɪŋ] would be associated with the word *biding*, which was not pictured on the screen. This was not true for all of our flapped forms, though; only five of our twelve flapped forms had real-word minimal pairs with medial /d/: *biding*, *siding*, *riding*, *tidal*, and *wider*.

to our attention that the participant was bilingual. All others were native, monolingual speakers of Canadian English born and raised in Canada. All reported normal hearing and normal or corrected-to-normal vision. Participants either were paid with a $10 gift card per session, or received course credit for their participation.

Procedure. Preliminary analyses from experiment 1 showed that participants' eye-tracking behavior in the two sessions of that experiment was comparable. For that reason, experiment 2 was conducted in only one session. Each participant completed the same recording task (words accidentally excluded in experiment 1 were included in experiment 2). This was followed by the eye-tracking task, in which each of the five recorded tokens for each word was presented twice, for a total of 480 trials (12 sets × 4 words per set × 5 tokens per word × 2 repetitions). This single session, including both recording and eye-tracking, lasted about an hour. All other aspects of the procedure were the same as in experiment 1.

Eye-tracking. The set-up and calibration for the eye-tracking task were identical to those in experiment 1.

4.2. Results.

Accuracy. As in experiment 1, participants were highly accurate. On average, participants made three errors in 480 trials (range: 0–9). Again, only correct trials were analyzed in the eye-tracking analysis.

Production. We again analyzed F1 and F2 at maximum F1 to determine whether participants' productions showed evidence of Canadian raising (Tables 7 and 8). Using a linear mixed-effects model with the same structure used in experiment 1, we found that unraised words had higher F1 and lower F2 than raised and opaque words. Raised and opaque words did not differ in F1 or F2, and word frequency was not a significant predictor of either. As is apparent from Figure 4, the participants in experiment 2 were also overwhelmingly Canadian raisers.

word-type	F1 maximum (Bark / Hz)	F2 at F1 maximum (Bark / Hz)
$\bar{x}_{raised}$	6.47 / 702	12.11 / 1795
$\bar{x}_{unraised}$	7.78 / 887	11.21 / 1563
$\bar{x}_{opaque}$	6.61 / 721	11.99 / 1764

Table 7. Mean F1 and F2 values in Bark and Hz for each word-type in experiment 2.

	est	*SE*	*df*	*t*	*p*
F1					
(intercept)	6.9	0.2	27.4	43.8	< 0.001
unraised vs. raised/opaque	1.2	0.1	40.1	10.1	< 0.001
raised vs. opaque	−0.2	0.1	33.9	1.5	0.153
log frequency (covariate)	−0.1	0.1	33.1	0.9	0.377
F2					
(intercept)	11.8	0.1	33.3	88.2	< 0.001
unraised vs. raised/opaque	−0.8	0.1	36.5	6.2	< 0.001
raised vs. opaque	0.1	0.2	34.3	0.8	0.424
log frequency (covariate)	0.009	0.1	34.4	0.1	0.906

Table 8. Test of fixed effects of word-type on F1 and F2 values in experiment 2.

Fixation measurements. Fixations to target by word-type are shown in Figure 5. The same comparisons of target fixations were carried out for experiment 2, including both target-fixation latency and target midpoint.

Target-fixation latency. Again we excluded trials in which the participant was fixating the target at 200 ms after the onset of the auditory stimulus. This resulted in the ex-

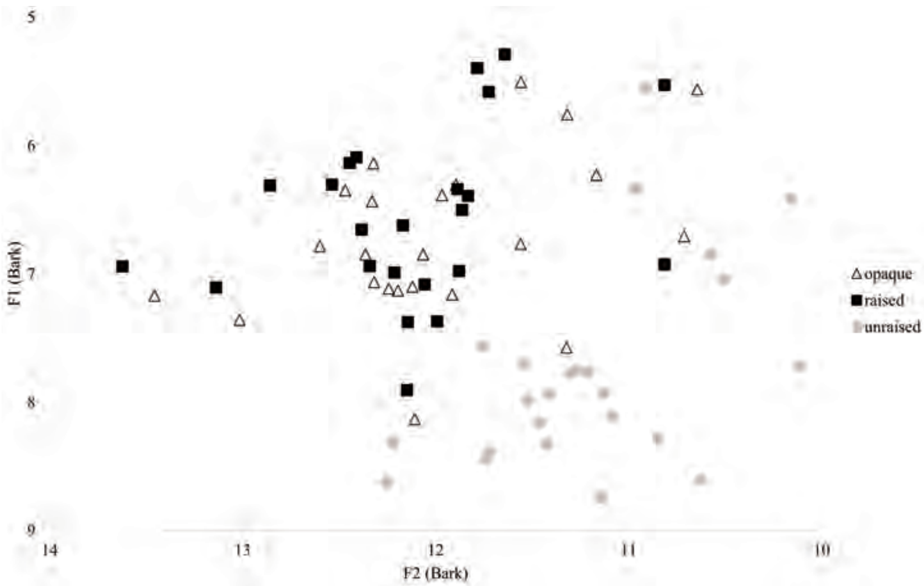

FIGURE 4. Productions by participant in experiment 2.

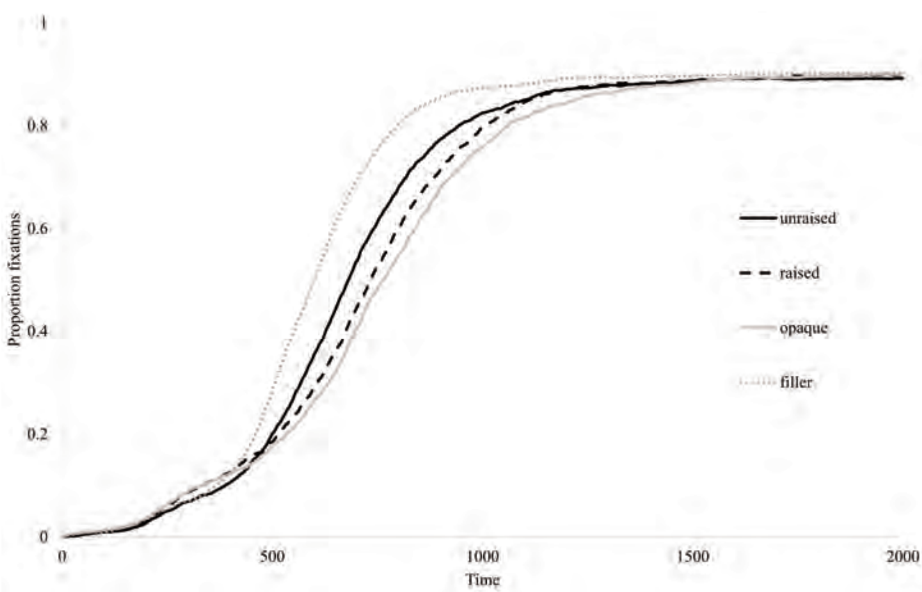

FIGURE 5. Target fixations by word-type in experiment 2.

clusion of 902 trials, or 7.8% of all trials. The linear mixed-effects model was identical to that used in experiment 1, except that because there were no flapped words (i.e. words like *biting* were always produced with raised vowels), one fewer Helmert contrast code was necessary. The data were thus coded to represent three word-types: unraised (+.66) vs. raised and opaque (−.33), and raised (+.5) vs. opaque (−.5). Results are given in Tables 9 and 10. Unraised words had a significantly shorter target-fixation latency than raised and opaque words; raised words had a significantly shorter latency than opaque words. Log frequency was also a significant predictor of target-fixation latency.

WORD-TYPE	MEAN IN MS (*SD*)
unraised	417 (93)
raised	448 (109)
opaque	480 (104)

TABLE 9. Mean target latency in experiment 2.

	EST	*SE*	*df*	*t*	*p*
(intercept)	746.7	22.3	32.0	33.5	< 0.001
unraised vs. raised/opaque	−39.2	8.2	22.1	4.8	< 0.001
raised vs. opaque	−55.7	9.6	24.7	5.8	< 0.001
log frequency (covariate)	−30.7	4.2	46.4	7.2	< 0.001

TABLE 10. Test of fixed effects of word-type on target-fixation latency in experiment 2.

TARGET MIDPOINT. We used the same target midpoint analysis used in experiment 1, but with the Helmert contrast codes defined in the target-fixation latency analysis above. The model structure was identical to that in the same analysis in experiment 1. Midpoints are given in Table 11, and results of the linear mixed-effects models in Table 12. Unraised words had significantly earlier midpoints than raised and opaque words; raised words had earlier midpoints than opaque words.

WORD-TYPE	MIDPOINT MEAN IN MS (*SD*)
unraised (*bible)*	664 (67)
raised (*biking*)	709 (84)
opaque (*b*[ʌɪɾ]*ing*)	743 (82)

TABLE 11. Average midpoint values for each word-type in experiment 2.

	EST	*SE*	*df*	*t*	*p*
Participant					
(intercept)	705.1	14.8	23	47.6	< 0.001
unraised vs. raised/opaque	−61.6	8.8	46	−7.0	< 0.001
raised vs. opaque	−33.9	10.1	46	−3.4	0.0016
Item					
(intercept)	777.6	28.3	31.8	27.5	< 0.001
unraised vs. raised/opaque	−53.75	20.9	21.2	−2.6	0.0175
raised vs. opaque	−59.73	25.2	21.6	−2.4	0.0271
log frequency (covariate)	−36.27	12.4	25.9	−2.9	0.0069

TABLE 12. Test of fixed effects of word-type on target midpoint by participant and item in experiment 2.

4.3. DISCUSSION. The results of experiment 2 replicate those of experiment 1. Participants were fastest to activate unraised words and slowest to activate opaque words, while raised words were intermediate. These results indicate that the flapped words in experiment 1, which were not characteristic of the participants' native dialect, did not drive the results.

5. GENERAL DISCUSSION.

5.1. ON THE NATURE OF PHONOLOGICAL REPRESENTATIONS. The results of the two experiments presented here support the claim that listeners store abstract segments in the lexical representations. In particular, they support representations similar to 2a whereby the raising diphthongs are stored unraised and surface flaps are stored as stops, and they use phonological inference to access these representations during word recognition. In words with raised diphthongs, like *biking*, the diphthong occurs in a raising context and the listener must work back to the abstract unraised vowel. In opaque words, the raised vowel occurs before a flap, so the listener must both undo the flap (to identify the underlying /t/) and then work back to the abstract vowel. It seems that each step requires additional cognitive resources, slowing processing.

Exemplar matching, at least in its most basic guise, does not provide an account of these results, because if listeners were matching surface forms to stored exemplars as in 2b, they would not be any slower to fixate an opaque word than a simple unraised one. But it is also more difficult to reconcile our findings with an account that posits an underlying contrast between raised and unraised vowels in some or all environments and a phonology built upon the contrast. For example, Pater (2014) argues for a raised/unraised contrast in the preflap environment, such that *title* has an underlyingly raised diphthong while *tidal* has an underlyingly unraised diphthong. A word outside the contrastive environment, like *biking*, could also have an underlyingly unraised diphthong on this account or, by lexicon optimization, could have an underlyingly raised diphthong. In the former case, we would predict that recognition of *biking* would proceed more slowly than recognition of *biting*, because the listener would have to use phonological inference to uncover the representation for *biking* but not for *biting*. In the latter case, we would predict recognition of *biking* and *biting* to proceed at the same rate. Neither of these scenarios is in line with our results.

Our findings may also have implications for the type of phonological grammar that speakers use to map from underlying input to surface output, though those implications remain murky. As mentioned above, phonological inference directly mirrors a serial rule-ordering approach, in which the rules of vowel raising and stop flapping apply in a counterbleeding order (Kiparsky 1973). In such grammars, a phonological change occurs because a process-oriented rule requires it: if a rule states that /aɪ/ → [ʌɪ] before voiceless stop, and the rule is fed the string /… aɪt …./, raising will happen with no regard to whether that voiceless stop will survive to the surface. In contrast, theories that apply all of their processes in parallel—such as optimality theory (Prince & Smolensky 2004 [1993]) and HARMONIC GRAMMAR (Legendre, Miyata, & Smolensky 1990a,b)—are not structurally designed to capture any serial nature of sound decoding in lexical recognition. Given the widespread agreement that constraint-based grammars are better at capturing crosslinguistic phonological generalizations, this may simply indicate that lexical recognition cannot be considered phonological production run ‘in reverse’. If, however, the serial nature of word recognition is interpreted as support for a serial, derivational model of underlying-to-surface phonology, it must further be noted that ANY framework which imposes changes via surface-oriented constraints—even the derivational theory of HARMONIC SERIALISM (McCarthy 2010)—is unable in its basic formulations to capture counterbleeding opacity. Given this state of phonological theorizing, we must leave open the extent to which serial word recognition informs any choice between phonological grammars.

To summarize our contribution to the larger phonological debate, we quote one of Pater’s (2014:237) main claims: that ‘patterns that usually are claimed to require intermediate derivational stages and abstract underlying representations, of which Canadian Raising and flapping is the canonical case, can at least sometimes be analyzed without these formal mechanisms. How many of them can be, whether they should be … are all questions for further research’. Our eye-tracking results provide one such type of counterevidence as to ‘whether they should be’—rather, that abstract underlying representations in raised/flapped words SHOULD be maintained.

5.2. ALTERNATIVE ACCOUNTS. While we have argued against an account that relies on pure storage and exemplar matching, some more nuanced possibilities remain. One alternative is to consider the morphological paradigms of our items. A simple exemplar account would store both *biking* and *biting* with raised vowels, predicting no need for slower processing of *biting*. However, a somewhat more complex account (e.g. Bybee 1988, 2000, 2001, 2002b, Hay & Baayen 2005, Pierrehumbert 2016) would link different morphological forms of a word, connecting, for example, the derived form *biting*, stored with a stem-final flap, with its morphological base form *bite*, stored with a final [t]. A similar link would connect *biking* to its morphological base *bike*, both stored with a stem-final [k]. On this account, the variation between /t/ and flap in the *bite* ~ *biting* paradigm could result in slower access of the derived form compared with the invariant *bike* ~ *biking* pair. However, this account does not explain why *biking* would be slower than *bible*: both of these word sets would have consistent representations of the diphthong and their following stops, so the slower activation we found for *biking*- vs. *bible*-type words would not be accounted for. More importantly, not all of the words in the study shared morphological structure. There were both monomorphs and bimorphemic words in each word-type (e.g. *bible* vs. *wiser* in the unraised set, *biking* vs. *cyclops* in the raised set, and *biting* vs. *vitals* in the unraised set). This makes it difficult to make generalizations about processing based on the morphological structure of our words.[9]

A different type of alternative account comes from the range of English dialects spoken and available to speakers living in Vancouver who participated in our study, and the potential 'purity' of their Canadian raising dialects, at least with respect to perception. All of the participants in our study are likely exposed to a great deal of American English, through media and in personal contact with American friends, relatives, coworkers, and teachers. Under an exemplar-based account, it is certainly possible that these listeners have stored both raised and unraised pronunciations of words like *biting* and *biking*, in contrast to their uniformly unraised exemplar cloud for words like *bible*. If speakers have a larger variety of vowel exemplars for raised words, that could explain their slower activation; the additional slowing in flapped words could also be the result of more exemplars if nonflapped exemplars are stored as well. Under this scenario, unraised words are recognized the fastest because all of their exemplars share a single vowel. Raised words like *biking* are slowed because listeners have both [bʌɪkɪŋ] and [baɪkɪŋ] exemplars, and opaque words like *biting* are slowest because listeners have all of [bʌɪɾɪŋ], [baɪɾɪŋ], [bʌɪtɪŋ], and [baɪtɪŋ].

While we cannot completely rule out this explanation, some research on variant recognition across dialects casts doubt on it. First, intervocalic flapping is overwhelmingly prevalent in potential flapping contexts in North American English (likelihood of flapping in these contexts is reported as anywhere from 93.9% (Patterson & Connine 2001) to 99% (Zue & Laferriere 1979)). While we expect that our Canadian English speakers hear many unraised tokens, it is hard to imagine that they hear very many unflapped tokens, calling into question the additional slowing explanation outlined above. With regard to the influence of dialect, it is clear that experience with a particular dialect influences a listener's ability to recognize words within that dialect, such that greater experience is likely to facilitate recognition (Otake & Cutler 1999, Sumner & Samuel 2009). On the one hand, Clopper and Walker (2017) suggest that listeners who have been exposed to multiple dialects may develop a strategy whereby competitors

[9] Unfortunately, our word sets were not balanced for the number of morphemes, making further analysis of this issue difficult. Nearly all (10/12) of the opaque words were bimorphemic, in an effort to signal underlying /t/ structure. By contrast, very few of the unraised words (3/12) were bimorphemic, and the raised words were split (6/12). This makes it impractical to do an analysis by number of morphemes. We suspect that there may be complex interactions between lexical frequency, word-type, and number of morphemes, but our item sets were unfortunately not designed to test that.

stay active longer, perhaps leading to a slowing in activation as we observed. On the other hand, research by Sumner and Samuel (2009) suggests that listeners who are familiar with but not speakers of a dialect may not store dialect variants in their underlying representations. This would suggest that our Canadian raising dialect speakers, who are familiar with nonraising dialects of American English but do not speak them, may not store these tokens as exemplars. Moreover, long-term exposure to an accent (as our participants would have to American English) leads to normalization (Floccia et al. 2006), perhaps decreasing the chance that our participants would be influenced by the American English variants. Finally, we note that listeners in our experiment 1 were especially slow to fixate the target picture when they heard unraised opaque words like [baɪɾɪŋ], which are standard in the American English dialect. This suggests that these exemplars were not accessed automatically, and thus may not have played a strong role in inhibiting recognition of raised opaque words.

Another alternative we have not yet discussed is the role of competitors during word recognition. A key tenet of almost every theory of word recognition is that listeners activate multiple competitors that are consistent with an acoustic input, eliminating competitors as the input unfolds and they are no longer consistent with it. The presence of multiple competitors thus slows activation of the target (Vitevitch et al. 1999). One might argue that in our experiment, *biking* and *biting* are stronger competitors for one another on the surface (since both have raised vowels), while *bible* does not have a strong surface competitor, as it is the only item of the four with an unraised vowel. This might explain the slower activation of *biking* and *biting* as compared to *bible*. Here again, however, the number of competitors does not account for why *biting* is activated more slowly than *biking*; there, the difference is in number of (transparent) phonological processes, rather than in number of competitors.

5.3. Other aspects of our results. There remains the contribution of phonological opacity. As already discussed, the cause of vowel raising in words like *biting* is not surface-apparent, in that its triggering voiceless stop at the end of *bite* has been obscured. It is not clear whether opacity itself plays a role in our findings. Within a phonological-inference account, there are at least two possible explanations for the additional slowing in opaque words that we are unable to tease apart. On the one hand, it could be the case that undoing two processes (raising and flapping) simply requires more processing than undoing a single process, and the delay associated with phonological inference is thus additive. Alternatively, though, it may be the case that listeners are delayed by the context for the raised vowel in opaque words. Here, the raised vowel occurs before a flap, which may not be a clear licenser of raising. This may slow the inference process, as the flap licenses both raised and unraised vowels. In the future, examining a transparent multistep interaction of processes (i.e. a feeding interaction) will be necessary to answer this question.

Appendix: Stimuli[10]

	Unraised			Raised	
Word	FPM	Log freq	Word	FPM	Log freq
bible	18.33	2.9713	biking	0.55	1.4624
silo	0.94	1.6902	cyclops	1.27	1.8195
fire	215.49	4.0410	fifer[11]	—	0.00043

[10] Frequency per million (FPM) and log frequency from the SUBTLEXus frequency corpus (Brysbaert & New 2009).

[11] Note that *fifer* did not appear in the SUBTLEXus corpus, so we assigned it a raw frequency value of 0.001.

	UNRAISED			RAISED	
WORD	FPM	LOG FREQ	WORD	FPM	LOG FREQ
liner	1.22	1.7993	license	32.06	3.2138
miser	0.49	1.4150	Mikey	0.02	0.30103
nylons	1.25	1.8129	Nike	1.00	1.7160
rising	8.41	2.6335	riper	0.06	0.6021
spiral	1.78	1.9638	spicy	3.31	2.2304
tiger	18.53	2.9759	typing	3.24	2.2201
timer	6.25	2.5051	typhus	0.51	1.4314
wiser	2.59	2.1239	wipers	0.43	1.3617
visor	0.59	1.4914	viper	2.53	2.1139
MEAN	23.00	2.29		3.75	1.54
SD	61.00	0.77		9.00	0.91

	OPAQUE			FILLER	
WORD	FPM	LOG FREQ	WORD	FPM	LOG FREQ
biting	4.90	2.3997	stapler	0.86	1.6532
sighting	1.55	1.9031	washer	2.04	2.0212
fighter	12.78	2.8149	lockers	1.45	1.8751
lighter	8.96	2.6609	cookie	16.71	2.9309
mighty	26.57	3.1323	sailor	12.39	2.8014
nightie	0.69	1.5563	clover	1.31	1.8325
writing	55.92	3.4553	honey	300.49	4.1854
spiting	0.02	0.3010	jacket	33.41	3.2317
tighter	4.08	2.3201	cherries	1.96	2.0043
title	18.57	2.9768	dragon	19.29	2.9934
whiter	0.73	1.5798	hammer	12.47	2.8041
vitals	3.61	2.2672	bubble	8.00	2.6117
MEAN	11.53	2.28		34.20	2.58
SD	16.19	0.86		84.43	0.74

REFERENCES

ALLOPENNA, PAUL; JAMES S. MAGNUSON; and MICHAEL K. TANENHAUS. 1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38.419–39. DOI: 10.1006/jmla.1997.2558.

BATES, DOUGLAS; MARTIN MAECHLER; BEN BOLKER; and STEVE WALKER. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1–48. DOI: 10.18637/jss.v067.i01.

BEDDOR, PATRICE SPEETER; KEVIN B. MCGOWAN; JULIE E. BOLAND; ANDRIES W. COETZEE; and ANTHONY BRASHER. 2013. The time course of perception of coarticulation. *The Journal of the Acoustical Society of America* 133.2350–66. DOI: 10.1121/1.4794366.

BERKSON, KELLY; STUART DAVIS; and ALYSSA STRICKLER. 2017. What does incipient /ay/-raising look like? A response to Josef Fruehwald. *Language* 93.e181–e191. DOI: 10.1353/lan.2017.0050.

BOERSMA, PAUL, and DAVID WEENINK. 2014. Praat: Doing phonetics by computer. Online: http://www.praat.org/.

BRYSBAERT, MARC, and BORIS NEW. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41.977–90. DOI: 10.3758/BRM.41.4.977.

BYBEE, JOAN L. 1988. Morphology as lexical organization. *Theoretical morphology: Approaches in modern linguistics*, ed. by Michael Hammond and Michael Noonan, 119–41. San Diego: Academic Press.

BYBEE, JOAN L. 2000. Lexicalization of sound change and alternating environments. *Papers in laboratory phonology 5*, ed. by Michael Broe and Janet Pierrehumbert, 250–68. Cambridge: Cambridge University Press.

Bybee, Joan L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan L. 2002a. Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition* 24.215–21. DOI: 10.1017/S0272263102002061.

Bybee, Joan L. 2002b. Regular morphology and the lexicon. *Language and Cognitive Processes* 10.425–55. DOI: 10.1080/01690969508407111.

Chambers, Jack K. 1973. Canadian raising. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 18.113–35. DOI: 10.1017/S0008413100007350.

Chambers, Jack K. 1989. Canadian raising: Blocking, fronting, etc. *American Speech* 64. 75–88. DOI: 10.2307/455114.

Chomsky, Noam. 1964. *Current issues in linguistic theory*. The Hague: Mouton.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Clopper, Cynthia G., and Abby Walker. 2017. Effects of lexical competition and dialect exposure on phonological priming. *Language and Speech* 60.85–109. DOI: 10.1177/0023830916643737.

Connine, Cynthia M. 2004. It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review* 11.1084–89. DOI: 10.3758/BF03196741.

Connine, Cynthia M.; Larissa J. Ranbom; and David J. Patterson. 2008. Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics* 70.403–11. DOI: 10.3758/PP.70.3.403.

Connine, Cynthia M.; Debra Titone; Thomas Deelman; and Dawn Blasko. 1997. Similarity mapping in spoken word recognition. *Journal of Memory and Language* 37. 463–80. DOI: 10.1006/jmla.1997.2535.

Dahan, Delphine; James S. Magnuson; Michael K. Tanenhaus; and E. Hogan. 2001. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes* 16.507–34. DOI: 10.1080/01690960143000074.

Dailey-O'Cain, Jennifer. 1997. Canadian raising in a midwestern US city. *Language Variation and Change* 9.107–20. DOI: 10.1017/S0954394500001812.

Ernestus, Mirjam, and R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79.5–38. DOI: 10.1353/lan.2003.0076.

Ernestus, Mirjam; R. Harald Baayen; and Rob Schreuder. 2002. The recognition of reduced word forms. *Brain and Language* 81.162–73. DOI: 10.1006/brln.2001.2514.

Eulitz, Carsten, and Aditi Lahiri. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience* 16.577–83. DOI: 10.1162/089892904323057308.

Farris-Trimble, Ashley, and Bob McMurray. 2018. Morpho-phonological regularities influence the dynamics of real-time word recognition: Evidence from artificial language learning. *Laboratory Phonology* 9:2. DOI: 10.5334/labphon.41.

Floccia, Caroline; Jeremy Goslin; Frédérique Girard; and Gabrielle Konopczynski. 2006. Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance* 32.1276–93. DOI: 10.1037/0096-1523.32.5.1276.

Fruehwald, Josef. 2016. The early influence of phonology on a phonetic change. *Language* 92.376–410. DOI: 10.1353/lan.2016.0041.

Gaskell, M. Gareth, and William D. Marslen-Wilson. 1996. Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22.144–58. DOI: 10.1037/0096-1523.22.1.144.

Gaskell, M. Gareth, and William D. Marslen-Wilson. 1997. Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes* 12.613–56. DOI: 10.1080/016909697386646.

Gaskell, M. Gareth, and William D. Marslen-Wilson. 1998. Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 24.380–96. DOI: 10.1037/0096-1523.24.2.380.

Gaskell, M. Gareth; Elsa Spinelli; and Fanny Meunier. 2002. Perception of resyllabification in French. *Memory & Cognition* 30.798–810. DOI: 10.3758/BF03196435.

Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22.1166–83. DOI: 10.1037/0278-7393.22.5.1166.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105.251–79. DOI: 10.1037/0033-295X.105.2.251.

Goldinger, Stephen D. 2007. A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken, 49–54. Online: http://www.icphs2007.de/.

Hall, Kathleen Currie. 2005. Defining phonological rules over lexical neighbourhoods: Evidence from Canadian raising. *West Coast Conference on Formal Linguistics (WCCFL)* 24.191–99. Online: http://www.lingref.com/cpp/wccfl/24/paper1223.pdf.

Halle, Morris. 1962. Phonology in generative grammar. *Word* 18.54–72. DOI: 10.1080/00437956.1962.11659765.

Harris, Zellig S. 1951. *Structural linguistics*. Chicago: University of Chicago Press.

Hay, Jennifer B., and R. Harald Baayen. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* 9.342–48. DOI: 10.1016/j.tics.2005.04.002.

Herd, Wendy; Allard Jongman; and Joan Soreno. 2010. An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics* 38.504–16. DOI: 10.1016/j.wocn.2010.06.003.

Hintzman, Douglas L. 1984. MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers* 16.96–101. DOI: 10.3758/BF03202365.

Hintzman, Douglas L. 1986. 'Schema abstraction' in a multiple-trace memory model. *Psychological Review* 93.411–28. DOI: 10.1037/0033-295X.93.4.411.

Hintzman, Douglas L. 1988. Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review* 95.528–51. DOI: 10.1037/0033-295X.95.4.528.

Hualde, José Ignacio; Tatiana Luchkina; and Christopher D. Eager. 2017. Canadian raising in Chicagoland: The production and perception of a marginal contrast. *Journal of Phonetics* 65.15–44. DOI: 10.1016/j.wocn.2017.06.001.

Idsardi, William J. 2006. Canadian raising, opacity, and rephonemicization. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 51.119–26. DOI: 10.1353/cjl.2008.0007.

Johnson, Keith. 1997. Speech perception without talker normalization. *Talker variability in speech processing*, ed. by Keith Johnson and John W. Mullenix, 145–65. San Diego: Academic Press.

Joos, Martin. 1942. A phonological dilemma in Canadian English. *Language* 18.141–44. DOI: 10.2307/408979.

Kahn, Daniel. 1980. *Syllable-based generalizations in English phonology*. New York: Garland.

Kiparsky, Paul. 1971. Historical linguistics. *A survey of linguistic science*, ed. by William Orr Dingwall, 576–649. College Park: Linguistics Program, University of Maryland.

Kiparsky, Paul. 1973. Abstractness, opacity, and global rules. *Three dimensions of linguistic theory*, ed. by Osamu Fujimura, 57–86. Tokyo: TEC.

Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10.421–41. Online: https://www.jstor.org/stable/4178120.

Kolinsky, Régine; José Morais; and Mireille Cluytens. 1995. Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language* 34.19–40. DOI: 10.1006/jmla.1995.1002.

Labov, William; Sherry Ash; and Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. New York: Mouton de Gruyter.

Lahiri, Aditi, and William Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38.245–94. DOI: 10.1016/0010-0277(91)90008-R.

Legendre, Géraldine; Yoshiro Miyata; and Paul Smolensky. 1990a. Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: An application. *Proceedings of the 12th annual conference of the Cognitive Science Society (CogSci 1990)*, 884–91. Hillsdale, NJ: Lawrence Erlbaum. Online: https://mindmodeling.org/cogscihistorical/cogsci_12.pdf.

Legendre, Géraldine; Yoshiro Miyata; and Paul Smolensky. 1990b. Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. *Proceedings of the 12th annual conference of the Cognitive Science Society (CogSci 1990)*, 388–95. Hillsdale, NJ: Lawrence Erlbaum. Online: https://mindmodeling.org/cogscihistorical/cogsci_12.pdf.

LoCasto, Paul C., and Cynthia M. Connine. 2002. Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception & Psychophysics* 64.208–19. DOI: 10.3758/BF03195787.

Luce, Paul A.; Conor T. McLennan; and Jan Charles-Luce. 2003. Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. *Rethinking implicit memory*, ed. by Jeffrey S. Bowers and Chad J. Marsolek, 197–214. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780192632326.003.0009.

Luce, Paul A., and David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19.1–36.

Marslen-Wilson, William D. 1990. Activation, competition, and frequency in lexical access. *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, ed. by Gerry T. M. Altmann, 148–72. Cambridge, MA: MIT Press.

Marslen-Wilson, William D.; Andy Nix; and Gareth Gaskell. 1995. Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes* 10.285–308. DOI: 10.1080/01690969508407097.

Marslen-Wilson, William D., and Paul Warren. 1994. Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review* 101.653–74. DOI: 10.1037/0033-295X.101.4.653.

McCarthy, John J.. 2010. An introduction to harmonic serialism. *Language and Linguistics Compass* 4.1001–18. DOI: 10.1111/j.1749-818X.2010.00240.x.

McClelland, James L., and Jeffrey L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18.1–86. DOI: 10.1016/0010-0285(86)90015-0.

McMurray, Bob; Richard N. Aslin; Michael K. Tanenhaus; Michael J. Spivey; and Dana Subik. 2008. Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance* 34.1609–31. DOI: 10.1037/a0011747.

McMurray, Bob; Meghan A. Clayards; Michael K. Tanenhaus; and Richard N. Aslin. 2008. Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review* 15.1064–71. DOI: 10.3758/PBR.15.6.1064.

McMurray, Bob; Vicki M. Samelson; Sung Hee Lee; and J. Bruce Tomblin. 2010. Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology* 60.1–39. DOI: 10.1016/j.cogpsych.2009.06.003.

Mielke, Jeff; Mike Armstrong; and Elizabeth Hume. 2003. Looking through opacity. *Theoretical Linguistics* 29.123–39. DOI: 10.1515/thli.29.1-2.123.

Mitterer, Holger, and Leo Blomert. 2003. Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics* 65.956–69. DOI: 10.3758/BF03194826.

Mitterer, Holger, and Mirjam Ernestus. 2006. Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34.73–103. DOI: 10.1016/j.wocn.2005.03.003.

Norris, Dennis. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52.189–234. DOI: 10.1016/0010-0277(94)90043-4.

Otake, Takashi, and Anne Cutler. 1999. Perception of suprasegmental structure in a non-native dialect. *Journal of Phonetics* 27.229–53. DOI: 10.1006/jpho.1999.0095.

Paradis, Carole. 1980. La règle de Canadian raising et l'analyse en structure syllabique. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 25.35–45. DOI: 10.1017/S0008413100009166.

Pater, Joe. 2014. Canadian raising with language-specific weighted constraints. *Language* 90.230–40. DOI: 10.1353/lan.2014.0009.

Patterson, David J., and Cynthia M. Connine. 2001. Variant frequency in flap production. *Phonetica* 58.254–75. DOI: 10.1159/000046178.

Peterson, Gordon E., and Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32.693–703. DOI: 10.1121/1.1908183.

Pierrehumbert, Janet. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2.33–52. DOI: 10.1146/annurev-linguistics-030514-125050.

Pinker, Steven, and Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28.73–193. DOI: 10.1016/0010-0277(88)90032-7.

Pinnow, Eleni, and Cynthia M. Connine. 2014. Phonological variant recognition: Representations and rules. *Language and Speech* 57.42–67. DOI: 10.1177/0023830913479105.

Pitt, Mark 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: 10.1016/j.jml.2009.02.005.

Pitt, Mark A.; Laura Dilley; and Michael Tat. 2011. Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics* 3.304–11. DOI: 10.1016/j.wocn.2010.07.004.

Prince, Alan, and Paul Smolensky. 2004 [1993]. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.

R Core Team. 2013. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: http://www.r-project.org/.

Spinelli, Elsa; James M. McQueen; and Anne Cutler. 2003. Processing resyllabified words in French. *Journal of Memory and Language* 48.233–54. DOI: 10.1016/S0749-596X(02)00513-2.

Sumner, Meghan, and Arthur G. Samuel. 2005. Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language* 52.322–38. DOI: 10.1016/j.jml.2004.11.004.

Sumner, Meghan, and Arthur G. Samuel. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60.487–501. DOI: 10.1016/j.jml.2009.01.001.

Toscano, Joseph C., and Bob McMurray. 2012. Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics* 74.1284–1301. DOI: 10.3758/s13414-012-0306-z.

Toscano, Joseph C., and Bob McMurray. 2014. The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience* 30.529–43. DOI: 10.1080/23273798.2014.946427.

Vance, Timothy J. 1987. ‘Canadian raising’ in some dialects of the northern United States. *American Speech* 62.195–210. DOI: 10.2307/454805.

Vitevitch, Michael S.; Paul A. Luce; David B. Pisoni; and Edward T. Auer. 1999. Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language* 68.306–11. DOI: 10.1006/brln.1999.2116.

Viviani, Paolo. 1990. Eye movements in visual search: Cognitive, perceptual, and motor control aspects. *Eye movements and their role in visual cognitive processes* (Reviews of oculomotor research 4), ed. by Eileen Kowler, 353–93. Amsterdam: Elsevier.

Zue, Victor W., and Martha Laferriere. 1979. Acoustic study of medial /t,d/ in American English. *The Journal of the Acoustical Society of America* 66.1039–50. DOI: 10.1121/1.383323.

Zuraw, Kie. 2000. *Patterned exceptions in phonology*. Los Angeles: University of California, Los Angeles dissertation.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28.417–72. DOI: 10.1007/s11049-010-9095-z.

[afarrist@sfu.ca]
[am.tessier@ubc.ca]

[Received 10 November 2017;
revision invited 17 December 2017;
revision received 30 May 2018;
revision invited 3 June 2018;
revision received 14 June 2018;
accepted 18 June 2018]