

## RESEARCH

# Auditory disruption improves word segmentation: A functional basis for lenition phenomena

Jonah Katz<sup>1</sup> and Melinda Fricke<sup>2</sup><sup>1</sup> West Virginia University, US<sup>2</sup> University of Pittsburgh, USCorresponding author: Jonah Katz ([katzlinguist@gmail.com](mailto:katzlinguist@gmail.com))

This paper presents evidence that *spirantization*, a cross-linguistically common lenition process, affects English listeners' ease of segmenting novel "words" in an artificial language. The cross-linguistically common spirantization pattern of initial stops and medial continuants (e.g. [gʊβa]) results in improved word segmentation compared to the inverse "anti-lenition" pattern of initial continuants and medial stops (e.g. [yuba]). The study also tests the effect of obstruent voicing, another common lenition pattern, but finds no significant differences in segmentation performance. There are several points of broader interest in these studies. Most of the phonetic factors influencing word segmentation in past studies have been language-specific and/or prosodic in nature: stress, intonation, final lengthening, etc. Spirantization, while often prosodically conditioned, is different from all of these patterns in that it concerns a segmental alternation. Moreover, the effects reported here are for speakers of a language, American English, that only sporadically displays spirantization, and not in the phonological contexts used in the experiment. This suggests that the results may reflect more general properties of speech perception and word boundary detection, rather than a perceptual processing strategy transferred directly from English. As such, the studies offer partial support for theories of lenition rooted in notions of perceptual-acoustic continuity and disruption.

**Keywords:** lenition; word segmentation; spirantization; statistical learning; phonetics; phonology

## 1 Introduction

The studies presented here investigate whether two cross-linguistically common lenition patterns have a facilitative effect on word segmentation for speakers of a language that does not display robust versions of these patterns. The term *lenition* is generally used to refer to phonological patterns characterized by relatively less sonorous segments occurring in word-initial position, and relatively more sonorous segments occurring word-medially. The experiments in this paper examine two distinct instantiations of lenition: intervocalic *spirantization*, characterized by word-initial voiced stops and word-medial voiced continuants, and intervocalic *voicing*, characterized by word-initial voiceless obstruents and word-medial voiced obstruents. We employ a well-studied paradigm from the language acquisition literature, the word segmentation paradigm, to ask whether there is any reason to believe that spirantization or voicing provide some benefit for detecting word boundaries. Adult native English listeners, whose language is not generally considered to display robust lenition, are exposed to approximately ten minutes of acoustically continuous, synthesized strings of nonce words that conform either to a lenition pattern or, in another condition, to its inverse, a "language" where the more versus less sonorous segments have swapped position. The question is whether the cross-linguistically common lenition patterns allow these listeners to more readily identify the "words" of the artificial

language, as compared to listeners exposed to “anti-lenition” patterns, which are cross-linguistically unattested. Such a finding would suggest that lenition provides a functional benefit during language learning and/or speech perception, perhaps explaining *why* lenition patterns are so cross-linguistically common and tend to have the form(s) that they do.

The latter point is related to an additional goal, which is to bring experimental evidence to bear on Katz’s (2016) boundary-disruption theory of lenition, a phonological theory developed on the basis of typological patterning that posits an as-yet-untested role for the acoustic-phonetic perception of “disruption”. The boundary-disruption theory of lenition, described in more detail in Section 1.1, is rooted in a typological examination of more versus less common lenition patterns, which ultimately culminates in the hypothesis that certain instantiations of lenition can be unified under the term *continuity lenition*. If continuity lenition patterns are indeed characterized by relative word- or constituent-internal acoustic continuity and relative acoustic-phonetic disruption at constituent boundaries, then patterns ascribed this label based on the typological data should be associated with better performance in an experimental paradigm that tests for listeners’ ability to detect word boundaries.

To preview our main findings: our results indicate that spirantization has a large effect on word boundary detection, while voicing, at least of the type examined here, seems to have little or no effect. We thus find partial but perhaps equivocal support for the boundary-disruption theory, while providing the first experimental evidence (to our knowledge) that at least some lenition patterns may serve a functional role in speech perception.

### **1.1 Lenition and disruption: A phonological theory**

There is a large and heterogeneous set of phonological patterns that are sometimes referred to as *lenition*, and no general agreement on how to define the term (see Honeybone 2008 for an extremely detailed history of the term and the concepts underlying it). Intervocalic spirantization and voicing, however, are included in the class of lenition processes in every definition with which we are familiar. Furthermore, these are of a sub-type referred to variously as *sonority-increasing* (Smith 2008), *vocalic* (Szigetvári 2008), *weak A* (Ségéral & Scheer 2008), and *continuity lenition* (Katz 2016). These lenition processes are defined by their tendency to occur between vowels if they occur anywhere in a language, the phonetic property of becoming more sonorous or louder in lenis realizations (Kingston 2008), and their extremely strong cross-linguistic tendency to produce complementary distribution of phones, very rarely resulting in positional neutralization of phonological contrasts (Gurevich 2003).

Spirantization generally refers to a pattern where stops in one context alternate or are in complementary distribution with continuants in another context. Although the word *spirantization* suggests that the continuant phones should be fricatives, the most common realization for voiced sounds is in fact an approximant without any appreciable noise component (e.g. Romero 1996; Ladd & Scobbie 2003; Kawahara 2006; Chong 2011; Bouavichith & Davidson 2013). We nonetheless retain the term *spirantization* because it is widely known and used in the linguistics community. Gurevich (2003) offers an excellent typological survey of spirantization as part of a broader typology of lenition, based on the earlier work of Kirchner (1998) and Lavoie (2001). Spirantization is the single most frequent lenition pattern attested in all three of those surveys: Gurevich counts 76 languages with spirantization patterns. This survey agrees with earlier work regarding which phonological environments are most likely to condition spirantization: intervocalic position is the most common, and phrase-initial position the least common. Example (1), from the Bantu language Kinande, shows a spirantization pattern similar to the one in the experiment described here.

- (1) *Kinande* spirantization (Katz 2016)
- | #_                       | V_V  |
|--------------------------|--|
| [boloβolo] ‘bit by bit’  | [oβoloβolo] ‘bit by bit’ (variant)             |
| [gereɣere] ‘perfect’     | [omuɣereɣere] ‘perfect person’ (human/class 1) |
| [embwa] ‘dog’ (class 9)  | [akaβwana] ‘young dog’ (diminutive/class 12)   |
| [eŋɣemu] ‘tax’ (class 9) | [eriyemula] ‘to pay a tax’                     |

Note that the segments written as voiced fricatives here are actually approximants. Following convention, we notate them as fricatives for visual and typographical ease. In general, the segments in question are realized as continuants in between two vowels or glides, and as stops phrase-initially (as well as in post-nasal *hardening* contexts, a pattern which is tangential to the current study).

Voicing lenition generally refers to patterns where obstruents are realized as voiceless word-or phrase-initially, and voiced in one or more other contexts. Like spirantization, the most common context for voicing is in between two vowels. And like spirantization, voicing lenition is quite common cross-linguistically: Gurevich reports on 39 languages with voicing lenition. Example (2), from Sanuma (Borgman 1990), shows a voicing pattern similar to the one in the experiment described here.

- (2) *Sanuma* optional voicing (Borgman 1990)
- | #_                     | V_V            |
|------------------------|----------------|
| [telulu] ‘dance’       | [hude] ‘heavy’ |
| [paso] ‘spider monkey’ | [iba] ‘my’     |
| [kahi] ‘mouth’         | [āga] ‘tongue’ |
| [tsinimo] ‘corn’       | [hadza] ‘deer’ |

Stops and the coronal affricate in Sanuma are realized as voiceless in word-initial position and voiced intervocally. This example also illustrates another common feature of voicing lenition: Sanuma voicing is variable, as are many similar processes.

Katz (2016) attempts to account for the positional, phonetic, and allophonic properties associated with these lenition processes using an approach grounded in *boundary-disruption* constraints. The overarching theory, based on earlier work by Keating (2006) and Kingston (2008), is that these particular lenition processes are organized to achieve certain phonological goals with regard to prosodic structure: marking prosodic boundaries with consonants that are more *disruptive* in the context of a stream of high-sonority sounds such as vowels; and marking the lack of prosodic boundaries with consonants that are less disruptive. This boundary-disruption alignment is hypothesized to have the global effect of demarcating the initial edges of prosodic constituents, potentially making them more salient relative to the continuation of a prosodic constituent. The boundary-disruption approach can be thought of as adopting the Gestalt theory of grouping (Wertheimer 1938) for the study of prosody. This general approach to perception and constituency has been influential in the history of cognitive science, and the idea that prosody operates at least in part on domain-general grouping principles is becoming increasingly prevalent in linguistics (e.g. Hunyadi 2006; Jeon & Nolan 2013; Kentner & Féry 2013).

The notion of *disruption* used here is essentially based on auditory similarity to vowels. Approximants are relatively loud, with strong periodic components and at least partial formant structure (Lavoie 2001; Parker 2002), making them relatively similar to vowels, and therefore not very disruptive when interpolated within a sequence of vowels. Voiced stops generally lack clear formant structure but sometimes allow lower-frequency acoustic energy to continue throughout closure (Lisker 1957), and to resume nearly immediately upon release (Lisker & Abramson 1964). Voiceless stops lack acoustic energy almost entirely during closure, are often longer than their voiced counterparts (Lisker 1957;

Luce & Charles-Luce 1985), and generally result in a “lag” for periodic energy to resume after release (referred to as *voice onset time*, Lisker & Abramson 1964). Therefore, both spirantization and voicing could be characterized as patterns that place more disruptive segments at the beginnings of prosodic constituents, and less disruptive ones internal to those constituents.

This reasoning, while grounded in basic and well-attested facts of acoustic phonetics, is based on a theory of fundamentally perceptual disruption that has rarely if ever been tested experimentally. In order to assess the general approach, then, it is important to test whether these particular phonetic disruptions really do make it easier to pick out constituent boundaries in connected speech. In Section 1.3, below, we describe an experimental paradigm that allows us to measure the contribution of various factors to the detection of prosodic boundaries. First, however, we turn to the question of boundary disruption in English. More specifically, we consider whether American English, the native language of the participants in our experiments, displays lenition patterns similar to those exemplified by our experimental stimuli. To the extent that it does, our experimental findings should be interpreted with caution, since prior linguistic experience could result in better word identification performance for artificial languages that are phonologically and/or phonetically similar to English.

## 1.2 Lenition in English

American English features several kinds of lenition. Perhaps the best known is flapping (Haugen 1938), the reduction of /t/ and /d/ to a ballistic tap or flap in certain environments (for the purposes of this discussion, we do not distinguish between tap and flap). Haugen (1938) describes the environment for flapping as following a vowel or sonorant consonant and preceding an unstressed syllabic segment. He describes the flap itself as being shortened, weakened, and more liquid-like than other stops. Subsequent research on flapping has shown that its presence, acoustic nature, and tendency to neutralize the /t-/d/ contrast are quite variable (e.g. Scharf 1960; Sheldon 1973; Umeda 1977; Zue & Laferriere 1979; see De Jong 2011 for a comprehensive overview). The general terms of Haugen’s description, however, have more or less stood the test of time, and the process has generally been considered robust, categorical, and neutralizing *enough* to be described in terms of a phonological rule (e.g. Kahn 1976; Hayes 1995).

Less attention has been paid to another type of leniting process present in American English: spirantization. In roughly the same phonological contexts as those that favor flapping of /t/ and /d/, voiced stops in particular have a tendency to be realized with shortened duration (Turk 1992), incomplete closure (Lavoie 2001), lack of an audible burst (Lavoie 2001; Warner & Tucker 2011; Bouavichith & Davidson 2013), and visible formant structure (Lavoie 2001; Walter 2007; Warner & Tucker 2011; Bouavichith & Davidson 2013). While precise estimates of the prevalence of these phonetically continuant forms differ depending on measurement criteria, experimental task, and place of articulation, non-coronal voiced stops in flapping contexts appear to be realized as approximants somewhere between 40 and 80% of the time. Where the following vowel is stressed, spirantization is rare: Bouavichith & Davidson (2013) and Walter (2007) report rates between 4 and 20%. Lavoie (2001), Walter (2007), and Bouavichith & Davidson (2013) all report higher rates of approximant realization for /g/ than /b/.

Finally, there is evidence that American English non-coronal voiceless stops occasionally undergo *voicing* in flapping contexts. Warner & Tucker (2011) report rates around 5–20%, depending on place of articulation and experimental task. These numbers, however, do not distinguish between stop realizations and continuant realizations. Warner and Tucker also explicitly mention cases with “even low-amplitude closure voicing” (Warner & Tucker

2011: 1609), which leaves open the possibility that the presence of voicing in some of these tokens may not be perceived. Bouavichith (2014) reports roughly 8% voiced-stop realizations of /p/ in this context, and no such realizations of /k/. While the possibility of intervocalic voiced stop realizations is not in doubt, it is worth noting that Warner & Tucker (2011) find partial *devoicing* of voiced stops in these contexts to be more common in two out of three tasks. This is consistent with earlier findings that the presence of closure voicing is one of the strongest cues to the voicing contrast in intervocalic position in English and that even stops with silent closures can be perceived as voiced in some cases (Lisker 1978). The overall conclusion to be drawn from this work is that, while voicing of non-coronal voiceless stops in flapping contexts is possible in American English, it is far less likely than voiced-stop lenition or flapping, and no more likely than the converse devoicing of voiced stops.

The presence of various lenition-like phenomena in American English has consequences for the interpretation of the experiment to be described in the next sections. In particular, given that the experiment investigates whether lenition processes confer an advantage for a particular type of recognition task, it will be important to ask whether or not performance differences stem from previous experience with English. Although it is impossible to entirely rule out such an influence, we have taken some steps to try to minimize the impact of English experience, and have included some post-hoc analyses (described in Section 3.2) that allow us to test whether specific aspects of English are influencing the results.

One precaution we have taken is our avoidance of flapping in the experimental stimuli. This is the most robust and categorical of the lenition phenomena just discussed, and is itself a kind of voicing lenition in the case of /t/. For our voicing stimuli, we omit coronal stops entirely, using the fricatives /θ/-/ð/ instead. For the spirantization stimuli, the coronal alternation is exemplified by a voiced dental stop alternating with a dental approximant rather than a tap; English speakers should have little experience with this type of segment, either as an underlying phonological category (because dental continuants are generally fricated in English) or as an outcome of lenition (because flaps are the most likely outcome for coronal stops).

A more general measure we take is to make the overall phonetic context of the experimental consonants quite unlike English flapping contexts. Consonants in the experiment are always followed by full vowels, with duration, intensity, and f<sub>0</sub> equal to the surrounding vowels. To the extent that these are perceived as English vowels, then, they should be perceived as bearing some degree of stress. Given the rarity of all the phenomena discussed above in stressed contexts, there is no reason for an English listener to expect lenited consonants here.

Finally, the consonants themselves are fairly different from English ones, particularly with regard to laryngeal features. Voicing of stops in the experiment does not vary by context: voiced stops are invariantly prevoiced, and voiceless ones invariantly feature a voiceless (simulated) closure and short-lag VOT. While these are plausible realizations of English stops in flapping contexts, they would be fairly unlikely in word-initial position, where most English voiced stops are not prevoiced and most voiceless ones are aspirated (e.g. Trager & Smith 1951; Cooper et al. 1952; Lisker & Abramson 1964).

### **1.3 Word segmentation: A paradigm for testing boundary detection**

The experimental paradigm used in the current studies is typically referred to in the language acquisition literature as *word segmentation*. The efficacy of this paradigm for studying short-term learning of word-like chunks in the laboratory was first demonstrated in a series of pioneering studies by Saffran, Aslin & Newport (1996a; b; see also earlier work

on non-speech linguistic analogs by Hayes & Clark 1970). These studies used synthesized speech to examine adult and infant listeners' ability to track transitional probabilities, or the relative statistical coherence, of acoustically continuous strings of syllables. Following a short period of exposure to an artificial "language" made up of CV syllables concatenated into short "words", which were in turn concatenated into "utterances", listeners were tested on their ability to distinguish between strings of syllables that always or very often occurred in adjacent position during the exposure phase (i.e. sequences of syllables that formed words) and those that did so less often (i.e. sequences of syllables that spanned word boundaries). The paradigm is referred to as *word segmentation* because it tests listeners' ability to divide a continuous acoustic stream into recombining units (or "segments") based on their relative statistical coherence; highly coherent sequences are more likely to be parsed into recombining segments than are relatively unlikely sequences. Such studies are careful to avoid presenting actual words, but a frequently cited example that can help clarify the paradigm is the four-syllable sequence *pretty#baby*, where the syllables *pre-* and *-tty* are highly coherent given their status as an English word, while the sequence *-tty#ba-* is less coherent given the intervening word boundary. Stated more simply, *pre-* is highly likely to be followed by *-tty*, whereas *-tty* may be followed by a wider variety of syllables due to its position at the end of the word (e.g. *pretty#doggy*, *pretty#kitty*, *pretty#ball*, etc.).

It is worth noting that while the paradigm is typically referred to as *word segmentation*, the notion of statistical coherence is certainly not restricted to words, as it can be applied to larger or smaller linguistic constituents, such as phonological segments, morphemes, feet, and prosodic words or phrases. Indeed, since the early studies of Saffran et al. (1996a; b), segmentation paradigms have been used to study numerous aspects of so-called *statistical chunking* phenomena, and have more recently led to the hypothesis that such statistical learning mechanisms may in fact be domain-general, accounting for findings involving visual, motor, and multi-modal learning (see Mareschal & French 2017 for a recent computational model that can be applied across domains, and other papers in the same volume, edited by Armstrong, Frost & Christiansen 2017, for a sampling of current perspectives on statistical learning more generally).

Most relevant for the current experiments, some prior work employing the word segmentation paradigm has examined whether particular acoustic cues can further bolster learning achieved via tracking of transitional probabilities alone. Saffran et al. (1996b), for example, compared a condition in which adults attempted to learn words from syllable transitional probabilities (as described above) to conditions in which either the first or last syllable of their three-syllable nonce words had been lengthened by 100 ms. The results showed that, relative to the non-lengthened condition, word-final lengthening provided a boost in word segmentation performance, whereas word-initial lengthening did not.

It is unclear whether these findings are related to the fact that the listeners in the study were native English speakers, and English prosody features a domain-final lengthening rule (e.g. Wightman et al. 1992), and/or whether they might reflect a more general strategy of so-called *prosodic bootstrapping* (Gleitman & Wanner 1982) that could be successfully applied not only to English, but to any language in which prosodic cues correlate with constituent boundaries in a predictable way (that is, presumably all natural languages; see Fisher & Tokura 1996 for a cross-linguistic comparison of acoustic cues to syntactic boundaries in infant-directed speech, and Tyler & Cutler 2009 for more recent work on language-specific versus universal aspects of adults' use of acoustic cues in a word segmentation paradigm).

While a small body of work has explored the role of acoustic-phonetic cues to constituent boundaries in both infant and adult word segmentation performance, these studies

have typically focused on prosodic or suprasegmental information (see, e.g., Bagou et al. 2002 for evidence that French-speaking adults learn artificial words more robustly when statistical coherence is reinforced by French-specific durational and intonational cues, Kim 2004 for similar findings in Korean, and Johnson & Jusczyk 2001 for a study explicitly pitting word-level stress *against* statistical coherence in a study of infant segmentation performance).

Comparatively less work has asked whether predictable phonetic variation affecting the realization of individual phonological segments is similarly beneficial for word segmentation performance. In their study pitting acoustic cues against statistical coherence, however, Johnson & Jusczyk (2001) did find that segmental coarticulation (like word-level stress) was weighted more strongly than statistical coherence in a word segmentation experiment with 8-month-olds. Other experimental paradigms have provided evidence that the presence of unreleased stops (Jusczyk et al. 1999), glottalization (Nakatani & Dukes 1977), and the light/dark lateral allophony attested in English (Nakatani & Dukes 1977) may aid in segmentation performance. Importantly, however, all of these studies have targeted patterns that are specific to listeners' native languages, making it difficult to draw conclusions regarding the general role of segmental acoustic variation and allophony in word boundary detection. The present study thus promises to address this gap in the word segmentation literature, while simultaneously shedding light on the question of whether the prevalence of a cross-linguistically common phonological pattern could be due in part to its potentially functional role in language processing.

#### 1.4 The current study

The current study asks whether the presence of spirantization or voicing lenition improve English speakers' performance on a word segmentation task. To test this, we trained each subject on an artificial language that either featured one of these lenition patterns or featured an anti-lenition pattern, where consonants are lenis initially and fortis medially. The basic hypothesis is that if lenition patterns help listeners detect constituent boundaries in speech, then learning words in the lenition languages should be easier than in the anti-lenition languages. We chose anti-lenition patterns as a comparison because they differ from lenition patterns only in the *content* of the pattern, not in its presence or absence. Comparing lenition languages to languages with no pattern would not allow us to determine whether any performance difference is due to inherent properties of lenition, or simply due to the presence of a pattern.

Even if an advantage for lenition languages is found, there will remain the question of whether the *inherent* properties of lenition are truly the operative factor. The literature reviewed in section 1.3 demonstrates that language-specific phonetic patterns can aid the detection of statistically coherent units in the word segmentation paradigm, although we are not aware of any investigation of a lenition pattern in this regard. Recall that American English displays very frequent flapping of coronal stops and somewhat frequent spirantization of non-coronal voiced stops. If English speakers are sensitive to these properties, then they could perform better on a spirantization language (and possibly voicing as well) just by applying their knowledge of English phonetics. A slightly different possible effect related to English exposure is that phonotactic generalizations over lexical items might make one type of language more similar to English than the other. Either of these would be a reasonably interesting finding in and of itself, but we attempt in our analyses to distinguish between these English-exposure hypotheses and the hypothesis that lenition itself is inherently useful for boundary marking.

To this end, we examine a number of variables pertaining to place of articulation. Because generalizations about both phonetic lenition and lexical distributions of consonants vary

by place of articulation in English, the simplest form of the English-exposure hypotheses would predict that some types of consonants should be more reliable boundary markers than others. This should affect subjects' response bias: for instance, if /d/ is particularly unlikely to show up as a voiced stop in domain-medial position in English, and subjects are applying that knowledge to the experiment, they should be less likely to select strings with medial [d] as words, regardless of whether such strings were statistically coherent in the training phase. The boundary-disruption theory, on the other hand, predicts no differences between consonants at different places of articulation, as long as their duration and intensity are roughly uniform.

## 2 Methods

### 2.1 Materials

Participants in the two studies were exposed to one of five possible artificial languages, composed of acoustically continuous and prosodically undifferentiated strings of syllables grouped into words and utterances. The phonetic details of the materials are described in section 2.2. Our global design is based on the results of Frank et al. (2010). This methodological investigation manipulated different aspects of the word segmentation paradigm, including the number of distinct words in a language, the variability of those words, and the length of the exposure period, in order to obtain a more general model of human word segmentation performance. We generally picked values for these parameters that resulted in intermediate performance in Frank et al.'s (2010) studies; in this way, listeners have room to demonstrate either improvement or decline as a function of the phonetic patterns of interest.

All languages had six distinct words, which ranged from two to four syllables. The syllables were of the form CV, combining the vowels [a], [i], and [u] with consonants dictated by the particular condition. For the spirantization conditions, the consonants were [b], [d], [g], [β], [ð], and [ɣ]. For the voicing conditions, they were [p], [θ], [k], [b], [ð], and [g]. During the exposure period, listeners were played 150 utterances (that is, 150 strings of four words each, concatenated into a continuous acoustic stream), separated by short (800 ms) pauses. The order of words was pseudo-randomized so that the same word was never repeated twice in a row in a single utterance, and the exposure period generally lasted about eight minutes.

After the exposure period, subjects completed a two-alternative forced-choice task in which they were asked which of two strings of syllables (again presented acoustically, with no accompanying orthographic representation) was "a word in the language you've just heard". The *target* in each trial was a word from the language, where most transitions between syllables had a probability of 1.0 in the exposure period (three syllables had to be used in more than one word, such that transitions involving these three had probabilities of 0.5 given the surrounding syllables). The *foil* consisted of one or more syllables composed of the end of one word, concatenated with one or more syllables from the beginning of a different word. In these foils, most word-internal sequences had transitional probabilities of 1.0, but cross-word sequences had an average transitional probability of 0.33 (with slightly larger or smaller values depending on exact details of randomization). There were a total of six targets and six foils. Each target was heard five times in the testing phase, paired with a different foil each time, for a total of 30 test trials. Subjects' likelihood of labeling a target as a word relative to labeling a foil as a word can be used to derive a measure of *sensitivity* to the target versus foil distinction, that is, an index of how much they've learned about the language in question.

For each experiment, subjects were exposed to either a lenition language or an anti-lenition language. In the lenition conditions, words featured fortis segments initially: stops in

the spirantization conditions, voiceless obstruents in the voicing condition. Medial consonants were lenis: continuants in the spirantization conditions, voiced obstruents in the voicing condition. The anti-lenition conditions were statistically identical to their lenited counterparts, but initial consonants were lenis and medial ones fortis. These anti-lenition patterns, to the best of our knowledge, are unattested in natural languages. Comparing performance on the lenition and anti-lenition conditions allows us to isolate the effect of lenition-related phonetic *content* on segmentation, rather than the effect of mere presence of a phonetic pattern, providing a direct test of the boundary disruption theory.

The list of words in each condition, three examples of utterances played during the exposure phase, and a test pair for the spirantization lenition condition are shown in Figure 1.

We ran two versions of the spirantization condition with two slightly different phonetic realizations, referred to here as “Spir1” and “Spir2”. Spir1 had more or less approximant-like continuants, while Spir2 had somewhat more intense and vowel-like continuants, comparable to glides. The purpose of this design was twofold: we were interested in the question of whether any putative lenition effect had a syntagmatic element, that is, whether a *larger* difference between fortis and lenis consonants would result in a larger learning advantage; and we also wanted to check whether any putative lenition effect would generalize across a somewhat larger range of lenis consonant sounds than just the three included in Spir1.

## 2.2 Synthesis

Syllables were synthesized using Keith Johnson’s Unix implementation of the KLSYN–88 synthesizer (Klatt & Klatt 1990). Formant targets for consonants and vowels were based on recordings of a female Madrileño Spanish speaker. After some experimentation, we found the best method for interpolating formant transitions and amplitude profiles of segments was to use the values given in the original Klatt “cookbook” manual. Using values for these parameters based on Spanish recordings instead resulted in stimuli that were hard to identify.

<b>Words:</b>			
<u>Spirantization</u>	<u>Anti-spirantization</u>	<u>Voicing</u>	<u>Anti-voicing</u>
baði	βadi	paði	baθi
guβa	γuba	kuba	gupa
daβuyi	δabugi	θabugi	δapuki
biyaðu	βigadu	pigaðu	bikaðu
giðaγuβi	γidagubi	kiðagubi	giθakupi
duβiðuγa	δubiduga	θubiðuga	δupiθuka
<b>Example utterances in spirantization condition:</b>			
guβagiðaγuβiguβabadi			
giðaγuβidaβuyiduβiðuγadaβuyi			
guβaduβiðuγagiðaγuβiduβiðuγa			
<b>Example test pair in spirantization condition:</b>			
daβuyi vs. βagiða			

**Figure 1:** Words used in the construction of each artificial language (top); examples of utterances in the spirantization lenition condition (middle); example of a test pair heard in the spirantization condition, with target first and foil second (bottom).

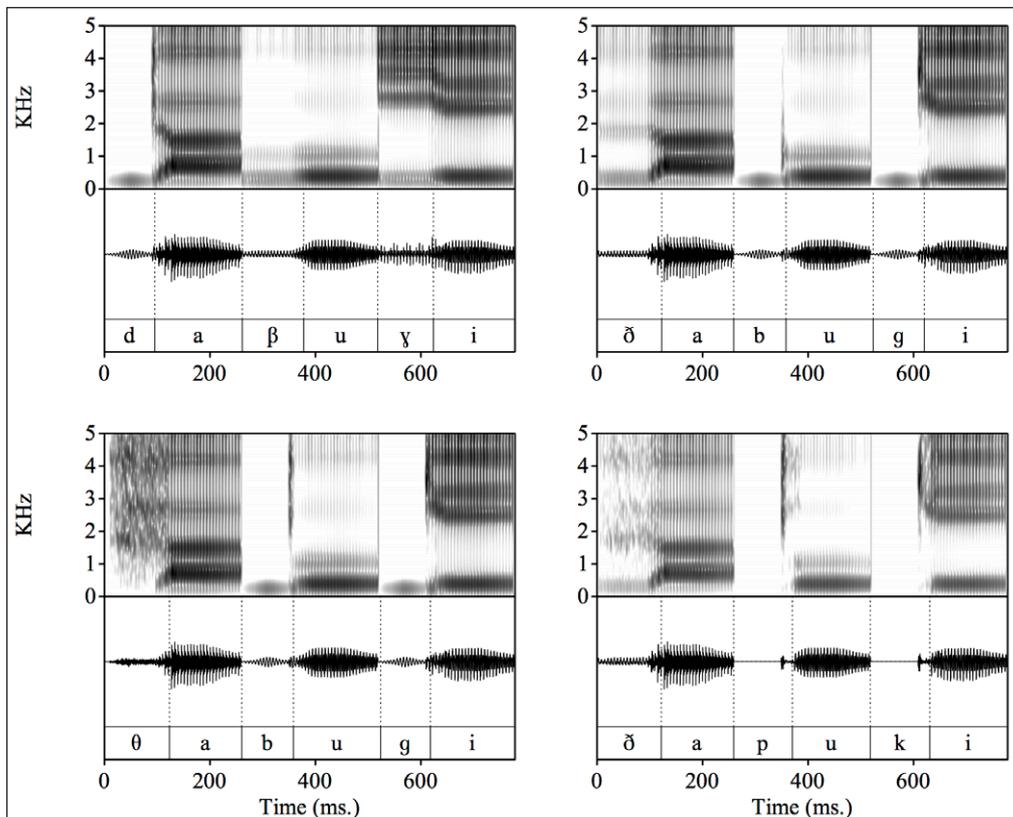
F0 was held flat at 165 Hz in all syllables (except during voiceless sounds). Consonants with their accompanying formant transitions were 120 ms long, with steady-state vowels of 240 ms. Voiced stops had prevoicing for their entire “closure” durations. Continuants in the spirantization and anti-spirantization conditions had no noise component; that is, they were realized as sonorants rather than fricatives. Voiceless stops had very light aspiration and a VOT of 10–20 ms. Continuants in the voicing and anti-voicing conditions, all of which were (simulated) dental, did have noise components. The voiced and voiceless stimuli in the voicing and anti-voicing conditions differed not only in phonetic voicing, but also in the amplitude of their bursts and presence of light aspiration (for stops), and amplitude of frication (for fricatives). Syllables were concatenated together with no intervening pauses to make words, and words were similarly concatenated together to make utterances. Utterances were separated by 800 ms of silence.

Example spectrograms and waveforms of one word from each condition are shown in Figure 2.

### 2.3 Procedure

Participants were seated at a workstation with AKG K240 headphones. The experiment was run in Open Sesame (Mathôt, Schreij & Theeuwes 2012). They were told that they would “listen to a foreign language for a few minutes” and then “be tested on the words that appear in that language”. The exposure period consisted of passive listening for about 8 minutes, and participants were not given any special instructions about what to do during that period other than to “listen to the speaker”.

After the exposure period, they were told: “In the next part of the experiment, you will hear two words at a time. Only one of them is a word from the language you heard.” They



**Figure 2:** Example spectrograms and waveforms for words in the spirantization (top row) and voicing (bottom row) experiments, with lenition (left column) and anti-lenition (right) patterns.

were told to push keyboard buttons to indicate whether the first or second word was from the language they heard. The total duration of the experiment was around 20 minutes for most participants.

## 2.4 Participants

Participants were 135 undergraduates at UC Berkeley, who participated for course credit. This sample included a large number of non-native English speakers. Results reported here are for the 90 native English speakers only (on average 17–19 per experimental condition). None of the subjects reported being diagnosed with any speech, hearing, or reading disorders.

## 2.5 Analysis

Data were analyzed with logit mixed-effects regression models using the lme4 package (Bates et al. 2015), v. 1–13, in the statistical platform R. Logistic regression estimates the log odds (or *logit*) of a binary outcome variable as a function of a number of predictors. Mixed-effects models estimate the influence of *fixed effects*, those for which all possible values are thought to be known, on the outcome variable, while taking into account the influence of *random effects*, those that are sampled randomly from an underlying population that is not of primary interest to a study. Fixed effects here are reported with the coefficient  $\beta$ , which gives the effect size, the model's estimate of standard error for that effect, a  $z$  statistic from the Wald test, which translates the effect size into units of standard error, and a  $p$ -value derived from the Wald test statistic.

In the current study, the question of interest is which of the presented auditory stimuli in a given test pair is labeled as a word from the language heard earlier. However, given the difficulty of the task and the likelihood that individual participants could have stronger or weaker biases toward, e.g., simply answering “the second stimulus” under conditions of uncertainty, we chose to model participants' probability of choosing the second stimulus presented during the test trial as our dependent variable. The result is that the overall model intercept corresponds to the overall log odds of a false alarm selecting the most recently heard stimulus as a word when it was actually a foil. When these false alarm parameters are subtracted from the corresponding hit rates, the resulting sensitivity parameters are corrected for any response bias associated with the most recently heard stimulus (or responding with a particular hand). We included random intercepts by subjects and by items; these account for differences among participants in the degree of “second stimulus” bias, and for differences in the degree of such a bias that could be associated with particular stimuli. We also included by-subject random slopes for sensitivity terms; these capture variance between subjects in overall ability to tell words from non-words.

We included fixed effects for the experimental manipulations of interest: whether the second stimulus in a pair was a target or foil, and the condition of the experiment (lenition vs. anti-lenition, spirantization, voicing, etc.), as well as task-related effects such as trial number (how far into the experiment the trial occurred) and whether or not there was an error or timeout on the preceding trial. These were entered both as main effects (revealing any effects of these variables on the “second stimulus” bias) and as interactions with the variable “second stimulus = target”.

The resulting models estimate the log odds of responding “second” on a trial where the target is first (thereby modeling the false alarm rate), then compare this to the log odds of responding “second” when the target is second (thereby capturing the hit rate). The difference between these two coefficients (in log odds) is a measure of sensitivity: how much more likely participants were to respond “word” to actual words than to part-words,

taking into account their bias to simply respond with the most recently heard stimulus. The difference in log odds between hits and false alarms is closely related but not exactly equivalent to the signal detection-theoretic measure  $d'$ . For empirical plots in section 3 showing means and by-subject variability, we use the better-known  $d'$  measure.

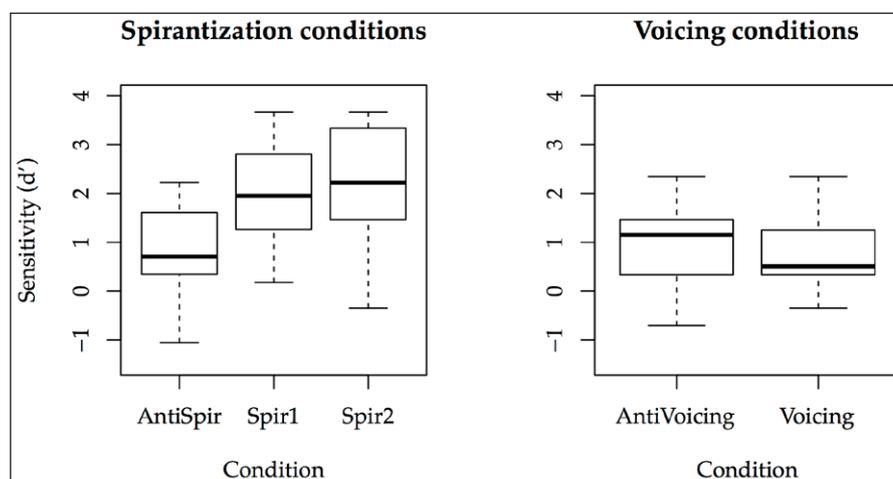
Models were fitted using the maximal possible random effects structure, following Barr et al. (2013). Full data are included in the Supplementary Materials for this paper.

### 3 Results

#### 3.1 Sensitivity

Sensitivity across all conditions is shown in Figure 3. These are empirical plots of by-subject  $d'$ , a measure of sensitivity consisting of the difference between z-transformed hits and z-transformed false alarms. Chance performance on this measure is zero. The statistical models reported below assess sensitivity in terms of the closely related difference between logit hit and logit false alarms (or equivalently, the log of the odds ratio of hits to false alarms). Median performance in all conditions is greater than chance. Sensitivity in both spirantization conditions is higher than the anti-spirantization condition (~83% vs. 64% accuracy). The two spirantization conditions do not differ very much from one another (82% vs. 84% accuracy). The voicing and anti-voicing conditions display very similar performance to one another: median  $d'$  is somewhat higher in the anti-voicing condition, but the distribution is more positively skewed in the voicing condition.

Fixed effects from logit mixed regression models for both experiments are given in Tables 1 and 2. The significant negative model intercepts indicate that participants in both experiments displayed false alarms less often than 50% of the time. Sensitivity is well above chance in the Spir1 condition, which was treated as the reference level for dummy-coded condition (effect 1.6). The difference in sensitivity between Anti-Spir and Spir1 is large and significant (effect 1.7). Sensitivity is somewhat higher in Spir2 than in Spir1 (effect 1.8), but this trend does not approach statistical significance. Sensitivity is above chance in the baseline voicing condition (effect 2.5). The model estimates that sensitivity in the voicing condition is about 0.32 logits higher than in the anti-voicing condition, but this trend is nowhere near statistical significance (effect 2.6).



**Figure 3:** By-subject  $d'$  in the spirantization (left) and voicing (right) experiments by condition. Horizontal lines are median values; boxes are interquartile intervals; whiskers are ranges.

**Table 1:** Fixed effects in the spirantization conditions.

	Fixed effects	$\beta$	SE	z	p
<b>False-alarm parameters</b>					
1.1	Intercept	-1.66	0.41	-4.08	<b>&lt;0.001</b>
1.2	Cond: AntiSpir	0.92	0.41	2.22	<b>0.03</b>
1.3	Cond: Spir2	-0.56	0.45	-1.25	0.21
1.4	Post-error	0.03	0.25	0.10	0.92
1.5	LogTrial	0.10	0.13	0.76	0.45
<b>Sensitivity parameters</b>					
1.6	TargetSecond	4.52	0.73	6.18	<b>&lt;0.001</b>
1.7	TargSec*AntiSpir	-2.57	0.74	-3.48	<b>&lt;0.001</b>
1.8	TargSec*Spir2	0.53	0.80	0.67	0.50
1.9	TargSec*PostErr	0.54	0.38	1.41	0.16
1.10	TargSec*LogTri	-0.38	0.20	-1.92	<b>0.05</b>

**Table 2:** Fixed effects in the voicing conditions.

	Fixed effects	$\beta$	SE	z	p
<b>False-alarm parameters</b>					
2.1	Intercept	-1.40	0.37	-3.79	<b>&lt;0.001</b>
2.2	Cond: AntiVoi	-0.13	0.37	-0.35	0.73
2.3	Post-switch	-0.16	0.22	-0.72	0.47
2.4	Trial	0.04	0.01	2.98	<b>0.003</b>
<b>Sensitivity parameters</b>					
2.5	TargetSecond	2.34	0.63	3.71	<b>&lt;0.001</b>
2.6	TargSec*AntiVoi	-0.32	0.63	-0.51	0.61
2.7	TargSec*PostSwitch	0.68	0.31	2.20	<b>0.03</b>
2.8	TargSec*Trial	-0.06	0.02	-2.91	<b>0.004</b>

A number of the task-related variables included here also affected sensitivity. In both experiments, sensitivity decreased during the course of the experiment as measured by trial number. For the spirantization conditions, the natural logarithm of trial number was the best predictor (effect 1.10). For the voicing conditions, “plain” (non-transformed) trial number explained more variance (effect 2.8). In the spirantization conditions, participants were slightly more accurate on trials following an error (effect 1.9), but this effect becomes non-significant after adding random slopes to the model. In the voicing conditions, sensitivity was significantly higher on trials where the correct answer was different from the preceding trial (“PostSwitch”, effect 2.7).

As mentioned in section 2.1, some of the target words had one syllabic bigram with a transitional probability of 0.5 rather than 1.0, while most of the targets were purely composed of bigrams with 1.0 transitional probabilities. While these “low probability” targets would still be substantially more probable in the training sequence than any of the foils, a reviewer suggests that we check whether sensitivity to these words is lower than the other targets. The answer is that the distinction does not appear to affect sensitivity: in the spirantization conditions, about 77% of both low- and high-probability targets are

correctly identified as words; in the voicing conditions, about 68% of the high-probability and 67% of the low-probability targets are correctly identified as words. This is qualitatively consistent with Saffran et al. (1996a), whose target stimuli had averaged transitional probabilities ranging from 0.31 to 1.0, and who found only a small difference between their highest and lowest probability targets (79% vs. 72% accuracy, respectively).

### 3.2 Testing for effects of English experience

In order to investigate the hypothesis that our results reflect experience with English phonology, we conducted a post-hoc analysis to determine whether response accuracy or bias differed by consonant place of articulation. The probability (pooled across subjects) of choosing each target word or foil as belonging to the exposure language was first calculated, resulting in “percentage hits” for each target, and “percentage false alarms” for each foil. (The Spir2 condition, the results for which were statistically indistinguishable from Spir1, was not included in these analyses.) These by-stimulus percentages were then averaged according to the position, place, and manner of the consonants in the stimulus, with the result that a given stimulus contributes a separate data point for each of its consonants. While this “double-counting” is not appropriate for statistical analysis, it does allow a qualitative look at the data as a function of all predictors that might be expected to affect the results, while avoiding problems with model convergence and statistical power that could result from adding more predictors to the existing models. The resulting plots are shown in Figure 4.

One way to investigate possible English influence in our results is to consider whether participants’ responses reflect the probability that a given sound will occur in a given position in English. Perhaps the strongest predictions can be made with respect to the interdental fricative /ð/; a reviewer points out that /ð/ never appears in word-initial position in English content words. If this knowledge of the English lexicon is impacting our results, then we should expect that target stimuli beginning with voiced coronal continuants (i.e. in the anti-spirantization and anti-voicing conditions) should show the lowest rates of hits, since listeners’ prior experience with English may have led to the inference that word-initial /ð/ is not possible, whereas no such inference may exist for /β/ or /ʏ/. Similarly, foils beginning with voiced coronal continuants (i.e. in the spirantization and voicing versions of the experiment) should show the lowest rates of false alarms.

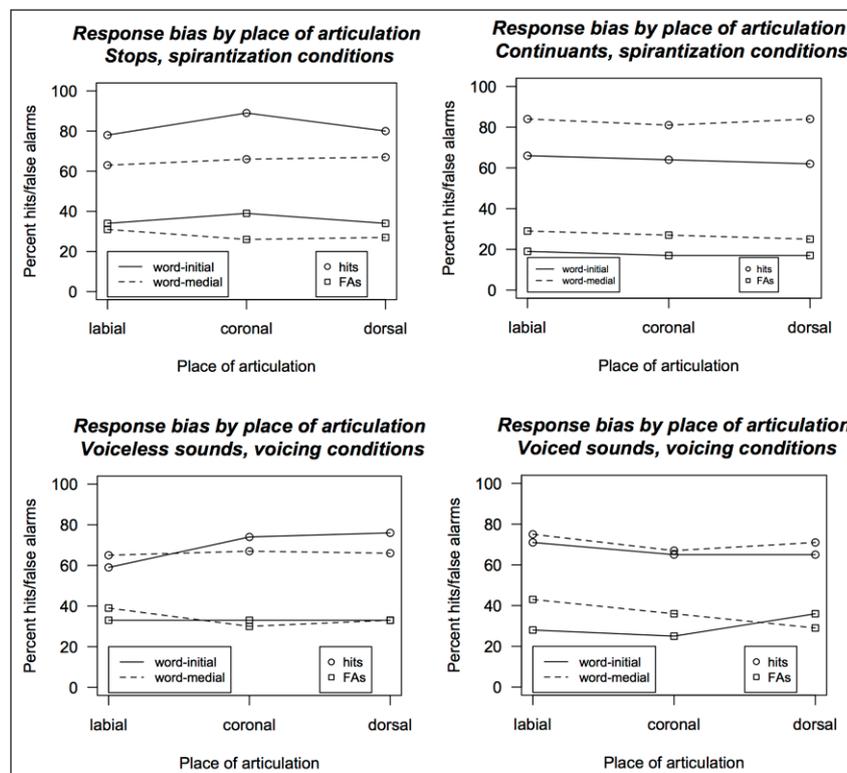
The top right panel of Figure 4 shows the hit and false alarm rates for stimuli with word-initial /ð/ in the spirantization conditions of the experiment. Contrary to the above predictions, there is no indication that targets beginning with /ð/ are any less likely to be correctly chosen (hits) than targets beginning with any other place of articulation. Likewise, there is also no indication that foils beginning with /ð/ (false alarms) are relatively less likely to be chosen. The bottom right panel of Figure 4 shows the hit and false alarm rates for stimuli with word-initial /ð/ in the voicing conditions of the experiment. Also contrary to predictions, word-initial /ð/ does not appear to affect hit rates or false alarm rates relative to other places of articulation. It appears as though participants in our experiments were not transferring their knowledge of where /ð/ can occur in English to our experimental task. To the extent that /ð/ shows the most restricted English distribution of any of the consonants in our experiments, we tentatively conclude that English-specific lexical statistics are not a primary driver of the present results.

English lexical statistics are not the only type of prior phonological knowledge that could have impacted our results; recall that while spirantization and voicing lenition are not generally considered robust phonological processes in English, previous work indicates that they do exist as low-level, variable phonetic processes. Thus with respect to the phonological alternations tested in our experiments, we should predict the greatest

number of hits for target words – and the greatest number of false alarms for foils – for places of articulation that are most likely to undergo (variable, low-level) phonetic lenition in English. Further, any such biases should be expected to hold steady across experiment conditions (i.e. regardless of whether the exposure period employed lenition or anti-lenition stimuli).

Concerning spirantization, recall from Section 1.2 that in contexts where coronal stops are realized as flaps, voiced non-coronal stops are reportedly realized with approximant-like phonetic characteristics approximately 40 to 80% of the time, with higher rates of approximant realization for /g/ than /b/ (e.g. Lavoie 2001; Walter 2007; Bouavichith & Davidson 2013). Recall also that in the spirantization conditions, we employed a /d/-/ð/ alternation in order to avoid any flapping, resulting in a maximally non-English-like pattern at the coronal place of articulation. If English experience is significantly contributing to our results, then, we should observe the greatest number of hits for stimuli in the spirantization conditions that contain word-medial /ɣ/, followed by word-medial /β/ and then /ð/. Similarly, the number of false alarms for foils should exhibit the same pattern (word-medial /ɣ/ > /β/ > /ð/). The predictions for the non-lenited consonants are the mirror image: target stimuli containing word-medial voiced stops should show the *least* number of hits for /g/, since it is the most likely to spirantize word-medially, followed by /b/ and then /d/, with the same pattern obtaining for the false alarms.

Word-medial continuants in the spirantization conditions are plotted in the upper right panel of Figure 4, and word-medial stops are plotted in the upper left panel of Figure 4. There is no indication that place of articulation for word-medial stops has any appreciable effect on either the hit rate or the false alarm rate in the spirantization conditions.



**Figure 4:** Proportion hits (for targets) and false alarms (for foils) for probe items containing consonants at different places of articulation in initial and medial positions. Separate charts show voiced stops (top left) and continuants (top right) in spirantization conditions, and voiceless obstruents (bottom left) and voiced obstruents (bottom right) in voicing conditions.

With respect to voicing lenition, previous work reports much lower rates of voicing lenition than spirantization in English: only 5–20% in Warner & Tucker (2011), and Bouavichith (2014) finds 8% voicing of /p/ to /b/, with no voicing of /k/ to /g/. (Recall that our voicing conditions employed a /θ/-/ð/ alternation at the coronal place of articulation in order to avoid flapping.) Thus if experience with English is a driving factor in our experiment, we should predict the highest number of hits for stimuli in the voicing conditions that contain word-medial /b/, followed by /g/, and the number of false alarms for foils in these conditions should show the same pattern. With respect to the non-lenited consonants, the voiceless obstruents, we would then predict the *least* number of hits for targets containing word-medial /p/, since it is the most likely to devoice word-medially, with more hits for targets containing word-medial /k/, and the same pattern for false alarms.

The bottom two panels of Figure 4 show the results for the voicing conditions of the experiment. For voiced stops (bottom right), there does seem to be a slight trend for word-medial /b/ to result in more false alarms than the other places of articulation. For the voiceless stops, however, (bottom left), word-medial voiceless /p/ does not appear to result in appreciably fewer hits or false alarms. Given the weakness of both the predictions and the trends regarding place of articulation and voicing lenition, we conclude that the results of the voicing experiments do not strongly favor the hypothesis that English knowledge impacted the outcome of our experiments.

In summary, there do not appear to be any obvious or large effects of place of articulation on subjects' responses. We tentatively conclude that prior experience with English did not play a significant role in the present experiment.

## 4 Discussion

The experiments reported on here find limited evidence that lenition patterns aid English listeners in segmenting the speech stream. Performance in the spirantization conditions is considerably better than the anti-spirantization condition, and this result is statistically quite robust. Performance in the voicing condition, however, is only marginally better than the anti-voicing condition, and this effect is highly variable, not reaching statistical significance. The effect on sensitivity of the slightly more lenis approximants used in the Spir2 condition relative to Spir1 is also small, variable, and non-significant. We found few asymmetries in response bias by place of articulation for targets or foils in any condition, suggesting that overall effects of phonetic/phonological patterns in the study are not being driven by subjects' preference for particular segments in particular positions.

### 4.1 Spirantization and boundary-disruption

The boundary-disruption approach holds that spirantization is widespread in human languages in part because it aids listeners in detecting linguistic constituents, by aligning moments of auditory disruption in the speech stream with constituent boundaries. A clear prediction arising from this proposal is that in a speech perception paradigm, spirantization should aid listeners in segmenting the speech stream into constituents. In the spirantization conditions, we found a large, robust effect in this direction. As such, these results can be seen as providing support for the boundary-disruption approach, although there are other possible interpretations, discussed in Section 4.3.

The existence of a spirantization effect is of interest in and of itself, as there have been few demonstrations in the word segmentation literature that allophonic variation in manner of articulation can have an effect on segmentation. This is also, to the best of our knowledge, the first investigation of a common lenition pattern in the word segmentation paradigm. One practical implication of this result is that, in studies where

researchers are investigating statistical and/or distributional factors in word segmentation, it is important to pay attention to and tightly control segmental characteristics of words in various conditions.

The finding that spirantization improves word segmentation performance is also interesting from the perspective of phonological theory. Most literature on lenition has tended to locate its functional motivation firmly in the articulatory sphere (e.g. Donegan & Stampe 1979; Bauer 1988; Kirchner 1998), but some more recent accounts have instead posited perceptual or information-processing accounts of lenition (e.g. Harris 2003; Kingston 2008; Katz 2016; Cohen Priva 2017). The current study is one of the first to show that lenition patterns can exert a significant effect on the processing of novel linguistic items. This result provides *prima facie* evidence that the perception and processing of lenition-fortition patterns is worth focusing on and exploring in more detail, along the lines of the theoretical work mentioned above.

There is little evidence here for a syntagmatic effect of lenition, whereby larger phonetic differences between fortis and lenis segments aid segmentation more. While a small trend in this direction was observed, it is fairly variable and not statistically robust. Performance was already quite high in the Spir1 condition (about 82% correct across all subjects), and if this is close to the “ceiling” for a task like this there may not have been enough “room” for listeners in the Spir2 condition to demonstrate increased sensitivity. Another possibility is that, while the difference between close and open approximants here was fairly large in acoustic terms, it may not have had much of a perceptual effect. A related idea is that the perceptual difference between stops and approximants is already so large that making the approximants even more sonorous simply doesn’t have much of an effect.

#### **4.2 Voicing and boundary-disruption**

The lack of a significant segmentation effect from voicing lenition is problematic from the perspective of the boundary-disruption approach. Because voicing increases the amount of low-frequency energy in a stop, rendering it more vowel-like, the approach predicts that voiced stops should be less disruptive to a stream of vowels than voiceless ones, making segmentation easier in the voicing than anti-voicing condition. One possibility is that the boundary-disruption approach is broadly correct for spirantization but not for voicing. This would be somewhat puzzling, considering that the two processes have many similarities in terms of their characteristic phonological environments and their limited interactions with phonological contrast, as well as the fact that they sometimes operate in tandem (e.g. Hyman 1972 on Fe’fe’; Ladd & Scobbie 2003 on Logudorese Sardinian).

As with all null results, there are numerous other possible interpretations. One is that voicing lenition *does* facilitate word segmentation, but the effect is much smaller than for spirantization, too small to produce statistically robust results in our experimental design. This would be consistent with the apparent fact that spirantization is more widely attested in the world’s languages than voicing lenition. While the typological surveys mentioned in Section 1.1 are not meant to be genealogically balanced or representative, it is notable that spirantization occurs about twice as often as voicing lenition. Gurevich (2003), for instance, reports 76 languages with spirantization and 39 with voicing. This is consistent with the idea that voicing lenition does aid segmentation, but not as efficiently as spirantization.

Another possibility is that our (admittedly quite unnatural-sounding) synthesized stimuli were in some way qualitatively unlike the voicing lenition found in natural languages. We chose to synthesize our stimuli in order to afford complete control over their acoustic characteristics, ensuring, for example, that no acoustic correlates of prosody could be present. It is possible, though, that the phonetic manipulations used to represent voicing lenition

were insufficient to create a perceptual difference in “disruption” between the two types of consonants. Impressionistically, the voiceless and voiced obstruents (especially stops) did seem harder to tell apart than the stop and continuant ones used in the spirantization conditions.<sup>1</sup> One way to address this issue in future work would be to use additional cues to enhance the distinction between voiced and voiceless obstruents. In work currently underway, we are adding duration differences to this condition, making voiceless obstruents about 1.5 times as long as voiced ones. This is the type of pattern generally found in languages with (at least contrastive) voicing. Interestingly, Katz (2016) suggests on independent grounds that the driving force behind so-called voicing lenition may in fact be shortening (cf. Lavoie 2001; Ennever et al. 2017). Katz suggests that if a durational constraint is driving voicing lenition, it would help explain both cross-linguistic patterns in how voicing lenition applies (or fails to apply) to obstruents in consonant clusters, as well as the fact that voicing lenition is so often optional or variable in languages where it is reported.

An additional possibility is that the English listeners in our experiment are able to use spirantization to aid segmentation, but not voicing, because English (or some other languages they’ve been exposed to) displays spirantization patterns but rarely voicing lenition. We turn to this possibility in the next section.

### 4.3 The role of language experience

As discussed in Section 1.2, English displays fairly frequent spirantization lenition in some contexts, but rarely voicing lenition. Given that English-speaking subjects in our experiment showed improved performance with a spirantization pattern but not a voicing one, one obvious hypothesis is that listeners are using their knowledge of English phonetics to guide expectations in these artificial languages. As noted above, we took several measures in designing the stimuli to make them quite different from English, in order to discourage listeners from using this type of parsing strategy. But of course we can’t rule out the possibility that these measures failed and listeners did, in the end, treat these stimuli in an English-like way. We provided some post-hoc analyses in Section 3.2 that would seem to argue against this interpretation. In this section, we discuss the results of these post-hoc analyses, as well as the language experience hypothesis more generally, in more detail.

The basic idea behind the language experience hypothesis is that, because English has a tendency toward relatively phonetically fortis voiced stops in domain-initial positions, and lenis approximants in medial positions, English listeners should default to the hypothesis that voiced stops mark constituent boundaries and approximants do not. This strategy, if applied to the spirantization languages in our experiment, would succeed in uncovering the “words” almost immediately, whereas in the anti-spirantization language, it would substantially impair performance. On the other hand, because the evidence for a parallel voicing lenition-voicing pattern in English is quite weak, listeners would be unlikely to pursue such a parsing strategy. This is consistent with finding no significant difference between performance on the voicing and anti-voicing conditions.

The language experience hypothesis is attractive because it need not posit any special machinery, such as boundary-disruption constraints and inherent auditory disruption, to explain our data. Instead, our findings would be consistent with the existing literature (reviewed in Section 1.3) demonstrating that phonetic or phonological patterns attested

<sup>1</sup> A reviewer brings up the possibility that English speakers may not have been able to tell apart the prevoiced and short-lag stops used here *at all*. We think this is likely too strong a claim, because English speakers have been shown to use closure voicing effectively as a cue in intervocalic position (e.g. Lisker 1978), because most subjects performed above chance in both voicing conditions, and because there was no advantage for stimuli with fricatives, which English speakers should be more likely to discriminate effectively.

in a listener's native language can aid in the word segmentation task. In this case, our study's novel contribution would consist in its demonstration that segmental allophonic variation should be counted among the language-specific phonetic and phonological patterns that the perceptual system can capitalize on in order to optimize performance in the word segmentation paradigm. Under this view, the language experience hypothesis is independently motivated and seems to explain our data rather well, while the boundary-disruption hypothesis has trouble explaining the voicing data and is not as well established in previous literature. However, we think there are several reasons, both *a priori* and empirical, why the language experience hypothesis is not entirely adequate for explaining our data.

The first reason pertains to the mismatch between the domain of English lenition and the domain of lenition in our experiment. As discussed in Section 1.2, English flapping and spirantization are largely limited to contexts where the following vowel is unstressed. As a result, unlenited voiced stops preceding stressed or full vowels are not especially likely to coincide with morpheme, word, or phrase boundaries in English. Because the syllables in our experimental materials contain full vowels with identical duration, intensity, and  $f_0$  (non-contours), prior English experience would tend to dictate *against* positing constituent boundaries on the basis of consonant fortition in our stimuli. This clearly runs counter to our finding that relative consonant fortition (i.e. boundary disruption) was particularly helpful in the Spir1 and Spir2 conditions of our experiment.

An additional prediction of the language experience hypothesis is that a listener's likelihood of positing boundaries should track statistical asymmetries present in the language. As discussed in Section 3.2, one possible source of such asymmetries concerns place of articulation: flapping is far more frequent than spirantization in English, and spirantization of /g/ is more likely than spirantization of /b/ (Lavoie 2001; Walter 2007; Bouavichith & Davidson 2013). English experience thus favors constituents with medial unspirantized [b] over unspirantized [g], and favors both of these over unspirantized [d]. Conversely, medial approximant [ɣ] should be favored over [β] because it is more likely to be the output of spirantization in English. Predictions about initial segments are less straightforward, because given a lenition-domain-boundary, the probability of fortis stops at any place of articulation is fairly high and the probability of lenis approximants/taps is fairly low. However, as a reviewer points out, given that English [ð] is unattested in word-initial position outside the determiner system, English influence could be manifested in a bias against constituents beginning with that segment. As shown in Figure 4, there is little to no evidence for any of these patterns in our data.

More generally, it is also possible that targets in the spirantization conditions were more well-formed in terms of English lexical statistics than their anti-spirantization counterparts. For instance, a CELEX search reveals that the frequency of /b/ in word-initial prevocalic position relative to word-medial intervocalic position is greater than the comparable ratio for /v/ or /w/, regardless of whether type or token counts are used. So perhaps words with initial /b/ were globally considered more well-formed than those with initial labial continuants. However, this logic breaks down to a certain extent when we consider other places of articulation. For one thing, initial velar continuants are quite rare in English and probably unattested in the lexicon, yet the patterning of (simulated) velar consonants in this experiment is not appreciably different from other consonants.

The upshot of this discussion is that, while English experience could in principle explain why we observed a spirantization advantage in our experiment, the role of English experience would have to be limited to aggregate statistics over manners of articulation (e.g., stops are frequent in initial position and approximants are frequent in medial position), and would need to ignore generalizations about place of articulation. While such emphasis

on one set of features to the exclusion of another set is by no means impossible, it also has no obvious explanation.

A final possibility is that the spirantization pattern conferred an advantage in word segmentation because as native English speakers living in California, our subjects were likely to have accumulated some degree of exposure to spoken Spanish, and Spanish spirantization is both robust and relatively similar to the pattern used here (indeed, the acoustic targets for our stimuli were based on Madrileño Spanish recordings). A first step towards assessing this hypothesis is to look at results for subjects who reported extensive Spanish experience. There were six subjects in our study who reported being proficient Spanish speakers: three in the anti-spirantization condition, two in the Spir1 condition, and one in the Spir2 condition. While this is obviously a very small sample, these listeners do show a trend in the expected direction: this group displayed pooled accuracy of 88% in the spirantization conditions and 53% in the anti-spirantization condition, while the rest of the subjects displayed values of 83% and 66%, respectively. It does appear, then, that extensive previous experience with a similar spirantization pattern is associated with a larger effect on performance, consistent with the findings related to language-specific phonetic cues to word segmentation summarized in Section 1.3.

Could the results for the remaining subjects be explained by incidental exposure to Spanish? Clearly we cannot definitively rule this explanation out, but we would argue that it does present several non-negligible weaknesses, mainly pertaining to the voicing conditions. If this general subject population has had significant enough prior exposure to Spanish (despite not reporting any) to internalize Spanish sound patterns, and they tended to draw on this exposure to complete the word segmentation task, then listeners in the voicing conditions should have had no trouble distinguishing between the prevoiced and short-lag voiceless stops used in the voicing conditions. Further, several varieties of Peninsular, Canary Islands, and Caribbean Spanish display variable voicing of voiceless stops in intervocalic position (see Hualde et al. 2011 for a review), with rates varying between 35% and 65% by speaker and variety. We are not aware of any data on Mexican or Central American varieties, which are most likely to be spoken in California, but assuming that voicing lenition is at least a possibility in the putative incidental Spanish exposure of our subjects, that exposure should also have conferred an advantage in the voicing condition over the anti-voicing one. This is not what we found. Performance in both voicing conditions was roughly equal to the anti-spirantization condition, the latter of which should be strongly disfavored by Spanish experience. In summary, if prior incidental Spanish experience can be invoked to explain performance in the spirantization conditions, then we must also explain why such experience was *not* a significant factor in the voicing conditions.

Future work could, of course, test whether incidental Spanish exposure might explain our results by running a replication study with a population that is unlikely to have much exposure to Spanish at all. The spirantization study is currently being run with grade-school children in West Virginia, who are extremely unlikely to have been exposed to Spanish beyond educational television programs. If the advantage for spirantization persists in this follow-up study, it would reinforce the argument that Spanish exposure is not the driving factor behind the present results either.

#### **4.4 Conclusion**

This study finds that English speakers' performance on a word segmentation task is improved by spirantization patterns in an artificial language, but not by a voicing lenition pattern. Methodologically speaking, the study thus demonstrates that, in principle, the word segmentation paradigm can be fruitfully used to examine the notion of auditory-

perceptual *disruption*. There are several possible ways in which previous linguistic experience could have contributed to these results. We have given several examples, however, of more detailed predictions based on linguistic experience that do not seem to be borne out by the data. In light of this, we think it is worth considering that the results may be due to properties of auditory continuity and disruption inherent to the spirantization pattern, as predicted by the boundary-disruption approach to lenition.

To further assess the merits of the boundary-disruption approach, future work should test further manipulations of the voicing process as discussed above, and extend the segmentation paradigm to other types of lenition such as tapping, shortening, and glottalization. Another possibility could be to reproduce the spirantization experiment with listeners that have absolutely no demonstrable exposure to such a pattern. It is worth noting, however, that we are not aware of any study that has looked for evidence of spirantization in a language with voiced stops and failed to find it at least some of the time. We suspect this is not an accident: if spirantization is caused by such general and universal factors as articulatory ease and/or perceptual continuity, it will not be easy to find a language that entirely lacks spirantization. One might try a language with no voicing contrasts, and thus no voiced stops to putatively spirantize, but many languages with no voicing contrasts do have voiced stop allophones in at least some contexts. Recent research shows that even in Gurindji, a language with no voicing contrast, stops are often realized as voiced and spirantized intervocally (Ennever et al. 2017). The only practical way to examine language-specific versus language-independent influences on such artificial language tasks, therefore, may be to turn to more specific patterns attested in particular languages, as we've done here.

## Additional Files

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

- **ExplanationOfDataFiles.txt.** Explains the fields in the data files. DOI: <https://doi.org/10.5334/gjgl.443.s1>
- **DataX1.txt.** Data from the spirantization conditions, in tab-delimited .txt format. DOI: <https://doi.org/10.5334/gjgl.443.s2>
- **DataX2.txt.** Data from the voicing conditions, in tab-delimited .txt format. DOI: <https://doi.org/10.5334/gjgl.443.s3>

## Acknowledgements

The authors wish to thank Keith Johnson, Bob Ladd, Sharon Peperkamp, Alice Turk, and audiences at the LSA annual meeting and LabPhon for helpful comments and discussion.

## Competing Interests

The authors have no competing interests to declare.

## References

- Armstrong, Blair C., Ram Frost & Morten H. Christiansen. 2017. The long road of statistical learning research: Past, present and future. *Philosophical Transactions of the Royal Society B* 372. DOI: <https://doi.org/10.1098/rstb.2016.0047>
- Bagou, Odile, Cécile Fougeron & Uli Frauenfelder. 2002. Contribution of prosody to the segmentation and storage of “words” in the acquisition of a new mini-language. *Presented at Speech Prosody*. Aix-En-Provence.
- Barr, Dale, Roger Levy, Christoph Scheepers & Harry Tilly. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>

- Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
- Bauer, Laurie. 1988. What is lenition? *Journal of Linguistics* 24. 381–392. DOI: <https://doi.org/10.1017/S002222670001183X>
- Borgman, Donald. 1990. Sanuma. In Desmond Derbyshire & Geoff Pullum (eds.), *Handbook of Amazonian languages II*. 15–248. Berlin: Mouton de Gruyter.
- Bouavichith, Dominique. 2014. *Segmental and prosodic effects on intervocalic voiceless stop lenition in connected speech*. New York: New York University BA Thesis.
- Bouavichith, Dominique & Lisa Davidson. 2013. Segmental and prosodic effects on intervocalic voiced stop reduction in connected speech. *Phonetica* 70. 182–206. DOI: <https://doi.org/10.1159/000355635>
- Chong, Adam. 2011. Lenition in Gaalpu: an Optimality Theoretic analysis. *Australian Journal of Linguistics* 31. 473–490. DOI: <https://doi.org/10.1080/07268602.2011.625601>
- Cohen Priva, Uriel. 2017. Informativity and the actuation of lenition. *Language* 93(3). 569–597. DOI: <https://doi.org/10.1353/lan.2017.0037>
- Cooper, Franklin, Pierre Delattre, Alvin Liberman, John Borst & Louis Gerstman. 1952. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America* 24. 597–606. DOI: <https://doi.org/10.1121/1.1906940>
- De Jong, Kenneth. 2011. Flapping in American English. In Marc Oostendorp, Colin Ewen & Elizabeth Hume (eds.), *The Blackwell companion to phonology*, 2711–2729. Malden, MA: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781444335262.wbctp0113>
- Donegan, Patricia & David Stampe. 1979. The study of Natural Phonology. In Daniel Dinnsen (ed.), *Current approaches to phonological theory*, 126–173. Bloomington, IN: Indiana University Press.
- Ennever, Thomas, Felicity Meakins & Erich Round. 2017. A replicable acoustic measure of lenition and the nature of variability in Gurindji stops. *Laboratory Phonology* 8(1). 1–32. DOI: <https://doi.org/10.5334/labphon.18>
- Fisher, Cynthia & Hisayo Tokura. 1996. Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development* 67(6). 3192–3218. DOI: <https://doi.org/10.2307/1131774>
- Frank, Michael, Sharon Goldwater, Thomas Griffiths & Joshua Tenenbaum. 2010. Modeling human performance in statistical word segmentation. *Cognition* 117. 107–125. DOI: <https://doi.org/10.1016/j.cognition.2010.07.005>
- Gleitman, Lila R. & Eric Wanner. 1982. Language acquisition: the state of the art. In Eric Wanner & Lila Gleitman (eds.), *Language acquisition: The state of the art*, 3–48. Cambridge: Cambridge University Press.
- Gurevich, Naomi. 2003. *Functional constraints on phonetically conditioned sound changes*. Urbana-Champaign, IL: University of Illinois dissertation.
- Harris, John. 2003. Grammar-internal and grammar-external assimilation. In Maria Josep Solé, Daniel Recasens & Joaquín Romero (eds.), *Proceedings of the 15<sup>th</sup> International Congress of Phonetic Sciences*, 281–294. Barcelona: Futurgraphic.
- Haugen, Einar. 1938. Notes on “voiced t” in American English. *Dialect Notes* 6. 627–634.
- Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago, IL: University of Chicago Press.
- Hayes, John R. & Herbert H. Clark. 1970. Experiments in the segmentation of an artificial speech analog. *Cognition and the Development of Language*, 221–234.

- Honeybone, Patrick. 2008. Lenition, weakening, and consonantal strength: Tracing concepts through the history of phonology. In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition*, 9–93. Berlin: Mouton de Gruyter.
- Hualde, José, Miquel Simonet & Marianna Nadeu. 2011. Consonant lenition and phonological recategorization. *Laboratory Phonology* 2. 301–329. DOI: <https://doi.org/10.1515/labphon.2011.011>
- Hunyadi, László. 2006. Grouping, the cognitive basis of recursion in language. *Argumentum* 2. 67–114.
- Hyman, Larry. 1972. *A phonological study of Fe'fe'-Bamileke*. Los Angeles, CA: University of California dissertation.
- Jeon, Hae-Sung & Francis Nolan. 2013. The role of pitch and timing cues in the perception of phrasal grouping in Seoul Korean. *Journal of the Acoustical Society of America* 133(5). 3039–3049. DOI: <https://doi.org/10.1121/1.4798663>
- Johnson, Elizabeth & Peter Jusczyk. 2001. Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory & Language* 44(4). 548–567. DOI: <https://doi.org/10.1006/jmla.2000.2755>
- Jusczyk, Peter, Elizabeth Hohne & Angela Bauman. 1999. Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics* 61. 1465–1476. DOI: <https://doi.org/10.3758/BF03213111>
- Kahn, Daniel. 1976. *Syllable-based generalizations in English phonology*. Cambridge, MA: MIT dissertation.
- Katz, Jonah. 2016. Lenition, perception, and neutralisation. *Phonology* 33(1). 43–85. DOI: <https://doi.org/10.1017/S0952675716000038>
- Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82. 536–574. DOI: <https://doi.org/10.1353/lan.2006.0146>
- Keating, Patricia. 2006. Phonetic encoding of prosodic structure. In Jonathan Harrington & Marija Tabain (eds.), *Speech production: Models, phonetic processes, and techniques*, 167–186. New York: Psychology Press.
- Kentner, Gerrit & Caroline Féry. 2013. A new approach to prosodic grouping. *The Linguistic Review* 30(2). 277–311. DOI: <https://doi.org/10.1515/tlr-2013-0009>
- Kim, Sahyang. 2004. *The role of prosodic phrasing in Korean word segmentation*. Los Angeles, CA: University of California dissertation.
- Kingston, John. 2008. Lenition. In Laura Colantoni & Jeffrey Steele (eds.), *Proceedings of the 3<sup>rd</sup> Conference on Laboratory Approaches to Spanish Phonology*, 1–31. Somerville: Cascadilla.
- Kirchner, Robert. 1998. *An effort-based approach to consonant lenition*. Los Angeles, CA: University of California dissertation.
- Klatt, Dennis & Laura Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87. 820–857. DOI: <https://doi.org/10.1121/1.398894>
- Ladd, D. Robert & Jim Scobbie. 2003. External sandhi as gestural overlap? Counter-evidence from Sardinian. In John Local, Richard Ogden & Rosalind Temple (eds.), *Papers in Laboratory Phonology VI*, 164–182. Cambridge, UK: Cambridge University Press.
- Lavoie, Lisa. 2001. *Consonant strength: Phonological patterns and phonetic manifestations*. New York: Garland.
- Lisker, Leigh. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33. 42–49. DOI: <https://doi.org/10.2307/410949>

- Lisker, Leigh. 1978. Rapid vs. ravid: A catalogue of acoustic features that may cue the distinction. In *Haskins Laboratories status report on speech research, SR-54*, 127–132. New Haven, CT: Haskins Laboratories.
- Lisker, Leigh & Arthur Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–422. DOI: <https://doi.org/10.1080/00437956.1964.11659830>
- Luce, Paul A. & Jan Charles-Luce. 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. DOI: <https://doi.org/10.1121/1.392651>
- Mareschal, Denis & Robert M. French. 2017. TRACX2: A connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philosophical Transactions of the Royal Society B* 372. DOI: <https://doi.org/10.1098/rstb.2016.0057>
- Mathôt, Sebastian, Daniel Schreij & Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44(2). 314–324. DOI: <https://doi.org/10.3758/s13428-011-0168-7>
- Nakatani, Lloyd & Kathleen Dukes. 1977. Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America* 62. 714–719. DOI: <https://doi.org/10.1121/1.381583>
- Parker, Steve. 2002. *Quantifying the sonority hierarchy*. Amherst, MA: UMass dissertation.
- Romero, Joaquín. 1996. *Gestural organization in Spanish: An experimental study of spirantization and aspiration*. Storrs, CT: University of Connecticut dissertation.
- Saffran, Jenny, Elissa Newport & Richard Aslin. 1996b. Word segmentation: The role of distributional cues. *Journal of Memory & Language* 35(4). 606–621. DOI: <https://doi.org/10.1006/jmla.1996.0032>
- Saffran, Jenny, Richard Aslin & Elissa Newport. 1996a. Statistical learning by 8-month-old infants. *Science* 274. 1926–1928. DOI: <https://doi.org/10.1126/science.274.5294.1926>
- Scharf, Donald. 1960. Distinctiveness of “voiced t” words. *American Speech* 35. 105–109. DOI: <https://doi.org/10.2307/453733>
- Ségéral, Philippe & Tobias Scheer. 2008. Positional factors in lenition and fortition. In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition*, 131–172. Berlin: Mouton de Gruyter.
- Sheldon, D. R. 1973. A short experimental investigation of the phonological view of the writer – rider contrast in U.S. English. *Journal of Phonetics* 1. 339–346.
- Smith, Jennifer. 2008. Markedness, faithfulness, positions, and contexts: Lenition and fortition in Optimality Theory. In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition*, 519–560. Berlin: Mouton de Gruyter.
- Szigetvári, Péter. 2008. What and where? In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition*, 93–130. Berlin: Mouton de Gruyter.
- Trager, George & Henry Smith. 1951. *An outline of English structure*. Norman, OK: Battenburg Press.
- Turk, Alice. 1992. The American English flapping rule and the effect of stress on stop consonant durations. *Working Papers of the Cornell Phonetics Laboratory* 7. 103–133.
- Tyler, Michael D. & Anne Cutler. 2009. Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America* 126(1). 367–376. DOI: <https://doi.org/10.1121/1.3129127>
- Umeda, Noriko. 1977. Consonant duration in American English. *Journal of the Acoustical Society of America* 61. 846–858. DOI: <https://doi.org/10.1121/1.381374>
- Walter, Mary Ann. 2007. *Repetition avoidance in human language*. Cambridge, MA: MIT dissertation.

- Warner, Natasha & Benjamin Tucker. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *Journal of the Acoustical Society of America* 130(3). 1606–1617. DOI: <https://doi.org/10.1121/1.3621306>
- Wertheimer, Max. 1938 [English translation of 1923 essay]. Laws of organization in perceptual forms. In Willis Ellis (ed.), *A source book of Gestalt psychology*, 71–88. London: Routledge.
- Wightman, Colin, Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3). 1707–1717. DOI: <https://doi.org/10.1121/1.402450>
- Zue, Victor & Martha Laferriere. 1979. Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America* 66. 1039–1050. DOI: <https://doi.org/10.1121/1.383323>

**How to cite this article:** Katz, Jonah and Melinda Fricke. 2018. Auditory disruption improves word segmentation: A functional basis for lenition phenomena. *Glossa: a journal of general linguistics* 3(1): 38.1–25, DOI: <https://doi.org/10.5334/gjgl.443>

**Submitted:** 02 June 2017    **Accepted:** 09 January 2018    **Published:** 23 March 2018

**Copyright:** © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press.

