

Online perception of glottalized coda stops in American English

Adam J. Chong^{a,b} and Marc Garellek^c

^aDepartment of Linguistics, Queen Mary University of London

London, UK

^bDepartment of Linguistics, University of California, Los Angeles

Los Angeles, California, USA

^cDepartment of Linguistics, University of California, San Diego

La Jolla, California, USA

Authors' email addresses

^aa.chong@qmul.ac.uk

^bmgarellek@ucsd.edu

Online perception of glottalized coda stops in American English

Abstract

In American English, voiceless codas /t/ and /p/ are often glottalized: they have glottal constriction that results in creaky voice on the preceding vowel. Previous claims suggest that such glottalization can serve to enhance /t/ or, more generally, voicelessness of coda stops. In this study, we examine the timecourse of word recognition to test whether glottalization facilitates the perception of words ending in voiceless /t/ and /p/, which is expected if glottalization is in fact enhancing. Sixty American English listeners participated in an eye-tracking study, where they heard resynthesized glottalized and non-glottalized versions of CVC English words ending in /p, t, b, d/ while looking at a display with two words presented orthographically. Target words were presented with a minimal pair differing in place of articulation (e.g. cop-cot), or voicing, (e.g. bat-bad, cap-cab). Although there is little evidence that glottalization facilitates recognition of words ending in /t/ or /p/, there is a strong inhibitory effect: words ending in voiced stops are recognized more slowly and poorly when the preceding vowel was glottalized. These findings lend little support to a listener-driven, enhancement-based explanation for the occurrence of coda glottalization in American English. On the other hand, they suggest that glottalized instances of coda /t/ and /p/, but not of coda /d/ and /b/, are perceived as equally good variants of these sounds.

Keywords: glottalization, eye-tracking, enhancement, phonological variants

1. Introduction

Coda glottalization is the process by which coda stops are produced either with simultaneous glottal constriction or with glottal constriction that has replaced the oral gesture. For example, in American English (as in many other varieties) the word ‘bat’ /bæt/ may be pronounced without coda glottalization [bæt] (or with another non-glottal variant of /t/, such as [ɾ]), or with coda glottalization [bæ̠t̚, bæ̠ʔ]. In this study, we focus on instances of coda glottalization where the glottal constriction co-occurs with the oral one, e.g. /bæt/ → [bæ̠t̚]. Using a word recognition task with eye-tracking, we examine the timecourse of processing of coda glottalization in American English, to determine whether listeners use voice quality information anticipatorily to perceive voicing and place of articulation contrasts in coda stops.

1.1. Where, and how often, does glottalization occur in American English?

In this study, we use the term “glottalization” to refer to a phonological phenomenon characterized by increased vocal fold constriction associated with a particular segment. (In English and other languages, there are additional sources of vocal fold constriction, e.g. phrasal creak, which act on larger prosodic units). Phonetically, glottalization is usually realized as creaky voice, with or without a glottal stop [ʔ]. Different sources of glottalization exist, including word-initial glottalization (e.g. of words beginning with a stressed vowel: Dilley et al., 1996; Garellek, 2013; Davidson & Erker, 2014), and coda glottalization, which we focus on here.

Glottalized variants of coda /t/ include cases where /t/ is realized as a glottal stop with no alveolar closure (/t/ → [ʔ]), and cases where /t/ is ‘reinforced’ with glottal closure in addition to the alveolar closure (/t/ → [t̚, t̚ʔ]). The reinforced instances of /t/ are often unreleased (Esling et al., 2005), though released variants with glottalization, which we fo-

cus on here, are also attested (Seyfarth & Garellek, 2015; Penney et al., 2018). Altogether, glottalized variants of /t/ are common in American English, especially before sonorants (Huffman, 2005; Pierrehumbert, 1994; Sumner & Samuel, 2005; Seyfarth & Garellek, 2015). For instance, Seyfarth & Garellek (2015) show that speakers from the Buckeye Corpus (Pitt et al., 2007) glottalize coda /t/ as glottal stop [ʔ] roughly 50% of the time. Glottalized variants of coda /t/ in general (both [t̚] and [ʔ]) are more frequent than any other variant of /t/, including the canonical [t] variant. Previous studies have come to similar conclusions (Pierrehumbert, 1994; Huffman, 2005): in a corpus of Long Island English, Huffman (2005) reports that coda /t/ glottalizes 58% of the time.

Studies have also shown that coda /p/ in American English undergoes glottalization to [p̚] (but not to [ʔ]), though glottalization rates for /p/ are much lower than for /t/ (Pierrehumbert, 1994; Huffman, 2005, Seyfarth & Garellek, 2015). In Long Island English, glottalization of coda /p/ occurs roughly 29% of the time (Huffman, 2005). Thus, glottalized variant [p̚], though attested in American English, cannot be considered the most common variant of /p/ in coda position, in contrast to glottalized variants of /t/. Lastly, glottalized variants of other voiceless sounds (e.g. /k/ or /tʃ/) and voiced sounds (e.g. /d/) are not attested in American English, but can be found in varieties of British English (Roach, 1973; Foulkes & Docherty, 2007).

1.2. Glottalization as enhancement of voice and place contrasts

Several researchers have proposed that glottalized variants of coda stops are due to phonetic enhancement. One enhancement-based explanation for coda glottalization is that it occurs as a mechanism to inhibit voicing and/or to cue voicelessness (Stevens & Keyser, 1989; Pierrehumbert, 1994, 1995; Keyser & Stevens, 2006; Stevens & Keyser, 2010; Gordeeva & Scobbie, 2013; Penney et al., 2018). The reason why voiceless coda stops might involve glottalization in order to inhibit voicing is that these very sounds are likely to undergo voic-

ing: utterance-medially, stop voicing is favored aerodynamically (Westbury & Keating, 1986); coda stops are also often immediately preceded by a vowel, which has strong voicing, and are frequently followed by a voiced onset. For instance, it is well known that coda /t/ is more likely to glottalize before a sonorant (Pierrehumbert, 1994; Huffman, 2005; Keyser & Stevens, 2006; Seyfarth & Garellek, 2015). If the stop is glottalized, it is by definition produced with increased vocal fold constriction. This constriction should help cease voicing, ensuring that the voiceless coda stop is in fact voiceless.

However, the enhancement account of /t/ glottalization cannot explain why /t/ and /d/ often neutralize to a coronal tap intervocalically. Moreover, Huffman (2005) scrutinized this account on empirical grounds. She showed that coda glottalization was prevalent even before voiceless consonants, which is unexpected if coda glottalization should be used to enhance voicelessness. Moreover, Huffman found that phrase-final coda stops were not more likely to glottalize before (phrase-initial) sonorants than obstruents; assuming that phrase-initial obstruents have weaker voicing than phrase-initial sonorants, this finding is unexpected if /t/ and /p/ could stand to be enhanced before sonorants. Therefore, though glottalization is likelier to occur before sonorants than obstruents phrase-medially, this effect is weakened phrase-finally (see also Seyfarth & Garellek, 2015); more importantly – and contrary to an explanation rooted in enhancement of voicelessness – glottalization still occurs on coda stops that precede voiceless sounds.

As an alternative enhancement-based explanation, Keyser & Stevens (2006) further propose that all voiceless stops are likely to glottalize phrase-finally because *voiced* stops are likely to devoice in that position (see, e.g. Westbury & Keating, 1986). To compensate for devoicing of voiced coda stops, voiceless stops undergo glottalization. Thus, according to this explanation, glottalization of phrase-final voiceless stops occurs not because voicelessness is difficult to produce or perceptually “imperiled” in that position, but as a conse-

quence of the fact that the voicing during voiced stops is weakened. Glottalization of voiceless stops thereby helps enhance the voicing contrast phrase-finally.

Note also that an explanation involving voiceless enhancement leaves open the question regarding distributional differences: i.e., in American English, why does coda /t/ glottalize so frequently, /p/ only sometimes, and /k/ not at all? This issue is addressed by Stevens & Keyser (2010), who propose that coda glottalization also occurs to cue closure and release specifically for coda /t/. Coronal closure and release are more subject to gestural overlap than labial or velar constrictions for /p/ or /k/. Since overlap weakens recoverability of a particular feature or segment, /t/ glottalizes when the voiceless closure and coronal release would be obscured by a non-coronal (Keyser & Stevens, 2006; Stevens & Keyser, 2010). Seyfarth & Garellek (2015), however, show that glottalization rates for coda /t/ before non-coronals are roughly the same as rates of utterance-final /t/, for which there is no chance of overlap.

Overall it is clear that coda glottalization in American English is strongly associated with voiceless stops, and that it occurs more frequently with coda /t/ than /p/, and rarely with /k/. And while enhancement-based explanations for coda glottalization may not be able to account for all instances of the phenomenon, they dominate in the literature. In this paper, we test a central assumption of these enhancement-based accounts, namely, that glottalization (as an enhancement gesture/feature) is *perceptually* beneficial to listeners. If so, we expect that listeners should be more likely to recognize words with coda /t/ and /p/ when they are glottalized. In other words, the presence of glottalization should result in faster and better recognition of words ending in voiceless stops, especially /t/.

1.3. Perception of phonological variants and coarticulation

In addition to testing predictions of enhancement theories, this study has implications for understanding how phonological variants are perceived. Several researchers have examined

listeners' ability to recognize familiar words produced with a glottal variant of /t/ both in word-medial (Pitt et al., 2011) as well as in word-final coda position (Sumner & Samuel, 2005). In particular, word-final glottal variants of /t/ (both [\widehat{t} or ?]) were as effective as canonical non-glottalized [t] at priming a semantically-related target. This provides evidence that glottalization does not hinder word recognition. Although glottalization does not hinder recognition of coda /t/, its effect on coda /p/ and voiced stops, however, is yet to be determined.

Glottalization is phonologically associated with a coda stop, but it is additionally (or even entirely) realized phonetically as creaky voice on the preceding vowel (Pierrehumbert, 1994; Huffman, 2005; Garellek, 2015; Seyfarth & Garellek, 2015). Thus, a word like 'bat', when glottalized, may be realized as [bæt, bæ̰]. Listeners might therefore use creaky voice (derived from glottalization) as a cue to anticipate the upcoming coda consonant. In the current study, we investigate this question using a variant of the visual world paradigm (Allopenna et al., 1998), where listeners' eye movements on a visual display are continuously monitored while they are listening to speech stimuli. This paradigm allows us to assess listeners' interpretation, shown through their visual fixations, of the incoming speech signal, and how this changes over time as the signal unfolds.

A number of studies have used eye tracking to assess listeners' perception of coarticulated speech. For example, Dahan et al. (2001) cross-spliced the initial CV of minimal pairs of words such as *neck* and *net* to create tokens which either contained matching coarticulatory cues to the final coda consonant (e.g. [nɛ] in *net* cross-spliced with [t] from another token of *net*) or mismatching cues (e.g. [nɛ] in *neck* cross-spliced with [t] from *net*). Although the mismatching tokens were eventually perceived as *net*, the formant transition cues in the vowel were more consistent with *neck*. Listeners, in fact, made more initial fixations to the competitor image of a neck when they heard *net* tokens with mismatch cues than when they heard tokens with matching cues (see also Dahan & Tanenhaus, 2004 on Dutch

and Gow & McMurray, 2007 on C-C coarticulation in English). Thus listeners make use of the information present in the speech signal immediately and their interpretations are incrementally updated as the speech signal unfolds. More importantly, the inappropriateness of a particularly coarticulatory cue can *hinder* the recognition process by garden-pathing listeners' initial interpretations.

In fact, listeners are not just sensitive to the appropriateness of coarticulatory cues for a given upcoming segment, but are also sensitive to the specific timing of when coarticulatory cues become available in the speech signal. Beddor et al. (2013) investigated the timecourse of listeners' perception of coarticulatory vowel nasalization in American English. Participants heard cross-spliced target words (e.g. *bent*) with nasalization starting early or late in the vowel while looking at two pictures on a display. Listeners' fixations converged on the target more quickly when nasalization started early in the vowel than when it started late in the vowel. These results thus show that lexical access follows closely the unfolding of acoustic cues in the speech signal (see also Salverda et al., 2014). Given the visual world paradigm's (Allopenna et al., 1998) success at elucidating listeners' online perception of coarticulatory cues, in this study, we use this paradigm to examine listeners' online perception of coda glottalization in American English.

1.4. Hypotheses of the present study

The previous work discussed above has shown that both the frequency of phonological variants and the coarticulatory information they provide can influence online word recognition. In the case of glottalization, we therefore expect that listeners should show online sensitivity to its acoustic presence, and that they should have prior expectations of which sounds are likelier than others to be glottalized. The research on phonetic enhancement also makes clear (but not yet verified) predictions regarding how listeners should perceive acoustic information deemed "enhancing": if an enhancement feature is present, the sound it enhances

should be perceived more rapidly and identified more accurately. If the presence of glottalization is in fact due to enhancement (either for /t/ specifically, or for voiceless stops more generally), we expect listeners to recognize a target word faster and/or better when it is present than when it is absent.

In sum, we examine here two hypotheses related to coda glottalization in American English. First, since glottalization occurs more frequently with /t/ than with /p/, glottalization should preferentially facilitate recognition of /t/ words more than /p/ words. Second, if glottalization is associated with voiceless stops more generally, as Keyser & Stevens (2006) suggest, then glottalization should facilitate recognition of voiceless stops (both /t/ and /p/) over voiced stops.

2. Methods

In this study, we use eye-tracking to measure online word recognition using visual fixations to printed words. In brief, eye movements are tracked as participants hear instructions to interact with a display, such as “look at the word *bat*”. The display contains two words, e.g. the target *bat* and a competitor like *bad*, represented as printed words. Following previous researchers (e.g. McQueen & Viebahn, 2007; Mitterer, 2011; Brouwer et al., 2012), we chose to use the printed-word variant of the visual world paradigm (rather than the traditional variant with images of objects), because of difficulty representing certain targets visually. Crucially, the target and other words presented on the screen differed minimally in their orthography; for example, English *bat* was paired with *bad*.

2.1. Participants

Sixty participants (20 M, 46 F; Mean age: 20.56, range: 18-38) were recruited from the UCLA Psychology Subject Pool and received course credit in an undergraduate psychology

or linguistics class. Participants were all adult native speakers of American English without any reported hearing deficits. All had normal or corrected-to-normal vision and could recognize the visual stimuli used. An additional 13 participants were also tested, but their data were excluded from data analysis because they were not native English speakers ($n = 7$), because of inattention during the task ($n = 4$), or because of failure of the system to save the experimental data ($n = 2$).

2.2. Target words

The target words were monosyllabic CVC English words. Words were chosen such that they had a minimal pair differing in terms of the coda obstruent. Pairs of words were also controlled for orthographic length and were organized into four groups, each consisting of four word pairs:

The Baseline group

In the first “Baseline” group, we tested whether the presence of glottalization resulted in differences in eye-tracking behavior even when glottalization is not associated with a regular glottalized variant of a coda. Thus for this condition we used words differing only in their final consonant, which could be either a voiceless alveolar fricative /s/ or a voiceless dental fricative /θ/, e.g. ‘moss’ vs. ‘moth.’ Neither fricative is regularly glottalized in coda position. The complete wordlist is shown in Table 1 (the pairs do not differ significantly in frequency, as assessed by a paired t-test).

Table 1. *Target words used in the Baseline condition. Frequencies come from the SUBTLEXus corpus (Brybaert & New, 2009).*

Word 1	Log frequency	Word 2	Log frequency
mass	2.945	math	2.9227
moss	2.1644	moth	2.0682

worse	3.6992	worth	3.7459
pass	3.7415	path	3.098

For these words, the initial consonant and vowel preceding the final fricative were always drawn from the same original token (always the alveolar, e.g. ‘moss’), to minimize the acoustic differences between words in these pairs. Although in natural speech, formant transitions before alveolar vs. dental fricatives differ slightly (Jongman et al., 1985), our pilot experiments revealed that both words in a pair (e.g. ‘moss’ and ‘moth’) sounded equally natural when resynthesized with the same pre-alveolar vowel. Overall then, for this group we predict no overall facilitative or inhibitory effect of glottalization, because (1) neither glottalized $[\widehat{s}]$ nor $[\widehat{\theta}]$ are regular variants of coda /s/ and /θ/ in American English, and (2) both fricatives are voiceless, which precludes the possibility of voicing enhancement. For example, if on a particular trial, listeners hear the word ‘mass’ and see both MASS and MATH on the screen, the presence of glottalization should not affect their identification of the target given the alternative on the screen; although the presence of glottalization might be unexpected on ‘mass’, it is not expected on the competitor MOTH either, and thus should not bias the listener towards the competitor, resulting in an overall null effect of glottalization.

The Place of Articulation (POA) group

The second “Place of Articulation (POA)” group tested whether the presence of glottalization would result in differences in eye-tracking behavior for words ending in coda /t/ vs. /p/. For this condition we used words differing only in their final consonant, which could be either a voiceless alveolar stop /t/ or a voiceless labial stop /p/, e.g. ‘rap’ vs. ‘rat.’ In American English, coda /t/ usually appears as glottalized $[\widehat{t}]$ or even $[\text{ʔ}]$, whereas coda /p/

glottalizes much less frequently (Pierrehumbert, 1994; Huffman, 2005; Seyfarth & Garellek, 2015). Thus, for this group we have two main predictions: if glottalization is associated with /t/ only, then its presence may be facilitative or neutral for identifying words with /t/ (vs. /p/), but not inhibitory. Further, its presence on a word with coda /p/ should be inhibitory when paired with a competitor ending in /t/. On the other hand, if listeners associate glottalization with both /t/ and /p/ (consistent with the fact that glottalization can occur with both), then we would expect no inhibitory effect of glottalization on identifying words with /p/. The complete wordlist for the POA group appears in Table 2 (the pairs do not differ significantly in frequency, as assessed by a paired t-test).

Table 2. *Target words used in the POA condition. Frequencies come from the SUBTLEXus corpus (Brysbaert & New, 2009).*

Word 1	Log frequency	Word 2	Log frequency
shop	3.4365	shot	4.0645
cop	3.6429	cot	2.0128
rap	2.8235	rat	3.2212
pop	3.5368	pot	3.0607

The Voicing groups

The third and fourth “Voicing” groups tested whether the presence of glottalization resulted in differences in visual fixations for words ending in voiceless vs. voiced coda stops. Thus for these conditions we used words differing only in their final consonant: in Group 4a, the final stop could be either a voiceless alveolar stop /t/ or a voiced /d/, e.g. ‘bat’ vs. ‘bad’; in Group 4b, the final stop could be either a voiceless labial stop /p/ or a voiced /b/, e.g. ‘tap’ vs. ‘tab.’ Recall that, in American English, only the voiceless codas glottalize, though coda

/p/ glottalizes much less frequently than /t/ (Pierrehumbert, 1994; Huffman, 2005; Seyfarth & Garellek, 2015). However, some researchers (Pierrehumbert, 1994; Keyser & Stevens, 2006) have claimed that glottalization occurs to enhance voicelessness, which should in principle apply for both /t/ and /p/. Thus, for this group we have two main predictions: if glottalization enhances voicelessness, then its presence should be facilitative for words with coda /t, p/ and inhibitory for /d, b/. Alternatively, if glottalization is associated with /t/ more than /p/, then its presence may be facilitative words with /t/, weakly facilitative or neutral for identifying words with /p/, and inhibitory for words with coda /d, b/. The complete wordlists for Groups 4a and 4b are shown in Table 3 (the pairs do not differ significantly in frequency, as assessed by a paired t-test).

Table 3. *Target words used in the Voicing conditions (Coronal and Labial contrasts). Frequencies come from the SUBTLEXus corpus (Brysbaert & New, 2009).*

Voiceless	Log frequency	Voiced	Log frequency
bat	3.0224	bad	4.4441
sat	3.1644	sad	3.5096
mat	2.2529	mad	3.7623
not	5.442	nod	2.3096
tap	2.8768	tab	2.4698
cap	2.9809	cab	3.2617
mop	2.3263	mob	2.8202
nap	2.8102	nab	1.5441

In total then, we had 16 pairs of target words. Correspondingly, we included 16 pairs of filler words, which differed in terms of their initial onset consonant. The reason for using pairs differing only in their onset was to prevent participants from learning that they

should focus only on the final consonant. As with the target word pairs, these words were always monosyllabic and came in non-glottalized/glottalized versions. A complete wordlist (including fillers) can be found in the Appendix.

2.3. Audio stimuli resynthesis

The stimuli were recorded by a phonetically-trained female speaker of Southern Californian English. She uttered each word in a carrier phrase, where the target word appeared in medial position to avoid phrase-final creak. Each word was then resynthesized using the Klatt synthesizer in Praat (Boersma & Weenink, 2015) as follows. First, natural tokens were segmented according to their onsets, vowels, and codas. (For words beginning with approximants /ɹ/ and /w/, the onset was not segmented due to difficulties identifying onset-vowel boundaries, and because such approximants can easily be resynthesized to sound natural.) The natural onsets and codas were extracted. Then, the vowel was extracted using a Hamming window (in our piloting, windowing improved the naturalness of the vowel resynthesis and consonant-vowel cross-splicing, described below). The vowel was then resynthesized using the Klatt synthesizer in Praat. We created two versions of each vowel: one glottalized (i.e., with creaky voice during the vowel), the other non-glottalized. The four parameters used to create both versions are shown in Table 4, and result in a prototypical creaky voice quality (Keating et al., 2015; Garellek, to appear) that is commonly found on vowels before glottalized /t/: constricted-sounding (as indexed by lower spectral tilt), low and irregular in f_0 , and where the creaky voice progresses over the vowel's duration (Garellek & Seyfarth, 2016). The precise values of each parameter, and the time points at which values were changed, were determined in our piloting to maximize the percept of glottalization while minimizing synthetic vowel quality.

The first parameter, Open Phase, changes the open quotient of each glottal pulse (the proportion of the glottal cycle during which the vocal folds are open); higher values of

Open Phase result in a more prominent first harmonic, which is associated with less creaky voice. We chose to manipulate this parameter from the vowel's midpoint rather than onset, because during piloting we found that a low level of Open Phase throughout the vowel resulted in a less natural-sounding glottalized token with too tense a voice quality. Thus, for our "glottalized" vowels, we had this parameter decrease linearly from the midpoint of the vowel to its end. This resulted in an increasingly constricted, creaky quality of the "glottalized" vowel compared to its "non-glottalized" counterpart.

We also manipulated the Spectral Tilt parameter, which raises or lowers the spectral tilt between H1 and the harmonic around 3000 Hz. Glottalized vowels typically have lower spectral tilt, especially in the latter half of the vowel (Garellek & Seyfarth, 2016). Consequently, we had the vowel start off with a Spectral Tilt value of 10 dB and end with a value of 0 dB. This parameter, like the Open Phase one, also resulted in a more constricted creaky voice quality for "glottalized" vowels compared with "non-glottalized" ones.

"Glottalized" tokens also had adjustments to the Flutter parameter, which at higher values produces jitter, or irregular pitch periods. Higher values of the parameter mean that the F0 becomes increasingly irregular, which is common during glottalization (Garellek & Seyfarth, 2016). The "non-glottalized" vowels had no Flutter, resulting in a regular falling F0 contour.

The final parameter manipulated was Double Pulsing, which like Flutter varied for "glottalized" tokens and held at 0 for "non-glottalized" ones. A non-zero value of this parameter creates sub-harmonics, resulting in period-doubling or "diplophonia", another common acoustic and perceptual feature of glottalization and creaky voice in general (Gerratt & Kreiman, 2001; Redi & Shattuck-Hufnagel, 2001; Keating et al., 2015). Higher values indicate greater period doubling, increasing the percept of irregular pitch. We chose to manipulate this parameter from the vowel's midpoint rather than onset, because during piloting we found that a high level of Double Pulsing throughout the vowel resulted in a less

natural-sounding glottalized token. Thus in our “glottalized” vowels, this parameter increased from 0 to 0.5 from the vowel’s midpoint to the end of the vowel.

Table 4. *Values of the Klatt Synthesizer parameters that were manipulated to create non-glottalized and glottalized versions of the stimuli.*

	Non-glottalized vowel	Glottalized vowel
Open Phase (OQ)	0.5 throughout	Initial value: 0.5 Middle value: 0.5 Final value: 0.2
Spectral tilt	10 throughout	Initial value: 10 Final value: 0
Flutter	0 throughout	0.95 throughout
Double pulsing	0 throughout	Initial value: 0 Middle value: 0 Final value: 0.5

Once the non-glottalized and glottalized versions of each vowel were resynthesized, we then spliced the original (non-synthesized) onset and coda back on to the synthetic vowels. All tokens had released codas, though these could differ in terms of intensity across items. If the onset or coda was judged too loud or quiet with respect to the synthetic vowel, we adjusted the consonants’ intensities before concatenating the two versions of the word once more. The vowel durations – which were not altered in the resynthesis – could differ across items, based on the way the speaker uttered each word. As we would expect, vowels tended to be shorter before voiceless stops than before voiced ones, and before coronals than before labials (Peterson & Lehiste, 1960; Luce & Charles-Luce, 1985). Crucially, non-glottalized and glottalized versions of each word always had the same onset and coda, and the same vowel duration. Lastly, all stimuli were scaled for peak intensity. Figure 2 illus-

trates the spectrographic differences between non-glottalized and glottalized versions of the audio stimuli. Sample audio files can be found as supplementary materials.

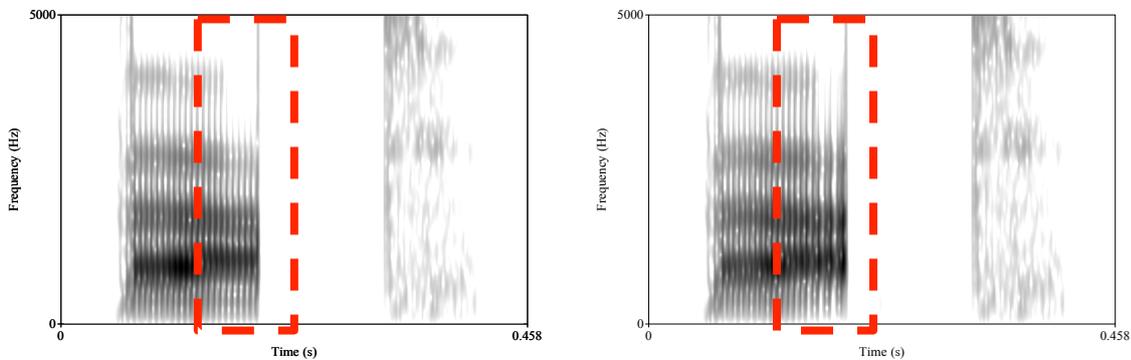


Figure 1. Spectrograms of non-glottalized (left) and glottalized (right) versions of the stimulus 'bat'. The dashed box illustrates where the second half of the vowel, where the differences between non-glottalized and glottalized versions are strongest. For the glottalized version, note the presence of irregular glottal pulses and lower spectral tilt (as evidenced by greater energy in the higher frequencies).

2.4. Visual stimuli

Visual stimuli consisted of printed text of the target words. These were created in Adobe Photoshop and were 88 pixels in size and in Times New Roman font. They were saved as .bmp files and were presented on a 1920 X 1080 ASUS HDMI monitor. A sample visual display is shown in Figure 2.

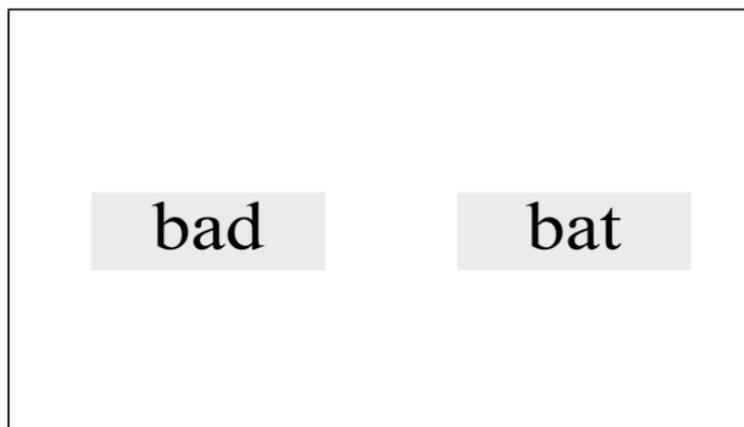


Figure 2. *Sample visual display with 'bad' (left) and 'bat' (right).*

2.5. Procedure

Each listener heard each target test word only once, in either glottalized or non-glottalized conditions. Listeners, however, heard both members of a given stimulus pair across two different blocks, such that a particular visual stimulus pair was not seen in the same block. The order of blocking was also counterbalanced. This yielded four experimental lists. Listeners heard onset fillers in both glottalized and non-glottalized conditions in all counterbalancing groups, split evenly across blocks. The blocking for these stimuli was kept consistent across each experimental list, with trials in each block being fully randomized. There were 96 trials in each experimental list (see Appendix for details).

Participants sat in front of an arm-mounted SR Eyelink 1000 (SR Research, Mississauga, Canada) set to track the left eye at a sampling rate of 250 Hz, at an approximate distance of 550 mm. At the start of each experiment, a 5-point calibration was conducted. They were instructed that they were going to see words on a screen and to listen very carefully to the instructions they were going to hear. At the start of each trial, both words appeared on the screen, one on the left and the other on the right, and they heard the instructions: "Look at the words on the screen". The text remained on the screen for 3 seconds. Then they heard the instructions "Now look at the cross" which coincided with the appearance of a cross in the center of the display. This was done to ensure that when the words were named in the test phase listeners were not already on the relevant target word. Following a 500-ms pause, they then heard: "Now look at the word...[*TARGET*]". The text stayed on the screen for 1500 ms following the offset of the target word. The side that the target image appeared on was counterbalanced across trials, such that they appeared equally on either side of the screen. In order to familiarize participants with the task, listeners

heard six practice trials which consisted entirely of onset fillers (which also appeared later in the experiment itself). Drift correction was conducted after every five trials. In total, each experimental session lasted approximately 25 minutes. The experiment is schematized in Figure 3.

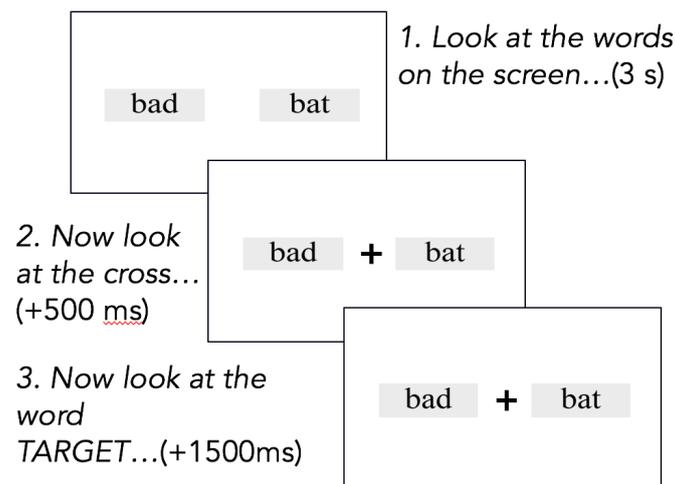


Figure 3. Schematic of the experiment, with sample display showing 'bad' and 'bat'.

2.6. Measures

The eye movements of participants were monitored over the course of each test trial until the trial ended. We were interested in looking at two measures, following Beddor et al. (2013). The first was the latency of the first *correct* fixation to the target word. Latencies were calculated from the onset of the vowel. We chose the vowel onset as the start of the analysis window (or approximant onsets, which were also resynthesized) as we are interested in within-item comparisons (e.g. non-glottalized *bat* vs. glottalized *bat*), where the phonetic differences between the resynthesized stimuli (i.e. presence of creaky voice) is realized from vowel (or approximant) onset to vowel offset. Because words are matched for non-approximant onsets, any disambiguating information would only start during the vowel via formant transitions. Trials in which listeners made errors in looks (i.e. they only looked to

the distractor and never to the target) were excluded from this analysis. We excluded any fixations that occurred in the first 200 ms from the vowel onset. Given the 200 ms it takes to initiate a saccade (Matin et al., 1993; Dahan et al., 2001), any fixation to the target region would not be due to any potentially disambiguating information. We further excluded trials in which the latency of the first correct fixation did not occur within the first 1000ms. Two rectangular interest areas were set 350 X 240 pixels around the printed target words, and were situated in distinct halves of the screen such that there was no overlap between interest areas. The interest areas were also larger than the target words themselves to ensure that fixations around the target word would also be counted.

The second measure was the proportion of fixations to the target image, which indicates how good a target auditory stimulus is at activating the target lexical representation. Thus if an image attracts more looks than a competitor when one auditory stimulus is heard compared to another, we can say that the former is better at facilitating lexical access than the latter. The time window for the analysis started from 200 ms following the onset of the vowel and ended at 1000 ms, where fixations to the target image generally plateaued. This resulted in an 800-ms analysis window. Finally, since the analysis of proportions is problematic to handle in a linear model, target fixation proportion data was first transformed using the empirical logit function, using the function in Barr (2008)¹.

¹ Empirical logit transformation = $\log\left(\frac{Y+0.5}{N-Y+0.5}\right)$, where N is the total number of samples in a given 50-ms frame and Y is the number of samples that fall in a given target interest area (Barr, 2008).

3. Results

3.1. Baseline

To ensure that our synthesized glottalized stimuli were not degraded compared to our synthesized non-glottalized stimuli, we first examine what listeners' responses were to these stimuli in our baseline condition where we do not expect there to be any differences in fixation latencies or proportions based on presence or absence of glottalization. Thus in the baseline conditions, trials consisted of pairs of target words with coda /s/ or /θ/: glottalization is not associated with either of these coda consonants, so we do not expect any facilitatory or inhibitory effect on word recognition in this sub-experiment.

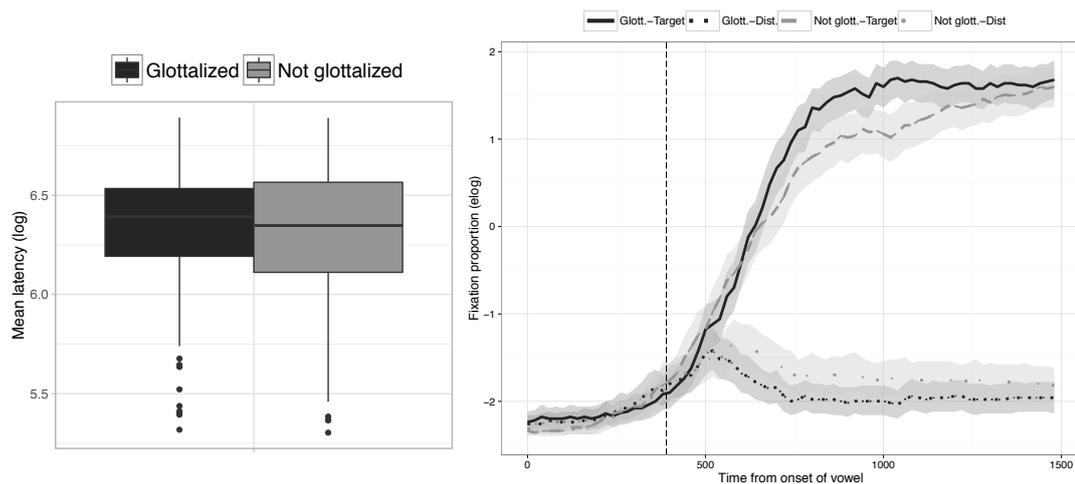


Figure 4. Results for latency (left) and fixation proportion (right) in the Baseline condition, where targets and competitors differed in terms of having coda /s/ vs. /θ/. Darker lines indicate glottalized words; dotted lines indicate looks to distractors. The vertical dashed line indicates the average offset of target words across all experimental groups.

The left panel of Figure 4 shows the log-transformed mean latencies of the first correct fixations on the trials where participants heard trials with coda /s/ or /θ/. Overall, as we predicted, listeners fixated to the target word equally quickly regardless of whether they

had heard a glottalized or non-glottalized token of the target word (Glottalized: Mean = 587, S.D. = 152; Non-glottalized: Mean = 579, S.D. = 169). A linear mixed-effects model (Bates et al., 2015) with Satterthwaite approximations was fit using the *lmerTest()* function (Kuznetsova et al., 2016) in R (R Core Team, 2015). with the log-transformed fixation latencies as the dependent variable and Glottalization (Glottalized vs. Non-glottalized) as the fixed effect. The model was also fit with random intercepts for participant and target word, as well as random slopes for Glottalization status for both participant and word. *P*-values are provided in the model outputs. (For all statistical analyses, the full model structures and results can be found as supplementary materials.) As was expected, the presence or absence of glottalization did not significantly affect the speed at which listeners looked to the target word when they heard a word with coda /s/ or /θ/ [$\beta = -0.07$, SE = 0.09, $t = -0.73$, $p > 0.05$].

Further confirmation of this result comes from looking at the timecourse of fixations to the target word in baseline trials (right panel of Figure 4). As is evident from the fixation behavior, fixations to the target word did not differ across time regardless of the presence or absence of glottalization in the auditory stimuli. A linear-mixed effects model was fit to the logit-transformed target fixation proportion, with Glottalization as a fixed effect as well as random intercepts for participant and word as well as random slopes for Glottalization. Fixations were averaged over the 800-ms analysis window that started from 200 ms following the onset of the vowel and ended 1000 ms after the vowel onset. As with the analysis of fixation latencies, there was no significant effect of glottalization on listeners' fixation proportions for target words with coda /s/ and /θ/ [$\beta = -0.41$, SE = 0.26, $t = -1.62$, $p > 0.05$]. As expected, the results of both the latency and fixation analysis indicate no specific baseline effects of glottalization on the resynthesized words.

3.2. Place of Articulation

In the Place of Articulation condition, listeners saw displays with minimal pairs of words with coda /t/ or /p/. Here we were interested in examining whether listeners were faster at recognizing /t/ and /p/ words when they were glottalized.

Table 5. Mean first correct fixation latencies (in ms) by Glottalization and Place of Articulation (one standard deviation in parentheses).

	Coronal stops	Labial Stops
Glottalized	491 (121)	561 (142)
Non-Glottalized	532 (125)	557 (152)

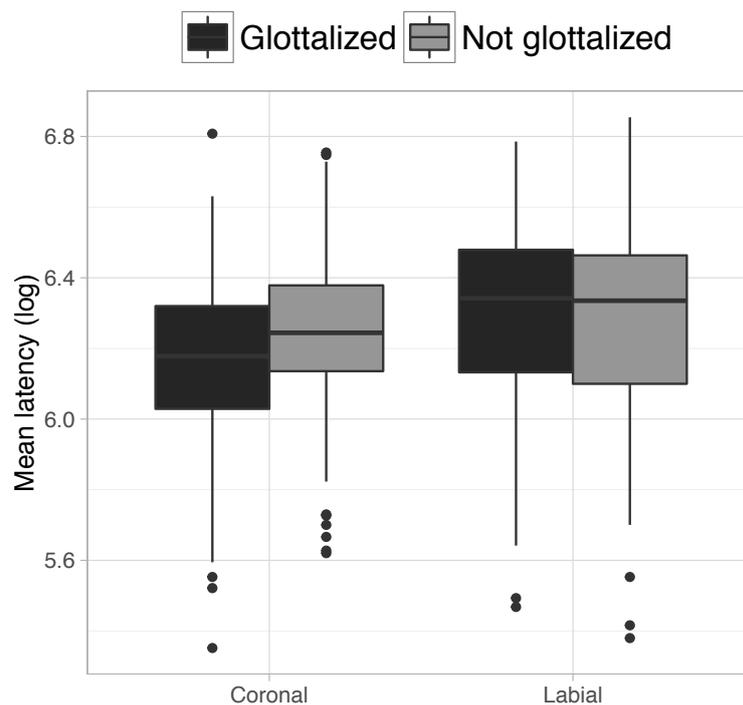


Figure 5. Results for latency in the Place of Articulation condition, where targets and competitors differed in terms of having coronal coda /t/ vs. labial coda /p/.

Table 5 and Figure 5 show the mean latencies of the first correct fixations on the trials where participants heard trials with coda /t/ or /p/. Log latencies were analyzed using a linear-mixed effects model with Glottalization and Place (both contrast-coded), as well as their interaction, as fixed factors. The model also included by-subject and item random intercepts as well as random slopes for Glottalization X Place by subject and a random slope for Glottalization by target word. There was no significant effect of Glottalization [$\beta = 0.13$, $SE = 0.11$, $t = 1.20$, $p > 0.05$] or Place [$\beta = 0.29$, $SE = 0.18$, $t = 1.62$, $p > 0.05$]. Finally, there was a marginal interaction, indicating that listeners were marginally quicker to fixate on the target when they heard a word ending in a glottalized /t/ than one ending in a non-glottalized /t/ [$\beta = -0.36$, $SE = 0.21$, $t = -1.77$, $p = 0.09$].

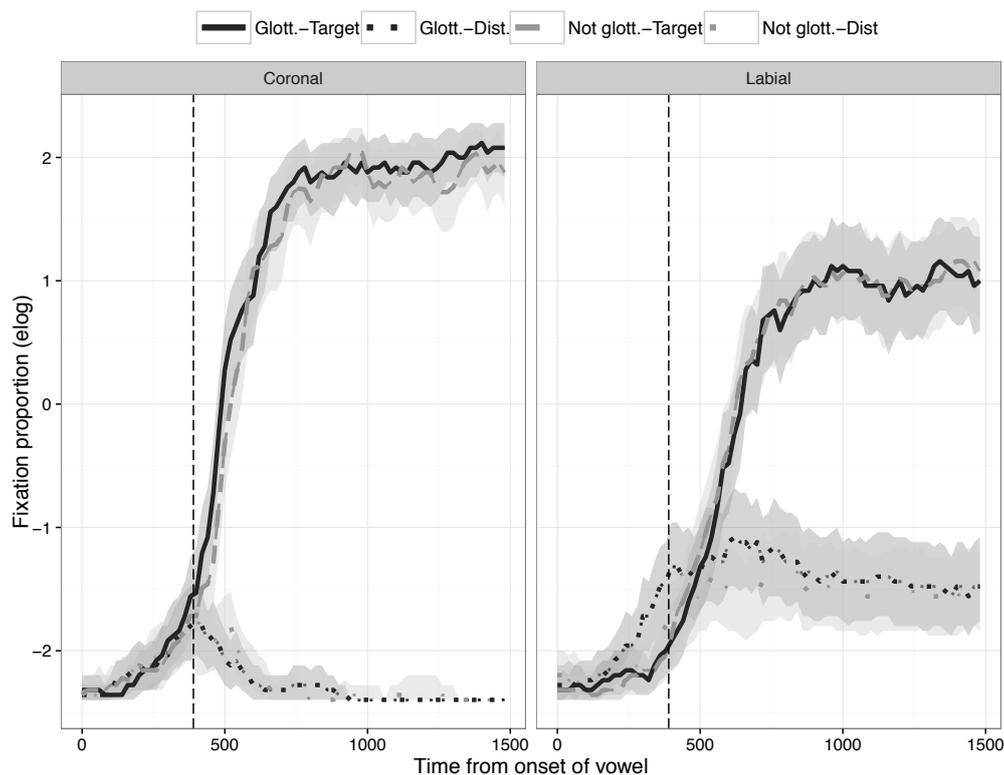


Figure 6. Fixation proportion for coronal /t/ (left) vs. labial /p/ (right). Darker lines indicate glottalized words; dotted lines indicate looks to distractors. The vertical dashed line indicates the average offset of target words across all experimental groups.

The timecourse of fixations to the target and distractor is shown in Figure 6. We analyzed fixation proportions using the same analysis window as above and with a linear-mixed effects model. For this model, we included as fixed effects Glottalization and Place, as well as their interaction. Random intercepts for subject and target word were also included, as well as random slopes for Glottalization and Place by subject and Glottalization by word. This was the maximal model to converge. Results show no significant effect of glottalization [$\beta = 0.04$, $SE = 0.21$, $t = 0.18$, $p > 0.05$] but there was a significant effect of place [$\beta = -1.56$, $SE = 0.56$, $t = -2.79$, $p = 0.03$], confirming the observations that words with coda /p/ were just recognized more poorly than words with coda /t/. The interaction was not significant [$\beta = -0.02$, $SE = 0.42$, $t = -0.04$, $p > 0.05$], indicating that the effect of glottalization did not differ between words ending in /p/ vs. /t/.

Overall then, listeners looked less to the correct target when they heard a word ending in coda /p/ than a word ending in coda /t/, suggesting that listeners expect a word-final /t/ more than a word-final /p/. This may be due to distributional properties of English: in the SUBTLEXus corpus (Brysbaert & New, 2009), there are 2.5 times more words ending in /t/ than /p/. That is, it is possible listeners might *a priori* expect a /t/-final word than /p/-final word when presented with both. One could also argue that this effect is driven by acoustic properties, for instance that coda /p/ is harder to identify than /t/. However, this is unlikely to be due to properties of the /p/-final stimuli independent of the distractor; as we will see in the following section, when the target word ends in /p/ and the competitor ends in /b/, listeners do in fact converge on /p/ responses by the end of the analysis window.

3.3. Voicing

In the Voicing condition, listeners saw displays with minimal pairs of /t/ vs. /d/, as well as /p/ vs. /b/. Here we address the question whether or not glottalization facilitates the recognition of voiceless stops over voiced stops, and further whether this differs by place of articulation.

Table 6. Mean first correct fixation latencies (in ms) by Voicing, Glottalization and Place of Articulation (standard deviations in parentheses).

	Coronal stops		Labial stops	
	Voiceless	Voiced	Voiceless	Voiced
Glottalized	493 (121)	607 (138)	489 (149)	607 (139)
Non-Glottalized	506 (137)	590 (151)	507 (138)	562 (135)

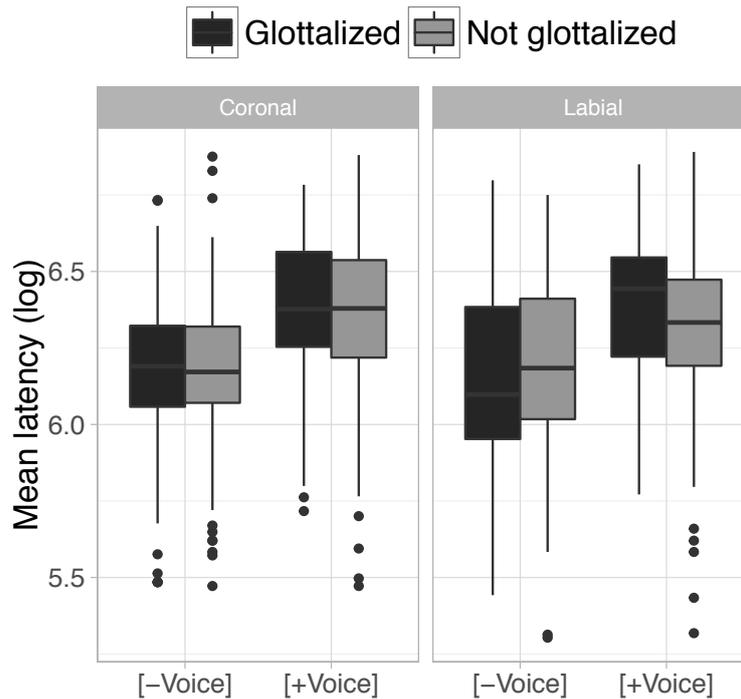


Figure 7. Mean latency (log-transformed) of first correct fixations for coronal /t/ and /d/ (left) and labial /p/ and /b/ (right).

Figure 7 shows the log-transformed mean latencies of first correct fixations in the Voicing condition. Overall listeners were slower at correctly fixating to the target image when the target word ended in a voiced stop (Table 6). A linear mixed-effects was fit to fixation latencies with the following fixed effects: Voicing, Place, Glottalization, as well as their interactions. All factors were contrast-coded. Random intercepts for subject and target word were also included, as well as by-subject random slopes for Glottalization X Place, Glottalization X Voicing, and Voicing X Place, as well as by-word slopes for Glottalization. This was the maximal model to converge. There was a significant effect of Voicing [$\beta = 0.60$, SE = 0.08, $t = 7.60$, $p < 0.001$], with listeners being slower at fixating to the correct target image when the target was a voiced stop, regardless of place of articulation or glottalization. There was also a significant two-way interaction of Glottalization by Voicing [$\beta = -$

0.28, SE = 0.12, $t = -2.30$, $p = 0.03$], with listeners being slower at fixating to the correct image when the target word was voiced and glottalized. No other effects were significant.

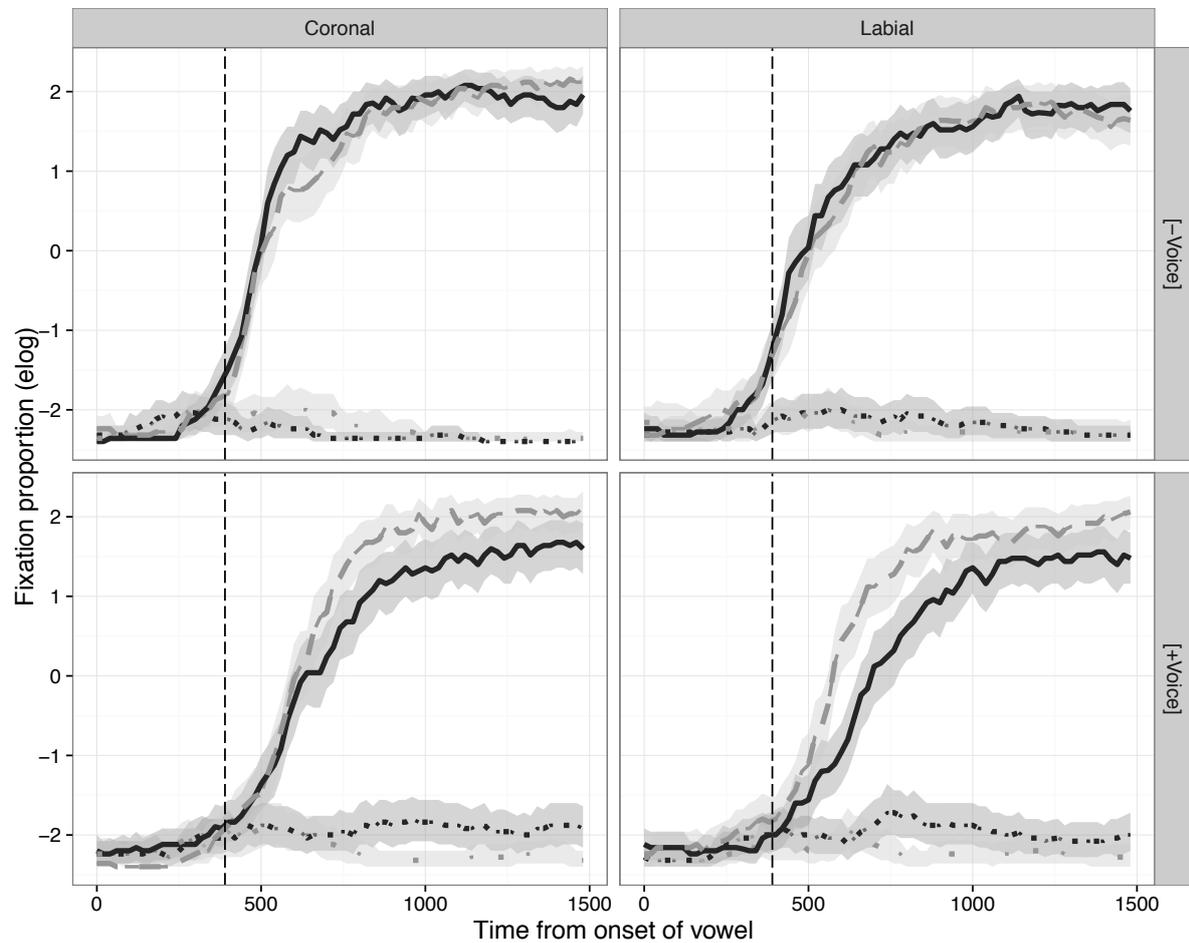


Figure 8. Fixation proportion for coronal /t, d/ (left) vs. labial /p, b/ (right). The top panels show results when the target was /t, p/, the bottom panels when the target was /b, d/. Darker lines indicate glottalized words; dotted lines indicate looks to distractors. The vertical dashed line indicates the average offset of target words across all experimental groups.

The timecourse of fixations to the target and distractor is shown in Figure 8. Logit-transformed fixation proportions were analyzed using a linear-mixed effects model with the following fixed effects: Voicing, Place, Glottalization, as well as their interactions. All factors were contrast coded. The random-effects structure was the same as for fixation latency.

This was the maximal model to successfully converge. Results show significant main effects of Voicing [$\beta = -0.66$, $SE = 0.17$, $t = -3.86$, $p = 0.002$] and Glottalization [$\beta = 0.33$, $SE = 0.13$, $t = 2.61$, $p = 0.01$]. There was also a significant two-way interaction of Voicing by Glottalization [$\beta = 0.87$, $SE = 0.27$, $t = 3.21$, $p = 0.003$], driven by the fact the non-glottalized voiced stops have higher fixation proportions than glottalized voiced stops (but no effect of glottalization is found for voiceless stops, as we discuss in more detail below). No other factors and interactions were significant.

Given the significant interaction between Glottalization and Voicing, we were further interested in examining when fixations to words ending in non-glottalized stops were higher than those to words ending in glottalized ones. (No timecourse analyses were presented earlier for lack of an effect of glottalization.) We used the *bdots* package (Seedorff et al., 2017) in R to ascertain the time window in which fixations to the target word differed as a function of glottalization. Following Oleson et al. (2015) a nonlinear curve is fit to the fixation proportions for each target word² in the glottalized and non-glottalized conditions. In our case, since we did not find a main effect of place, both /t/ and /p/, as well as /d/ and /b/ are collapsed into one analysis per voicing group. Through bootstrapping using the *logistic.boot()* in the *bdots* package function in R, a set of estimates and standard errors is obtained across all the items within a glottalization condition. This then yields a time window in which two curves (e.g., glottalized vs. non-glottalized voiceless stops; glottalized vs. non-glottalized voiced stops) significantly differ from each other.

Our analysis revealed no time windows in which participants looked significantly more at the target word when glottalization was present for words ending in voiceless stops.

² Although Oleson et al. (2015) present a by-subject analysis, we chose here to conduct a by-item analysis for reasons of statistical power. Given that each subject contributes few data points to each experimental condition, the *bdots* package fails to adequately model the by-subject data. Thus, our by-items analysis allows us to include more data points per item, allowing the *bdots* package to more successfully fit the fixation curve.

For voiced stops on the other hand, participants looked significantly more at the target word when they heard the non-glottalized production starting from 520 ms following the onset of the vowel and lasting to the end of the trial (adjusted $p = 0.006$, see also figure in supplementary materials). Allowing for the 200 ms time delay to execute a saccade in response to an auditory stimulus, this means that participants are responding to what they heard at about 320 ms following the vowel onset, which is before the average onset of the coda consonant (400 ms), but nonetheless within the second half the vowel where glottalization would be most prominent. This therefore provides further evidence that participants' fixation behavior was due to their perception of the glottalization on the vowel, particularly when the strength of glottalization increased.

4. Discussion

4.1. Summary of results

Using a printed-text variant of the visual world paradigm, we investigated the timecourse of perception of coda glottalization in American English. Specifically, we examined the predictions of two enhancement accounts for why glottalization occurs. First, several researchers (e.g., Stevens & Keyser, 1989; Pierrehumbert, 1994, 1995; Keyser & Stevens, 2006; Stevens & Keyser, 2010; Gordeeva & Scobbie, 2013; Penney et al., 2018) have claimed that glottalization enhances voicelessness; thus, we predicted that glottalization should facilitate recognition of voiceless stops (both /t/ and /p/) when paired with voiced ones. When the target word ended in a voiceless stop, we found that listeners fixated to the target image equally quickly, regardless of the absence or presence of glottalization. That is, glottalization did not facilitate the recognition of words with coda voiceless stops. Interestingly, we found that glottalization inhibited the recognition of words with coda voiced stops, as evidenced by slower target fixation latencies as well as lower target fixation proportions.

Second, glottalization most commonly occurs with /t/ (Pierrehumbert, 1994; Huffman, 2005; Seyfarth & Garellek, 2015), which Stevens & Keyser (2006) explain is due to the fact that coronal articulations are fastest to produce, which means that /t/ is likely to be subject to more overlap with non-coronal articulations and would therefore benefit from glottalization. Thus, we hypothesized that glottalization should facilitate the recognition of words ending in coda /t/ when paired with words ending in coda /p/. Contrary to this prediction, listeners were not any better at recognizing the target /t/-final word when it had glottalization. The same result was found for /p/-final words. Thus, although glottalization most commonly occurs with /t/ and less so with /p/, this distributional asymmetry does not result in faster or better recognition of words ending in /t/, nor does it result in slower or poorer recognition of words ending in /p/.

Overall, we interpret these results as follows: first, given that recognition of words ending in voiced stops is inhibited by glottalization, listeners must (a) be sensitive to its presence in the acoustic signal, and (b) know that glottalization is a cue to word-final voiceless stops but not to word-final voiced ones. Second, because listeners do not use glottalization to facilitate recognition of words ending in voiceless stops, this suggests that glottalization, though perceptible, does not enhance the recognition of words ending in voiceless stops in general, or /t/ specifically. In the remainder of the discussion, we review the implications of these results for the perception of coarticulation and explanatory theories of coda glottalization.

4.2. Glottalization as coarticulation

As we discussed earlier, listeners are sensitive both to the *appropriateness* of coarticulatory cues for an upcoming sound (Dahan et al., 2001; Dahan & Tanenhaus, 2004; Gow & McMurray, 2007), as well as to the specific timing of when coarticulatory cues become available in the speech signal (Beddor et al., 2013; Salverda et al., 2014). For instance, Bed-

dor et al. (2013) showed that listeners' fixations converged on a CVNC word (e.g. *bent*) more quickly when nasalization started early in the vowel than when it started late in the vowel.

In our study, creaky voice can also be construed as a coarticulatory cue to stop identity: if listeners know, based on distributional evidence, that the presence of creaky voice is a cue to an upcoming voiceless coda stop (especially /t/), then they should be able to use the non-modal phonation as an anticipatory cue to recover an upcoming glottalized stop. In other words, they should look faster at words ending in /t/ than /p/ when glottalized, and faster at words ending in /t/ or /p/ than /d/ or /b/. However, we find little support for this in our study: when they hear a glottalized word ending in a voiced stop, listeners fixate less on the target image but when they hear a glottalized word ending in a voiceless stop, they do not use glottalization to fixate sooner on the target.

Thus, our results seem to be at odds with those in previous work, particularly those in Beddor et al. (2013). Firstly, we do not believe that the absence of an effect of glottalization in certain conditions can be attributable to a lack of statistical power. While Beddor et al.'s study had many more trials per participant (360 vs. 96), the current study tested nearly three times the number of participants (23 vs. 60). Moreover, the number of stimuli per condition was comparable across both studies. Instead, we believe that the absence of an effect of glottalization stems from intrinsic differences between cues to nasalization vs. glottalization. In our study, "glottalized" stimuli became progressively creakier as the vowel unfolded; two of the four glottalization parameters were used only from the vowel midpoint onward. We created glottalized stimuli in this way because of ecological validity and for methodological reasons: vowels before glottalized stops are creakiest nearest to the stop formation – unlike vowels in phrasal creak, where the creaky voice quality remains more stable (Garellek & Seyfarth, 2016). Thus, we created stimuli that sounded typically glottal-

ized (i.e., typical of vowels before glottalized stops), and which were unlikely to sound like stimuli produced with an overall creaky voice setting.

On the other hand, nasalized vowels in Beddor et al. (2013) varied according to where nasalization began (early vs. late onset), which the authors manipulated by varying the point at which oral vowels were cross-spliced with nasalized ones. Although it is unclear how the degree of nasalization (i.e., its spatial distinction with respect to oral vowels) varied over time, the results of their study imply that acoustic cues to nasalization must be perceptible as soon as they begin. This may be due to differences in listener sensitivity to glottalization vs. nasalization cues; for instance, listeners may be worse at perceiving changes in spectral tilt and f_0 regularity (i.e., the cues to glottalization) than they are at perceiving changes associated with nasalization (including formant frequencies and bandwidths, as well as spectral slope: Beddor, 1993; Styler, 2017). An alternative explanation is that, regardless of how well listeners perceive certain acoustic parameters, the cues to nasalization are more closely associated with the linguistic nasal gesture than the acoustic cues to glottalization with stop identity. This is likely the case, given that cues to nasalization relate directly back to nasal gestures (nasal voice is a cue to the presence of a nasal gesture), whereas cues to glottalization relate only indirectly back to a coronal or labial gesture (creaky voice is a cue to the presence of a glottal constriction gesture, which in turn cues the presence of a coda voiceless stop). Moreover, glottal constriction gestures in English can be attributed to multiple linguistic phenomena, including coda stop glottalization, glottalization of word-initial vowels, hiatus environments, and phrasal creak (Dilley et al., 1996; Garellek, 2013; Davidson & Erker, 2014; Garellek, 2015); in the case of coda stop glottalization, glottal constriction is also optional, unlike vowel nasalization before nasal consonants, which is nearly categorically present (though variable in degree; Zellou 2017). So differences in the online perception of nasal coarticulation (Beddor et al., 2013) vs. glottal coarticulation in our study might stem from the fact that perceiving glottal constriction is only

indirectly tied to retrieving coda stop gestures, that glottal constriction is optional, and that it may also be attributed to other linguistic sources.

4.3. Glottalization as phonetic enhancement and phonological variation

A central goal of this study was to examine the validity of a critical assumption of listener-based enhancement theories for glottalization: namely, that listeners should be able to perceive and utilize enhancing cues. Using a paradigm designed to probe listeners' online perception of the incoming signal, we found no evidence in support of this assumption. On the other hand, our findings support Huffman (2005), who argued that enhancement alone could not account for all instances of coda glottalization. Huffman instead posited that coda glottalization might be an optional gesture associated with /t/ and /p/ in coda position. Although listener-based enhancement accounts of glottalization cannot account for our results here, it still is possible that glottalization is used to enhance the articulation of coda stops /t/ and /p/. In other words, speakers may produce glottalization to facilitate reaching an articulatory target (namely voicelessness), but that glottalization gesture itself does not necessarily aid the listener. Note that, if this is the case, it is still unclear why coda /k/ would not also benefit from glottalization as a voicelessness enhancement strategy. Keyser & Stevens (2006: p.43) suggest that the tongue surface stiffens during coda /p/ and /k/, which accelerates the intraoral pressure buildup and thus facilitates voicelessness, obviating the need for glottalization. Of course, that in turn does not explain why coda /p/ is sometimes glottalized, though (following the logic in Keyser & Stevens, 2006; Stevens & Keyser, 2010) perhaps it is subject to more articulatory overlap than /k/. The hypothetical need for enhancing coda /t/ in English is all the more puzzling, given that it has low informativity and is generally subject to the most patterns of lenition (Cohen Priva, 2017).

Nonetheless, the possibility that glottalization is an articulatory enhancement strategy could be viewed as similar to the role of creaky voice on the low-falling tone in White

Hmong; though speakers reliably produce this tone with creaky voice (Esposito, 2012; Garellek, 2012), listeners ignore the creaky voice during word identification (Garellek et al., 2013). In that study, the authors hypothesized that creaky voice might be a means of achieving a low pitch, even if listeners do not use it independently of f_0 . Similarly, in this study we find that glottalization is not used by listeners to perceive voiceless stops, though coda stops (especially /t/) are usually glottalized in spontaneous speech (Huffman, 2005; Seyfarth & Garellek, 2015). Thus, glottalization might enhance the articulation of voicelessness and/or /t/, but this additional gesture is not being used by listeners as an enhancing cue.

Although we did not find evidence in support of listener-based enhancement as an explanation for the occurrence of coda stop glottalization, our results do in fact accord with studies of recognition of phonological variants. Firstly, we found that words with glottalized coda /t/ are recognized as well as those with canonical non-glottalized codas, which replicates the findings of Sumner & Samuel (2005) that both canonical non-glottalized coda [t] as well as glottalized coda [\widehat{t}] are equally effective in priming a given target word. Our current study both replicates this main finding using a different paradigm, and also extends the findings to the recognition of glottalized [\widehat{p}] variants, which Sumner & Samuel (2005) did not examine. Unlike glottalized [\widehat{t}], glottalized [\widehat{p}] is not the most frequent variant of /p/ in coda position (Huffman, 2005). Interestingly, we found that [\widehat{p}] is just as effective at facilitating lexical access as its canonical [p] counterpart. This runs contrary to an account of variant recognition that is driven purely by frequency of occurrence of variants in contexts (e.g. exemplar models: Goldinger, 1996). Sumner & Samuel (2005) also found that there was no advantage for [\widehat{t}] in terms of lexical access, despite it being more frequent in coda contexts. In our study, we similarly did not find any facilitation of word recognition when listeners heard glottalized [\widehat{t}] compared to non-glottalized [t]. These results are consistent with the robust finding that canonical forms are privileged, despite their actual pro-

duction frequency across contexts (McLennan et al., 2003, 2005; Ranbom & Connine, 2007; Pitt, 2009; Pitt et al., 2011; Ranbom et al., 2009; Tucker, 2011).

Finally, our findings provide evidence that glottalization is not associated with voiced stops, as evidenced by slower and poorer recognition of target words ending in coda $[\text{ʔd}]$ and $[\text{ʔb}]$. That is, glottalized voiced stops are less appropriate variants of /d/ and /b/. It is an open question, however, how much of a perceptual mismatch glottalized voiced stops are. Previous findings from word recognition studies have shown that the degree to which word recognition is impeded in both adult and child listeners is dependent on the degree of mismatch between an auditory signal and the lexical representation of a word (e.g. Andruski et al., 1994; McMurray et al., 2002; Gaskell, 2003; White & Morgan, 2008), with smaller mismatches resulting in a smaller decrement in performance. Given that glottalization is not phonemic in English, it is possible that glottalized voiced stops are a smaller perceptual mismatch for stored representations of words with coda voiced stops than whole phoneme substitutions (e.g. “mog” for “mob”). Future work will aim to address this possibility.

4.4. Conclusion and future directions

Using a printed text version of the visual world paradigm, we investigated the perceptual consequences of coda glottalization in American English. Previous researchers have claimed that glottalized coda stops arise from the speaker’s need to enhance voicelessness (or /t/ in particular) in environments where these stops are likely to become voiced or to overlap with other sounds (Pierrehumbert, 1994, 1995; Keyser & Stevens, 2006; Stevens & Keyser, 2010). Listeners treat glottalized variants of /p/ and /t/ as equally good instances of these sounds, and they treat glottalized variants of /b/ and /d/ either as unexpected variants or as mispronunciations. However, listeners do not use glottalization to fixate more on /t/ or /p/, which implies that glottalization is not used to enhance the acoustic attributes of these

sounds and consequently facilitate word recognition. More generally, these results show how the visual world eye-tracking paradigm can be used to examine assumptions of explanatory models of phonological variation that predict facilitative or inhibitory effects on word recognition.

Acknowledgements

We are grateful to Associate Editor Holger Mitterer and two anonymous reviewers for their helpful feedback, and to Kip Wilson for editorial assistance. We would also like to thank Amanda Ritchart, who recorded the stimuli, Scott Ragland and Stefan Wooding for their help with the stimuli resynthesis, as well as members of the UCLA Phonetics Lab and Sarah Creel for comments and discussion. An earlier version of this project was presented at the Fall 2016 meeting of the Acoustical Society of America.

References

- Allopenna, P. D., 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*, 419–439.
- Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*, 163–187.
- Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.
- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1-48.
- Beddor, P. S. (1993). The perception of nasal vowels. In M. K. Huffman, & R. A. Krakow (Eds.), *Nasals, Nasalization, and the Velum* (pp. 171–196). Academic Press.
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *Journal of the Acoustical Society of America*, *133*, 2350–2366.
- Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.04. Retrieved January 5, 2015 from <http://www.praat.org/>.

- Brouwer, S., Mitterer, H., & Huettig, F. (2012). Can hearing puter activate pupil? Phonological competition and the processing of reduced spoken words in spontaneous conversations. *The Quarterly Journal of Experimental Psychology*, *65*, 2193–2220.
- Brysbaert, M., & New, B. (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*, 977–990.
- Cohen Priva, U. (2017). Informativity and the actuation of lenition. *Language*, *93*, 569–597.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534.
- Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 498–513.
- Davidson, L., & Erker, D. (2014). Hiatus resolution in American English: the case against glide insertion. *Language*, *90*, 482–514.
- Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*, 423–444.
- Esling, J. H., Fraser, K. E., & Harris, J. G. (2005). Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). *Journal of Phonetics*, *33*, 383–410.
- Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong phonation. *Journal of Phonetics*, *40*, 466–476.
- Foulkes, P., & Docherty, G. (2007). Phonological variation in England. In D. Britain (Ed.), *Language in the British Isles* (pp. 52–74). Cambridge: Cambridge University Press.
- Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics*, *40*, 152–161.
- Garellek, M. (2013). *Production and perception of glottal stops*. Ph.D. thesis UCLA.
- Garellek, M. (2015). Perception of glottalization and phrase-final creak. *Journal of the Acoustical Society of America*, *137*, 822–831.
- Garellek, M. (to appear). The phonetics of voice. In W. Katz & P. Assmann (Eds.), *Handbook of Phonetics*. New York: Routledge.
- Garellek, M., Keating, P., Esposito, C. M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *Journal of the Acoustical Society of America*, *133*, 1078–1089.
- Garellek, M., & Seyfarth, S. (2016). Acoustic differences between English /t/ glottalization and phrasal creak. In *Proceedings of Interspeech 2016* (pp. 1054–1058). San Fran-

cisco.

- Gaskell, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, *31*, 447–463.
- 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.
- Gordeeva, O. B., & Scobbie, J. M. (2013). A phonetically versatile contrast: Pulmonic and glottalic voicelessness in Scottish English obstruents and voice quality. *Journal of the International Phonetic Association*, *43*, 249–271.
- Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. *Papers in laboratory phonology*, *9*, 173–200.
- Huffman, M. K. (2005). Segmental and prosodic effects on coda glottalization. *Journal of Phonetics*, *33*, 335–362.
- Jongman, A., Blumstein, S. E., & Lahiri, A. (1985). Acoustic properties for dental and alveolar stop consonants: A cross-language study. *Journal of Phonetics*, *13*, 235–251.
- Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.
- Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language*, *82*, 33–63.
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-30.
<https://CRAN.R-project.org/package=lmerTest>
- Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America*, *78*, 1949–1957.
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*, 372–380.
- McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *29*, 539–553.
- McLennan, C. T., Luce, P. A. & Charles-Luce, J. (2005). Representation of lexical form: Evidence from studies of sublexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1308–1314.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*, B33–B42.
- McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, *60*, 661–

671.

- Mitterer, H. (2011). The mental lexicon is fully specified: evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 496–513.
- Oleson, J. J., Cavanaugh, J. E., McMurray, B. & Brown, G. (2015). Detecting time-specific differences between temporal nonlinear curves: Analyzing data from the visual world paradigm. *Statistical Methods in Medical Research*, 0, 1-22.
- Penney, J., Cox, F., Miles, K., & Palethorpe, S. (2018). Glottalisation as a cue to coda consonant voicing in Australian English. *Journal of Phonetics*, 66, 161–184.
- Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32, 693–703.
- Pierrehumbert, J. (1994). Knowledge of variation. In *Papers from the parasession on variation, 30th meeting of the Chicago Linguistic Society* (pp. 232–256). Chicago: Chicago Linguistic Society.
- Pierrehumbert, J. (1995). Prosodic effects on glottal allophones. In O. Fujimura, & M. Hirano (Eds.), *Vocal fold physiology: voice quality control* (pp. 39–60). San Diego: Singular Publishing Group.
- 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.
- Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech (2nd release)*. Department of Psychology, Ohio State University Columbus, OH.
- Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, 39, 304–311.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57, 273–298.
- Ranbom, L. J., & Connine, C. M. (2011). Silent letters are activated in spoken word recognition. *Language and Cognitive Processes*, 26, 236–261.
- Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29, 407–429.
- Roach, P. J. (1973). Glottalization of English /p/, /t/, /k/ and /tS/ – a re-examination. *Journal of the International Phonetic Association*, 3, 10–21.
- Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken- word recognition. *Journal of Memory and Language*, 71, 145–163.

- Seedorff, M., Oleson, J. J., Brown, G., Cavanaugh, J. & McMurray, B. (2017). *bdots*: Bootstrapped differences of time series. R package version 0.1.13.
<https://CRAN.R-project.org/package=bdots>.
- Seyfarth, S., & Garellek, M. (2015). Coda glottalization in American English. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.
- Stevens, K. N., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language*, 65, 81–106.
- Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, 38, 10–19.
- Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *Journal of the Acoustical Society of America*, 142, 2469–2482.
- Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52, 322–338.
- Tucker, B. V. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics*, 39, 312–318.
- Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, 22, 145–166.
- White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, 59, 114–132.
- Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, 61, 13–29.

Appendix A. Experimental stimuli list: test items

List A1				
Group	Target	Competitor	Glottalization	Block
Baseline	mass	math	Non-glot.	1
Baseline	moth	moss	Glott.	1
Baseline	worse	worth	Glott.	1
Baseline	path	pass	Non-glot.	1
Baseline	math	mass	Glott.	2
Baseline	moss	moth	Non-glot.	2
Baseline	worth	worse	Non-glot.	2
Baseline	pass	path	Glott.	2
Place	shop	shot	Non-glot.	1

Place	cot	cop	Glott.	1
Place	rap	rat	Glott.	1
Place	pot	pop	Non-glott.	1
Place	shot	shop	Glott.	2
Place	cop	cot	Non-glott.	2
Place	rat	rap	Non-glott.	2
Place	pop	pot	Glott.	2
Voicing	bat	bad	Non-glott.	1
Voicing	sad	sat	Glott.	1
Voicing	mat	mad	Glott.	1
Voicing	nod	not	Non-glott.	1
Voicing	tap	tab	Non-glott.	1
Voicing	cab	cap	Glott.	1
Voicing	nap	nab	Glott.	1
Voicing	mob	mop	Non-glott.	1
Voicing	bad	bat	Glott.	2
Voicing	sat	sad	Non-glott.	2
Voicing	mad	mat	Non-glott.	2
Voicing	not	nod	Glott.	2
Voicing	tab	tap	Glott.	2
Voicing	cap	cab	Non-glott.	2
Voicing	nab	nap	Non-glott.	2
Voicing	mop	mob	Glott.	2

List A2

Group	Target	Competitor	Glottalization	Block
Baseline	mass	math	Non-glott.	2
Baseline	moth	moss	Glott.	2
Baseline	worse	worth	Glott.	2
Baseline	path	pass	Non-glott.	2
Baseline	math	mass	Glott.	1
Baseline	moss	moth	Non-glott.	1
Baseline	worth	worse	Non-glott.	1
Baseline	pass	path	Glott.	1
Place	shop	shot	Non-glott.	2
Place	cot	cop	Glott.	2
Place	rap	rat	Glott.	2

Place	pot	pop	Non-glot.	2
Place	shot	shop	Glott.	1
Place	cop	cot	Non-glot.	1
Place	rat	rap	Non-glot.	1
Place	pop	pot	Glott.	1
Voicing	bat	bad	Non-glot.	2
Voicing	sad	sat	Glott.	2
Voicing	mat	mad	Glott.	2
Voicing	nod	not	Non-glot.	2
Voicing	tap	tab	Non-glot.	2
Voicing	cab	cap	Glott.	2
Voicing	nap	nab	Glott.	2
Voicing	mob	mop	Non-glot.	2
Voicing	bad	bat	Glott.	1
Voicing	sat	sad	Non-glot.	1
Voicing	mad	mat	Non-glot.	1
Voicing	not	nod	Glott.	1
Voicing	tab	tap	Glott.	1
Voicing	cap	cab	Non-glot.	1
Voicing	nab	nap	Non-glot.	1
Voicing	mop	mob	Glott.	1

List B1

Group	Target	Competitor	Glottalization	Block
Baseline	mass	math	Glott.	1
Baseline	moth	moss	Non-glot.	1
Baseline	worse	worth	Non-glot.	1
Baseline	path	pass	Glott.	1
Baseline	math	mass	Non-glot.	2
Baseline	moss	moth	Glott.	2
Baseline	worth	worse	Glott.	2
Baseline	pass	path	Non-glot.	2
Place	shop	shot	Glott.	1
Place	cot	cop	Non-glot.	1
Place	rap	rat	Non-glot.	1
Place	pot	pop	Glott.	1
Place	shot	shop	Non-glot.	2

Place	cop	cot	Glott.	2
Place	rat	rap	Glott.	2
Place	pop	pot	Non-glott.	2
Voicing	bat	bad	Glott.	1
Voicing	sad	sat	Non-glott.	1
Voicing	mat	mad	Non-glott.	1
Voicing	nod	not	Glott.	1
Voicing	tap	tab	Glott.	1
Voicing	cab	cap	Non-glott.	1
Voicing	nap	nab	Non-glott.	1
Voicing	mob	mop	Glott.	1
Voicing	bad	bat	Non-glott.	2
Voicing	sat	sad	Glott.	2
Voicing	mad	mat	Glott.	2
Voicing	not	nod	Non-glott.	2
Voicing	tab	tap	Non-glott.	2
Voicing	cap	cab	Glott.	2
Voicing	nab	nap	Glott.	2
Voicing	mop	mob	Non-glott.	2

List B2

Group	Target	Competitor	Glottalization	Block
Baseline	mass	math	Glott.	2
Baseline	moth	moss	Non-glott.	2
Baseline	worse	worth	Non-glott.	2
Baseline	path	pass	Glott.	2
Baseline	math	mass	Non-glott.	1
Baseline	moss	moth	Glott.	1
Baseline	worth	worse	Glott.	1
Baseline	pass	path	Non-glott.	1
Place	shop	shot	Glott.	2
Place	cot	cop	Non-glott.	2
Place	rap	rat	Non-glott.	2
Place	pot	pop	Glott.	2
Place	shot	shop	Non-glott.	1
Place	cop	cot	Glott.	1
Place	rat	rap	Glott.	1

Place	pop	pot	Non-glot.	1
Voicing	bat	bad	Glott.	2
Voicing	sad	sat	Non-glot.	2
Voicing	mat	mad	Non-glot.	2
Voicing	nod	not	Glott.	2
Voicing	tap	tab	Glott.	2
Voicing	cab	cap	Non-glot.	2
Voicing	nap	nab	Non-glot.	2
Voicing	mob	mop	Glott.	2
Voicing	bad	bat	Non-glot.	1
Voicing	sat	sad	Glott.	1
Voicing	mad	mat	Glott.	1
Voicing	not	nod	Non-glot.	1
Voicing	tab	tap	Non-glot.	1
Voicing	cap	cab	Glott.	1
Voicing	nab	nap	Glott.	1
Voicing	mop	mob	Non-glot.	1

Appendix B. Experimental stimuli list: onset filler items

All onset fillers in both glottalization and non-glottalized forms were used with the same blocking order across all four experimental lists.

All Experimental Lists			
Target	Competitor	Glottalization	Blocking
note	dote	Glott.	1
neat	seat	Non-glot.	1
seed	deed	Glott.	1
food	mood	Non-glot.	1
night	light	Glott.	1
moat	boat	Non-glot.	1
duck	tuck	Glott.	1
ball	wall	Non-glot.	1
bore	wore	Glott.	1
bear	wear	Non-glot.	1
call	gall	Glott.	1
tear	dear	Non-glot.	1

make	wake	Glott.	1
day	say	Non-glott.	1
bull	pull	Glott.	1
love	dove	Non-glott.	1
dote	note	Glott.	1
seat	neat	Non-glott.	1
deed	seed	Glott.	1
mood	food	Non-glott.	1
light	night	Glott.	1
boat	moat	Non-glott.	1
tuck	duck	Glott.	1
wall	ball	Non-glott.	1
wore	bore	Glott.	1
wear	bear	Non-glott.	1
gall	call	Glott.	1
dear	tear	Non-glott.	1
wake	make	Glott.	1
say	day	Non-glott.	1
pull	bull	Glott.	1
dove	love	Non-glott.	1
note	dote	Non-glott.	2
neat	seat	Glott.	2
seed	deed	Non-glott.	2
food	mood	Glott.	2
night	light	Non-glott.	2
moat	boat	Glott.	2
duck	tuck	Non-glott.	2
ball	wall	Glott.	2
bore	wore	Non-glott.	2
bear	wear	Glott.	2
call	gall	Non-glott.	2
tear	dear	Glott.	2
make	wake	Non-glott.	2
day	say	Glott.	2
bull	pull	Non-glott.	2
love	dove	Glott.	2
dote	note	Non-glott.	2

seat	neat	Glott.	2
deed	seed	Non-glott.	2
mood	food	Glott.	2
light	night	Non-glott.	2
boat	moat	Glott.	2
tuck	duck	Non-glott.	2
wall	ball	Glott.	2
wore	bore	Non-glott.	2
wear	bear	Glott.	2
gall	call	Non-glott.	2
dear	tear	Glott.	2
wake	make	Non-glott.	2
say	day	Glott.	2
pull	bull	Non-glott.	2
dove	love	Glott.	2