The status of the word-final domain and linguistically-constrained language learning cues

Megan Rouch
The status of the word-final domain and linguistically-constrained language learning cues

A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor of Arts in Linguistics from the College of William and Mary

by

Megan Rouch

Accepted for Honors

Dr. Anya Lunden, Director

Dr. Kaitlyn Harrigan

Dr. Daniel Parker

Dr. Brian Castleberry

Williamsburg, VA
April 22, 2019
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgments</td>
<td>pg. 2</td>
</tr>
<tr>
<td>Abstract</td>
<td>pg. 3</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>pg. 4</td>
</tr>
<tr>
<td>2. Word-final phenomena/phonology</td>
<td>pg. 9</td>
</tr>
<tr>
<td>3. Statistical learning</td>
<td>pg. 23</td>
</tr>
<tr>
<td>4. Methods</td>
<td>pg. 27</td>
</tr>
<tr>
<td>5. Results</td>
<td>pg. 30</td>
</tr>
<tr>
<td>6. Conclusion</td>
<td>pg. 36</td>
</tr>
<tr>
<td>References</td>
<td>pg. 41</td>
</tr>
</tbody>
</table>

Rouch 1
Acknowledgments

I would like to thank Professors Anya Lunden and Kate Harrigan for their support, help, advice, and invaluable guidance during the entirety of the last year. I would not have had any idea how to lift the car, or what kind of car to lift, without you, and not only did you help me with that, but also helped me to remain fond and excited throughout. I would also like to thank the other members of my thesis committee, Professors Dan Parker and Brian Castleberry, for their support and clarifying questions throughout the researching and writing process. I thank the Computational and Experimental Linguistics Lab at the College of William and Mary for their help and suggestions during the development of the thesis and presentation along the way. I would like to thank Professor Felix Wang from University of Nevada Las Vegas for lending me the current draft of his paper in prep. Finally I want to thank Becki, my roommates, all my other friends, and my parents for their support in the last year; you kept me going through the long nights, reminded me to take breaks, and also reminded me that my results were just as important as the literature review.
Abstract

Past experimental and theoretical research have argued that phonological devoicing as a word-final phenomenon only has inherent phonetic motivation to exist utterance-finally and that its presence at lower domain levels is due to analogy (Hock 1991, 1999; Hualde and Eager 2016). This evinces the idea that the word is not a domain tied to phonology inherently. Juxtaposing, final devoicing (FD) is found phonetically at various prosodic domain levels. Final lengthening (FL) is also found at various prosodic levels, though is explained as only having a push to exist utterance-finally through a process similar to the “slowing down of a machine” or “tempo change...local to the final gesture in the phrase” (Klatt 1976 and Edwards et al. 1991 respectively). The present artificial language learning (ALL) experiment gives evidence that these cues may have an inherent reason for existing at lower domain levels. Participants were exposed to 7 minutes of AL stimulus in one of three conditions: transitional probabilities (TP) or the likelihood that syllables will occur in sequence together, final lengthening (FL) or greater duration of final segments, and final devoicing (FD) or a trailing off of voicing in word-final segments. They were then tested on their ability to segment the AL words out of the speech stream. The results show there is a perceptual effect to having phonetic FL at the right edge of words for listeners, in that it increases novel word segmentation ability above the level of only TPs. The same effect was not found for FD. The FL results indicate that phenomena in the word-final area may be used for perceptual reasons, providing evidence speakers use phonetic effects at the word-level.
1. Introduction

Phonological final devoicing (FD hereafter) is the process by which voiced sounds at the ends of words or syllables categorically devoice, and is attested in many languages such as German and Dutch (Hock 1999). The prevailing explanation for phonological devoicing is that it may occur word-finally only by analogy or secondary extension (an interchangeable term with analogy which has no discernible difference) from phonetic devoicing in utterance-final position. This is said to be because it is only utterance-finally that this phenomenon has an underlying motivation to occur inherently: approaching a pause, where though the vocal folds’ movement towards silence in pause is although not exactly the same as voicelessness, best resembles voicelessness (Hock 1976, 1991, 1999; Hualde 2013; Hualde and Eager 2016; Keating 1988; Myers and Padgett 2014). Devoicing occurring at the right edge of the word (word-finally), instead of “finality” (Hock’s term for utterance-finally), has been seen as problematic, due to the fact that this devoicing can even occur intervocally, wherein one might expect the opposite, that voicing be sustained and unhindered (Hock 1991, 1999; Hualde and Eager 2016). While Hock and Keating discuss the origin as being utterance-finally, Hualde and Eager cite the phrase-final domain as an origin of analogy (Hock 1991, 1999; Hualde and Eager 2016; Keating 1988). Additionally, this is unreconciled with studies that do show non-phonological, i.e. phonetic, word-final devoicing as seeming to be inherent in production at least for specific languages (e.g. Nakai et al. 2009 in Northern Finnish). Only phonological devoicing word-finally to my knowledge has been theoretically explained, with phonetic devoicing at the word-level going unmentioned in articles which address the former.

Phonetic FD is reported to occur at the various prosodic levels (Gordon and Munro 2007; Hock 1976, 1996, 1999; Johnson and Martin 2001 for utterance-final phonetic devoicing in Creek; Nakai et al. 2009 for mid-phrase at the word-level, and phrase-final breathiness in Northern Finnish). Phonetic FD consists of voiced sounds at the ends of words trailing off into “breathiness” or “breathy voice” approaching voicelessness at the phonetic level. There is an increase based on domain level, with breathiness being found at lower levels, but generally increasing in amount at higher levels (Gordon and Munro 2007). This final devoicing has also been cited as a potential cue to prosodic boundaries (Gordon and Munro 2007).

Secondary extension, or extension by analogy, of phonetically-motivated phonological devoicing from utterance- or phrase-final position to lower domain levels such as the word- or syllable-level, is hypothesized as happening in two or three stages. First the consonants of a language, which have phonetic devoicing at the utterance-level, must be categorically,
phonologically encoded as devoicing utterance- and/or phrase-finally, and from there this pronunciation is also encoded phrase-medially, so that all obstruents are now devoiced in all final positions (Hualde and Eager 2016). It is not clear whether this is proposed to be done by each individual word, encoding the right edge of words as being devoiced in phrase-final position, and then that encoding is applied in all other positions, or whether uniformly across the domains, first phrase-final and then down to word-final. In either scenario, only phonological devoicing is discussed; though phonetic devoicing is also found at these levels, its origin is unclear. This poses the question of whether phonetic phenomena may also be extended by analogy from the utterance- or phrase-final level down to lower domains, an idea which may be incongruous with phonetics traditionally, since phonetic phenomena are generally defined by their natural occurrence as a by-product of speech production, or inherent presence in the speech stream. This would entail generalization even down to syllable-final devoicing in some German dialects (Hock 1999). Other proposals of secondary extension such as in Hock 1999 are even less fully described by what means this process may occur.

Another phenomenon similar to FD which is attested to occur at many domain levels is final lengthening (henceforth FL). This is an established phenomenon cross-linguistically wherein sounds at the ends of prosodic domains are lengthened (Crystal and House 1988; Johnson and Martin 2001; Lunden 2017; Nakai et al. 2009). FL is highly cited as behaving in a similar way to phonetic devoicing, occurring gradationally at many prosodic domain levels, although it does not have a prevalent phonological counterpart in the same way FD does¹. This is assumed to be a universal tendency, although it can be suppressed in certain languages for the maintenance of contrastive vowel length (Gordon et al. 2010; Gordon and Munro 2007). Wightman et al. (1992) have shown that FL occurs in greater amounts the higher the hierarchical prosodic domain level (utterance being the largest, word the smallest). However, Cambier-Langeveld (1997) has shown Dutch in particular that this is not always the case, as she found no significant difference between some prosodic levels’ FL amounts. Similarly, while Wightman et al. (1992) found a distinction between four different domain levels lower than the intonational phrase based on lengthening, Fougeron and Keating (1997) found a distinction between only two in American English both with real English words and then repetitive syllables which stood in for the syllables of those real words. Additionally, the ordering of amounts of

¹ There are some languages which have FL on final vowels as opposed to the usual shortening, which could be considered a phonologization of FL, e.g. Tigre (Palmer 1962) and Scottish English (Wells 1982; Scobbie et al. 1999) and American Sign Language (Perlmutter 1993), however this is not as ubiquitous or directly analogous to phonetic FL as phonological FD is to phonetic FD; it is a more tenuous relationship.
lengthening found by Fougeron and Keating (1997) is counterintuitive, in that some lower domain levels have more lengthening than levels above them, which does not fit the predicted hierarchical relationship.

This universal phenomenon of FL is thought to have its origin in the acoustics, phonetics, and mechanics of the production of sound. The assumption is that the lengthening phenomenon is similar to the slowing down of a machine to a stop as a speaker comes up to a pause in their speech, which in turn may signal the end of the speech at that point to the listener phrase- or utterance-finally (Klatt 1976). Edwards et al. (1991) note that in the kinematics of FL, final segments are longer because of a larger and faster initial opening gesture which drags on. Originally, this signal was linked to syntactic boundaries (Klatt 1975), it has been shown to actually be linked to prosodic boundaries (Cambier-Langeveld 1997). In Edwards et al. (1991)’s study, participants said words in different phrases of sentences at varying production speeds, and they note durations of vowels were statistically significantly longer phrase-finally for all subjects, including even schwa and reduced vowels due to production speed. Although there is lengthening at lower prosodic levels including word-finally, it has not been explained how it might exist there, if its origin is in the utterance-final area. This utterance-final motivation rings very similar to the hypothesized cause of FD. Theories have gone unaddressed to my knowledge as to how this phenomenon surfaces at lower prosodic domains wherein there is no immediate approach to pause, however I suggest that it is either inherently present at all prosodic levels as could be indicated by perceptual usefulness.

FL has also been tested and found to be facilitative for users of some languages in parsing novel artificial language words, or parsing phrases and words (Elordieta et al. 2005 for intonational phrases in Spanish and European Portuguese). Artificial language learning (ALL) studies have been helpful in assessing the perceptual qualities or lack thereof of the relevant prosodic cues, and therefore could be used to explore possible evidence that these cues exist inherently throughout the speech stream at lower domain levels than the utterance (e.g. at the word-level). They thus also give evidence that the word-boundary is a domain in its own right.

The ALL paradigm is one approach to investigating how humans (adults as well as infants) may parse words from a continuous speech stream which has no pauses to indicate word boundaries. The experimenter creates a small group of novel words for the AL, concatenates them into a continuous speech stream, and then exposes the participant to that speech stream for a set amount of time (typically 2 minutes for infants and 7 for adults). They are then tested on their knowledge of the AL words or whatever independent variable the study
is focusing on, through programmed quizzes for adults, or preferential looking or the head-turn preference paradigm for infants. ALL experiments are meant to imitate the natural language language-learning process in an adequately natural way while manipulating variables, allowing for language acquisition and different facets of psycholinguistics to be studied.

One skill that language-users and -learners have at their disposal is statistics-taking using transitional probability (TP), wherein syllables within words always occur in sequence and therefore have a higher TP or probability of occurring in that sequence than syllables which are not in the same word, such as which never co-occur or only occur across a word boundary. Both infants and adults have been shown to be capable of parsing novel words out of an AL speech stream using TPs alone (Saffran et al. 1996b; replicated by Tyler and Cutler 2009), although when the words are of different syllable lengths, infants’ statistic-taking abilities are disrupted (Johnson and Jusczyk 2003; Johnson and Tyler 2010). However, adults are still capable of segmentation of novel words of different syllable lengths using only TPs (Saffran et al 1996b; replicated by Tyler and Cutler 2009).

Words have been further investigated with phonetic and prosodic markings in other ALL experiments, such as the one done by Kim et al. (2012), in which speakers of Korean and Dutch were tested on their ability to segment words out of an AL speech stream with the addition of prosodic cues in four conditions besides TPs: initial syllable F0 (pitch) rise, final syllable F0 rise, FL, and FL final syllable pitch rise. They showed that speakers of both languages were able to use FL at a higher than chance level, as well as at a higher level than the purely TPs condition, to encode word boundaries and separate novel trisyllabic words. Tyler and Cutler (2009) showed that a right-edge lengthening cue did the same for speakers of English, Dutch, and French with both trisyllabic and four syllable novel words present in the speech stream. They had three conditions, TP only, and left-edge cue and right-edge cue, wherein they would have initial-lengthening or FL respectively. It was found that regardless of language background, participants were capable in the right-edge cue condition of using the FL at a higher-than-chance rate, and also a rate higher than just the TP condition, to separate the words better than either other condition. It is important to note that the right-edge FL cue improved their performance at word segmentation on average by about 13%. Saffran et al. (1996b) showed that for adult speakers of English, a FL cue improves word segmentation performance of novel AL words above the level of TPs alone. Langus et al. (2012) demonstrate that FL and intonational pitch arc declination prosodic cues assist Italian speakers in parsing AL phrases and utterances respectively, and show that listeners are capable of perceiving different prosodic
cues as cuing different prosodic levels. Ordin et al. (2017) have additionally demonstrated that for Basque and German speakers, an FL prosodic cue improves novel word segmentation beyond only TPs. After all of this study, it is interesting to note that to my knowledge, FD has not been tested in word segmentation the way that FL has.

Though lengthening itself is regarded as a universal phenomenon, it has been asserted that the production and especially the perception of FL, specifically the perception of FL as a cue to word segmentation, may be language specific. Cambier-Langeveld et al. (1997) demonstrates that for Dutch speakers, word acceptability as sounding naturally pronounced is not contingent on a proportionally larger amount of lengthening as domain level increases. D’Imperio et al. (2005) show that lengthening is a potential phrasal boundary marker in Spanish. Nakai et al. (2009) show that in Northern Finnish, FL occurs but in altered amounts in order to maintain the short- versus long-vowel contrast. Turk and Shattuck-Hufnagel (2007) show that final consonants of the final rhyme of a final word are typically only slightly, though reliably, longer than the nucleus of that final rhyme, and White and Turk (2010) demonstrate that stress patterns in English have a large effect on how lengthening word-finally is realized. Finally, in contrast to the results for German and English speakers, and approaching for Basque speakers, Ordin et al. (2017)’s study showed that FL is not facilitative or impeding for Italian speakers, though lengthening on the middle, penultimate, syllable facilitated, and initial-lengthening disrupted, their ability to parse novel AL words. They postulate that this has to do with the stress system in Italian, in that the FL cue may be universally present, however when other cues to word boundaries which are language-specific, such as penultimate lengthening due to stress in Italian, are present, these overrule the inherent and universally-present FL cue (Ordin et al. 2017).

Despite the fact that FL and FD have been related to the utterance-final domain and discussed as being linked to some single underlying cause (e.g. approach to a pause) but also occur at the right edge of words, the two phenomena have not been considered as being linked, with the exception of a proposal by Blevins (2004). Blevins (2004) hypothesizes that both phonetic devoicing phrase-finally and thusly phonological might be a by-product of FL. This could be because the supralaryngeal pressure during a stop makes voicing unsustainable thus resulting in devoicing at the end of the sound, or elongation of a voiced stop may cause speakers to perceive it as voiceless and then encode it as such. Even after this connection between the two, it has not been reconciled how these phonetic phenomena are found at lower domain levels than even phrase-finally and may act as perceptually salient cues for language users, and yet
frequently are proposed to only have a phonetic/acoustic basis to occur utterance-finally, so may only appear at lower domain levels phonologically through analogy or secondary extension.

In this thesis I will assess potential evidence about whether devoicing and lengthening are related and what their origins are, whether those be organic to wherever they occur, or utterance-finally followed by analogy down to lower prosodic domains. The experiments carried out in this thesis seek to provide evidence supporting the idea that word-final phenomena such as FL or FD actually originate independently at the word-level, as opposed to through analogy. If speakers can use FL and FD in an artificial language learning experiment at a significantly higher level in the word-learning task than those exposed to the version with only TPs, then this would indicate an inherent helpfulness, and thus that since they are there already they are then used, highlighting inherent acoustic uniqueness or phonetic-pushes and perceptual relevance of the prosodic word boundary. Speakers may expect the existence of these phenomena word-finally, which would seem unlikely to be the case if the cues were not inherent at lower domains in the speech stream, and only existed there after analogy from higher up.

2. Word-final phenomena/phonology

2.1 Final devoicing

Final devoicing (FD), either phonetic or phonological, has been shown or reported to exist at many different domain levels cross-linguistically (Hock 1999 for utterance-final, word-final, and syllable-final; Nakai et al. 2009 and Nakai 2013 for word-final/phrase-medial and phrase-final), except for, again and like FL, where suppressed language-specified for contrastive-vowel length maintenance. In its phonological form, it is a categorical lack of voicing; however, when there is phonetic devoicing present at various domain levels, it is more gradientially varied. Voicing at the end of a word trails off into an almost-voiclessness, or “breathy voice,” as shown in figures 1 and 2 from Nakai et al. (2009)². The breathy voice here is occurring at both utterance-final and phrase-medial (word-final) levels, and is affecting the final vowel of the word/syllable.

---

In figure 1 of utterance-final FD, we see the voicing trail off to a definite, though brief, voicelessness leading into the silence following. However, in figure 2, the voicing in the spectrogram below the waveform fades into nonexistence by the very end of the vowel, with breathy voice in-between, at the word-level phrase-medially.

We see more examples of phrase-medial breathy voice in the following figures 3 and 4 from Lunden (2012). 353 four-syllable nonsense words from recordings done by Lunden (2012) were combed through, and all final devoicing/”breathy voice” demarcated in a Praat TextGrid, then extracted with a Script and statistics and analysis run on that extracted data in SPSS. These recordings from Lunden (2012) were done as part of a study in which four-syllable nonce words were pronounced with antepenultimate stress, phrase-medially in both a question and an answer in order to get the most naturalistic pronunciation possible (e.g. Q: Which bafasada did your brother examine? A: My brother examined the bafasada that was particularly dirty). Speakers were instructed to phrase the relative clause with the nonsense word, and pronunciations which did not fit this were excluded.
A summary of the breathy voice present in the data examples 3 and 4 come from is shown in figure 5, wherein the three vowels words could have ended in are present on the x-axis, and the percentage of voicelessness or breathiness present in the final vowel is on the y-axis. What this summary and respective analysis of the data summarized here shows is that 19 out of 22 participants and on average 15.4% of total words had over 10 milliseconds of voicelessness or “breathy voice” present in the final vowels of their naturalistic nonce word pronunciations.
Blevins (2004)’s theory, opposite the prevailing theory of analogy from utterance-level phonetic-push FD caused by “assimilation” to silence (Hock 1991, 1999; Hualde and Eager 2016; Keating 1988; Lieberman 1967; Myers and Padgett 2014), proposing basically that the longer a sound is lengthened coming up to a prosodic boundary, the harder it is going to be to hold out voicing on that sound, and therefore more devoicing at the end of that lengthened segment will occur. When the vocal folds are set up in the formation that would allow voicing, just after closure of an obstruent, voicing must stop because of pressure against the laryngeal area after the oral closure if no movements are made tract-externally to create more space and allow for further vocal fold vibration (e.g. relaxing the soft tissue of the oral cavity, bringing the tongue root forward, or lowering the jaw). In a way this is still a kind of extension, from phonetic utterance- or phrase-final devoicing, to phonologically encoded word-final devoicing, although again, and notably, the presence of phonetic devoicing at all prosodic domain levels often goes unaddressed, and the presence of FL at multiple domain levels is likewise unaddressed as potential push for that FD at those prosodic levels other than utterance-finally.

The connection or correlated relationship made between FL and FD proposed by Blevins (2004) could be carried further, and I suggest that since FL is present at many different domain levels, there is this natural push for FD at those domain levels because of the FL present there.
The connection Blevins (2004) proposes has yet to be tested experimentally, and though FL and FD clearly have similarities, the phenomena have typically not been discussed together outside of this particular claim or an acknowledgment by Nakai et al. (2009) and Nakai (2013), or at lower levels than utterance-finally.

Devoicing has also been hypothesized as another type of domain-final weakening (Hock 1999). It is worth noting that this kind of weakening is antithetical to the direction of weakening found word-internally, namely voicing (the opposite of devoicing or breathiness) (Hock 1976a 138-139 with references, 1999). Likewise, devoicing as weakening domain-finally is different than its status word-internally, where it would be strengthening. However, Hock (1999) resolves this confound by proposing that weakening can be defined differently depending on the domain which it is affecting. While we find these two types of weakening frequently cross-linguistically, it is nevertheless somewhat mysterious why one weakening might entail devoicing word-finally instead of voicing, where presumably the confound of intervocalic pressure or general pressure to maintain voicing may still exist, as word-final segments often occur between sonorant, vocalized sounds, if one assumes a word-initial sonorant in the onset of the following word. This weakening is juxtapositional to the other. Presumably, either there is a different phonetic push word-finally than word-medially, provoking the idea that either there is something about the word-final position that gives this phonetic push for devoicing to be inherently present, or it must get there by analogy; either explanation would resolve the conflict of opposing directions for these two types of weakening. Extension of Blevins (2004) seems to be coherent with the first idea, and prevailing analogy theories with the second. If there were a phonetic push word-finally, then that would explain the inherent existence of FD at various domain levels, in which case the source of FD at various lower prosodic levels is phonetically-motivated “weakening.” If there is truly not a phonetic push, then it must be getting there through phonological encoding utterance-finally which then analogizes to the word-level.

Thinking through the theory of analogy, it seems it would be hard to tell whether devoicing is phonetically conditioned at the utterance-final level and transfers to lower boundaries, or if it is present at all domain levels proportionally increasing as it rises in hierarchy height, and therefore of course is present most strongly utterance-finally. It would also seem implausible that the edges of words were simultaneously not a place that phonology cropped up naturally but at the same time could be targeted by it for phonological FD and other word-final phenomena to be analogized to.
2.2 Final lengthening

Final lengthening (FL) is a universal phonetic phenomenon which occurs at various domain levels, with the amount often discussed as becoming greater at each higher domain level (Cambier-Langeveld 1997; Klatt 1975; Wightman et al. 1992) except for where suppressed in certain languages. Figures 6-9 show spectrograms of four-syllable CV nonce words from the same study as mentioned in 2.1, in phrase-medial position as read by native English speakers with antepenultimate stress (Lunden 2012). The spectrograms trace the effect of word-internal position on the syllable “fa” [fa]/[fa] (note though that in each position of each word, the vowels are undergoing the same durational changes depending on word-internal position; it is just convenient to trace one syllable in particular). We see that the [a] vowel of the initial CV syllable is one fifth the length of the final vowel. Additionally, the penultimate syllable’s vowel is half the length of the final syllable’s. Even the vowel of the antepenultimate and stressed syllable of figure 7 is shorter than the vowel in the final syllable. Each figure’s [fa] vowel length is noted in milliseconds below the spectrogram and waveform.

Figure 6. Initial: [fadəsbə] 45 ms.
Figure 7. Antepenultimate, with stress: [dəfəbəsə] 136 ms.

Figure 8. Penultimate: [səbəfədə] 95 ms.

Figure 9. Final: [sədəbəfə] 250 ms. Lunden Recordings (2012)
This effect is summarized in the following box plot graph, figure 10, of the recordings from which these spectrograms were drawn. The final syllable of a word is consistently going to remain notably long due to FL. In the box plot, one quarter of the data per each syllable is contained within each quadrant and/or line section per box, with the median data point lying on the middle line within each box. As before in section 2.1, the box plot includes data represents a total of 353 nonce word pronunciations spoken by 22 participants.

![Box plot graph](image)

**Figure 10.**

Lunden Recordings (2012)

We see here that, though the antepenultimate syllable was pronounced with stress, its vowel can still on average be shorter than the final vowel which is being affected by FL. The final vowel is consistently one of the longer, if not the longest, vowel present in the word due to this effect.

Figure 11 (from Klatt 1976)\(^3\) demonstrates an instance of phrase-final lengthening on the word “bad” taken as its own noun phrase (as in “the bad,” like “the rich” or “the poor”) in the first spectrogram, compared with “bad” taken as an adjective and only part of a larger noun phrase in the second spectrogram. We can see that in the phrase-final “bad,” lengthening causes

an increase of the length of the word by 120 milliseconds (to 300 total milliseconds) as compared to the mid-phrase “bad” which is 180 milliseconds.

Figure 11. Klatt (1976)

Here we have seen examples of FL both at the word-level and at the phrase-level, as is the focus of discussion of FL in this paper. This phenomenon at the word-level is not discussed as often, with much of the prior literature has focused on phrase- or utterance-final lengthening.

One of the foundational studies on FL was carried out by Wightman et al. (1992), in which they segmented a corpus of 35 pairs of phonetically similar although syntactically ambiguous sentences read aloud by four professional FM radio news presenters, created by Price et al. (1991). They then measured normalized duration, the duration of a segment as a standard deviation amount away from the mean of the phone in that segment, and compared these durations at the coded prosodic levels of the corpus (Wightman et al. 1992). They examined both FL as well as initial-lengthening as possible candidates for analysis; while
initial-lengthening is a phenomenon which occurs, it is not as punctuated, likely to occur, or perceptually important, as FL (Wightman et al. 1992). The authors find that FL aligns with statistically significant differentiation of four different domain levels lower than intonational phrase. They suggest that, at least in production, durational cues can distinguish lower level domain boundaries. They found that FL is only measurable in the rhyme at the end of a domain/boundary; in other words, in the vowel and any coda consonants. In Dutch, FL has been reported as affecting only the rhyme of the final syllable of a word, unless the rhyme of that syllable only consists of a schwa, in which case lengthening begins in the rhyme of the penultimate syllable (Cambier-Langeveld et al. 1997).

On the other hand, while Wightman et al. (1992) do find evidence of the existence of word-initial lengthening, they note that word-initial lengthening is not correlated to boundary size. While it is present, it may not serve the same perceptual functions as FL, and is not as salient an effect, as initial-lengthening affects only the initial consonant and not the vowel (Keating et al. 2004). Figure 6 demonstrates an example of this, in which the vowel of the initial syllable is not affected by lengthening, and is often below any amount that could be taken as an “average length” of respective vowels. Fougeron (1998) as cited in Keating et al. (2004) entertains the theoretical that initial-lengthening may be for a perceptual reason, but concludes it is not and is more a by-product of articulation in the initial position.

Though Wightman et al. (1992) found FL present and statistically predictive of domain level, they did not test the perceptual qualities of it; they do not have data regarding the actual perception of these durational cues, and whether this lengthening in production could be either an artifact of mechanisms of production or under direct control of the speaker in an effort to differentiate those prosodic levels for listeners. Prior to this, Klatt (1975, 1976) proposed that lengthening seems to be a necessity of perception which is increased hierarchically further at each higher domain level. This is a pattern found throughout other phenomena such as child-directed speech and sign language (Koponen and Lacerda 2003 and Grosjean 1979 respectively), and even music (Lindblom 1978). Klatt (1975) proposes both the possibility that due to a necessity of perception, FL is present in ever-increasing amounts, or the other possibility, that it is a by-product of acoustic sound production and a slowing down. He, and others, conclude it is more likely that a prosodic boundary results in a slowing-down effect, “analogous to coming to a stop after any mechanical movement,” (Klatt 1975; Wightman et al. 1992) and that lengthening as a result provides a cue to that boundary which follows, no matter the domain level (Cambier-Langeveld 1997; Klatt 1975). As seen in figure 11 prior, this
suggestion by Klatt is born out in the differentiation between the phrase-level and word-level instances of “bad,” wherein when “bad” is a noun phrase by itself, the lengthening comes in to indicate it as the end of the phrase, while there is not lengthening found at the end of the word when it is an adjective present in a larger noun phrase (Klatt 1976). This is a contrast of hierarchical prosodic domain resulting in a difference of meaning signaled by FL as a prosodic cue.

While the theory of differentiation of different hierarchical levels has not been tested experimentally, many investigations assessing the usefulness of FL as a word segmentation cue have been carried out. Word-final lengthening has been shown to be a useful perceptual prosodic cue that signals the boundaries of words in Dutch by Cambier-Langeveld (1997). It has facilitated novel artificial language word segmentation for speakers of English, Dutch, and French (Tyler and Cutler 2009), German, English, and likely Basque (Ordin et al. 2017), and English speakers (Saffran et al. 1996b).

While the three studies show facilitative effects of FL for novel word segmentation, Ordin et al. (2017) show language-specific effects of FL as a universally-present but variably-used perceptual prosodic cue in language. In their experiment, 24 speakers per condition each of German, Italian, Spanish, and Basque were exposed to speech streams of CV trisyllabic words in one of four conditions: TP only, antepenultimate lengthening, penultimate lengthening, and FL. The results showed that German and Basque speakers segmented words at a higher level than the already higher-than-chance baseline TP score in the FL condition, and were not helped or hindered in either of the other two conditions. However, Italian speakers were impeded by antepenultimate lengthening, helped by penultimate lengthening, and FL had no effect on their word segmentation. Spanish speakers were impeded by initial-lengthening, and middle and final lengthening had no effect. The result from this is that while lengthening can be regarded as a naturally occurring cue, it is also used, left alone, or may disappear as a cue in a language in favor of other prosodic cues, and then become an impediment to word segmentation in those cases (Ordin et al. 2017). Ordin et al. (2017) suggest that TP evaluation and computation of prosodic regularities, such as prosodic structure and cues, are parallel but separate processes, and that words can be extracted based solely on TP alone, but words which fit into a “prosodic frame” will be extracted faster than those which do not fit into the prosodic structure of a language. Ordin et al. (2017) cite that the helpfulness of penultimate lengthening and lack of helpfulness by FL for Italian speakers in AL segmentation experiment might be to do with the stress system of Italian, in which penultimate stress is the most common stress in Italian, and
Italian stress is primarily signalled by duration (D’Imperio 2002). Therefore, while there might be FL present in the speech stream of Italian, it may not be the strongest marker or cue of word boundaries compared to penultimate stress in the form of penultimate lengthening for Italian in particular, speaking to the idea of there being linguistically-salient cues which are language-specifically instantiated. Meanwhile, they suggest, German and Basque stress are not as regularly predictable, so FL (perhaps taken as more of a phrasal cue in the mind of Ordin et al. (2017)) may become a more reliable phenomenon to pay attention to. This would not discount usefulness of FL which was present at only the word-level, though it is a different view of FL.

Notable in the field is Nakai et al. (2009). Their study focuses on FL in Northern Finnish at the ends of utterances. They had speakers produce speech in which test words were in both phrase-medial or utterance-final position, and then evaluated the effects of FL on vowels in Northern Finnish, in which there is typically a contrast between long and short vowels. Their results showed that lengthening neutralized in final position, however their explanation for it is of great interest in that they postulate FD may be the cause. FD present in the speech stream, they assume, allowed for misidentification of vowels as short, as the voiced part of vowels is the important part perceptually (Myers and Hansen 2007), thus contributing to sound change through misperception on the part of listeners (Nakai et al. 2009). This is particularly interesting because it is another of the few cases in which a study or postulation considers FL and FD in some linked way.

As we have seen, much research has been done assessing the existence and usefulness of FL at various prosodic levels. This investigation has shown that FL is found at various prosodic domain levels and that it is facilitative in novel word segmentation for various languages and provides perceptual cues to prosodic domain differentiation.

2.3 The status of the word final position

Generally, it seems that any final phenomena (e.g. phonetic FL or phonetic or phonological FD) would be ungrounded if their root’s phonetic inherency at the right-edge of words was in question. Exactly what may be transferred by analogy or secondary extension has not been clarified by the proponents of the theory. It is unclear whether in these hypotheses only phonological effects may transfer, or also phonetic phenomena in the case of gradential FD as well as FL which are regarded to both only have a motivation utterance-finally; typically no
bridge is made between these similar phenomena, excepting Blevins (2004), which are occurring, one could hypothesize, in an alike fashion with perhaps similar steps.

Following the explanation of analogy to the word-level, this has been hypothesized to indicate that the word-final area is not domain inherently tied to phonology (Hualde et al. 2015), and nothing phonological may happen there naturally. If the word has no inherent phonological significance, not only would nothing be able to occur inherently at the right edge of words, but it seems to raise the question of how analogy could possibly target this area; this demands, then, that speech users encode words individually with certain phonetic effects utterance-finally after an indeterminate amount of repetitions with those effects, for those effects to read lower domain levels from that part of the lexical entry, but this overrides the analogy idea.

Hualde and Eager (2016) say there is only a phonetic push for phonological devoicing utterance-finally due to approach to pause, thus necessitating secondary extension to explain word-final phonological devoicing, given the assumption that anywhere where something is properly conditioned for the surfacing of phonetic effects will have those phonetic effects. Hualde and Eager (2016) state that the origin of “word-final devoicing [is unclear],” questioning why “in languages like Russian, German and Catalan[,] word-final consonants are voiceless even intervocalically, when followed by a vowel in the next word, a context that should favor voiced realizations,” citing Keating (1988), as the source of phonetic push for FD in the utterance-final area. Having said this they then go on to advocate for the analogy hypothesis. If, however, word boundaries are an environment for phonetic effects, then they could provide the basis for phonological processes. The phonetic FD found in the data of Lunden (2012) and Nakai et al. (2009) at the word-level suggests that there may be a phonetic motivation for such devoicing word-finally. Although the underlying origin of this word-final phonetic FD is at this point hypothetical, its presence provides a basis for phonological FD. If this were the case, it would remove the need for the analogy hypothesis.

One study which has been done in particular to address the theory of domain generalization by analogy was carried out by Myers and Padgett (2014). In this study participants were exposed to nonsense utterances of a made-up language in which there were two different testing groups, one where word-final obstruents of words in the utterance-final position were devoiced (the devoiced learning group), and the other where words in utterance-final position were voiced (the voiced learning group). Neither group of participants ever heard the learning words in any other location than utterance-finally in the learning phase. Then, in the testing phase they were then asked if utterances containing test words that were
similar to the learning words (though not exactly the same words, where half had a devoiced and half a voiced obstruent word-finally) were in the language that they listened to, wherein the test words occurred both utterance-medially (as neutral learning words had) and utterance-finally. The thought was that participants may be able to a) learn the pattern for their language regarding utterance-final voicing of the word-final obstruent, b) generalize this pattern up to encode even utterance-medial word-final obstruents the same way, and c) that this learning would follow the cross-linguistic markedness of voiced versus voiceless codas. Their predictions were correct, in that both groups learned their patterns and generalized them up to the utterance-medial position, however the voicing group did not learn their pattern as well as the devoicing group, since devoicing is a natural pattern as evidenced by its presence cross-linguistically as a phonological rule (Myers and Padgett 2014).

While this shows strong evidence that humans are capable of generalization through the prosodic hierarchy by analogy, it leaves open the question of whether this was a demonstration of analogy from utterance-level (e.g. utterance- to word-), or whether participants were directly encoding that any phenomenon happening utterance-finally is inherent to every level (e.g. word- and therefore everywhere). The question can be posed whether there is any utterance-final-only phonology, or rather, if anything which happens utterance-finally is actually happening at all domain levels because it is a word-final phenomenon, and inextricably the word-final area is sometimes utterance-final.

Motivation utterance-finally for these phonetic phenomena can presumably be extended to phrase-medially, but there is no slowing/approach to pause phrase-medially so the phonetic push here is less clear. There may be the possibility that these phonetic phenomena are present for the explicit purpose of marking or denoting word boundaries in some capacity for perceptual reasons. This would be a stronger claim for the inherent uniqueness of word edges, specifically the right-edge of words. In this case, I propose that these phenomena could be “created” or added in organically by speakers, as organically as language itself is created, for the benefit of listeners. This would then be a reciprocal mechanism of language itself, where these cues are put in for perceptual reasons, and so therefore continue to exist for perceptual reasons, to the point where they are inextricably woven into the speech stream. At that point, any difference between these originally “purposeful” phenomena and true acoustic by-products of speech production would be nearly impossible to tease apart. This perceptual necessitation and resulting creation founds an underlying perceptual motivation for the phonetics occurring. It has been proposed
by Klatt (1975) that FL may be increased at higher domains for the sake of perception. My hypothesis would be in keeping with this, though more extreme.

This could surface multiple ways. There could be a perceptual push for FL, which then provides a phonetic push for FD, or there could be a perceptual push for both. We could also consider a perceptual push for only FD, and phonetic push for FL, however this seems less likely, as FD seems the more predictable or justifiable of the two in terms of a phonetic reasoning (that phonetic push being FL tied in with Blevins (2004)’s idea).

On the other hand, to elaborate on earlier thoughts, as Blevins (2004) suggested at the utterance- or phrase-level, and which we may presumably extend, there may be a phonetic push for both, in that in a way the end of a word would be a place to slow down cognitively anyway, giving a phonetic push for FL, which then may provide a phonetic push for FD. In this scenario, I do not suggest either would be specifically input into the speech stream for perceptual reasons, but rather since they are already present for phonetic reasons, they are then used for perceptual reasons after their already inherent existence. Speakers realize they are there, so then it becomes encoded in language to use them. This could of course be the case with analogy being the reason for their existence throughout the speech stream as well.

The study done in this thesis is not necessarily able to answer these questions or speak to one hypothesis over another directly, and does not shed light on the motivational origins of FL or FD, however it can be a piece of evidence towards understanding their role word-finally, showing simply if they are perceptually salient, no matter which part of the chicken-and-egg relationship is the origin of these phonetic phenomena/potential perceptual cues.

3. Statistical learning
3.1 Word-parsing and segmentation through statistical cues

Questions of how infants learn the infinitely complex, generative, and opaque system that is language, particularly of word segmentation, are continuously discussed in the field of language acquisition, as when looking at the fluid and fluent speech streams produced by native speakers which infants are exposed to, there are no regular gaps in the sound between words, as shown in figure 12 from Saffran (2003)4.

---

4 Reused with permission from SAGE Publications.
Fig. 1. A speech waveform of the sentence “Where are the silences between words?”
The height of the bars indicates loudness, and the x-axis is time. This example illustrates the lack of consistent silences between word boundaries in fluent speech. The vertical gray lines represent quiet points in the speech stream, some of which do not correspond to word boundaries. Some sounds are represented twice in the transcription below the waveform because of their continued persistence over time.

Figure 12. Saffran (2003)

Much of the research which has been carried out to assess possible skills for this notably challenging task have shown that infants can use statistical or pattern learning to learn at least some language. One of the principal studies done which showed that infants are capable of segmenting or encoding words solely off of TP statistical information was that of Saffran et al. (1996a). In this study, two groups of 24 8-month-old infants were exposed to 4 trisyllabic AL nonce CV words produced by a speech synthesizer and concatenated for 2 minutes in two different experiments. In the testing phase of the first using a head turn preference paradigm, infants were tested on their ability to differentiate between the words they heard and sequences that never occurred, and between words they heard and part-words, or syllable sequences which crossed word boundaries in the second. There were no other cues to word boundaries present other than TP, and yet infants were able to distinguish between words they had heard and non-words or part-words at a greater-than-chance level.

Conversely, further investigation by Johnson and Jusczyk (2003) and Johnson and Tyler (2010) demonstrates that infants are not capable of parsing apart an AL speech stream using just TPs, or statistical pattern-learning, alone, when the words in the AL are of different syllable-lengths, even if the TP amounts between conditions are kept the same and the only variable is how many syllables are in some of the words. They conclude that these ALL experiments are not true-to-life, not only in that 2 minutes of exposure is an unreal amount for any infant, but also that a language as simple as the ones necessary for these experiments to work does not naturally exist; natural languages are far more complex, with infinitely greater lexical and syntactic variation (Johnson and Jusczyk 2003; Johnson and Tyler 2010).
Because the statistical learning mechanism has been shown to fail when even a minute amount more of complexity is introduced, this suggests that the statistical learning mechanism must be guided by some innate and likely linguistically-specific constraints in order to be usable in a real-world natural-language language-learning environment (Yang 2004). This especially seems to be true because of adults’ capability to use this TP-taking skill in the same task run by Saffran et al. (1996b) on infants, as well as taking it a step further and completing the multiple-length words task that infants could not in Johnson and Jusczyk (2003) in both Johnson and Tyler (2010) and Tyler and Cutler (2009). Something which seems to aid the idea that TP-learning is a generalized method of learning available to the human brain which can be applied to language learning, but that there are other mechanisms or input which assist in this effort more, is that in the adult segmentation experiment done by Saffran et al. (1996b), performance improved from greater-than-chance with just TPs, to significantly better with the added prosodic cue of “word-final lengthening.” This provokes the current study’s focus on FL and FD as word-final potentially facilitative cues for word parsing in language-learning broadly.

One thing which has not been investigated as a potential word segmentation cue is aforementioned phonetic final devoicing. While FD has, as previously mentioned, been measured phonetically in speech streams of various languages at various prosodic levels, and is in fact encoded word-finally as a phonological rule (perhaps a very robust word-segmentation cue in the specific languages where this occurs) it has not been tested as a novel AL word-learning cue the same way that FL has.

Nickels et al. (2013) found that L2 parsers of second language, at least when the language is closely related to their native language, use prosodic cues much like native speakers, and even to a further extent. This study was done with German- and English-speaking populations of participants sorting out syntactic ambiguities created by cross-spliced prosodic boundaries, wherein the German speakers more firmly rejected the mismatch conditions than the English speakers, presumably because the German users are more cognisant of prosodic boundary cues from their own language than English users are, so they mapped that native use of a universally present cue onto the task even in a language that was not their mother tongue (Nickels et al. 2013). Discussion is nonspecific about what exactly became those prosodic cues when they cross-spliced their stimuli, however one can assume the cues present were those which are present seemingly cross-linguistically (which have been discussed): intonation, and FL as well as perhaps FD which are the focus of this paper. This returns to the idea that these cues may be present universally, though get used in a language-specific manner which may be stronger or

Rouch 25
weaker depending on the language, as was mentioned with different uses of lengthening, particularly in Ordin et al. (2017).

3.2 Phonetic gradient versus phonological categorization

The classic example of phonetically gradiential variation in tandem with a phonological effect or categorization is of voice onset time (VOT) producing a phonologically categorical divide between voiced and voiceless phonemes in English. In many studies (e.g. Ganong 1980; Keating et al. 1981; Reetz and Jongman 2009) it is shown that though VOT varies phonetically per speaker or per production, there is a certain point at which perception of that VOT (around 35 milliseconds) has a threshold across which it will be categorized as one or the other, higher will be perceived as voiceless and lower voiced. Therefore we see an example of phonology categorically dividing perceptual objects into a dichotomous system. Phonetic phenomena have a variable gradentiality to them with no defined boundaries, which phonology can interpret and divide into discrete categories.

FD has both a phonetic and phonological presence; it is present in a phonetically varied way at many different domain heights, and as a phonological rule in some languages. Presumably it is similar to VOT in that it varies phonetically and at a certain point crosses the threshold necessary to be perceived as phonologically devoiced from a voiced sound. Note that there is no exact parallel for FL; FL is present at many domain heights phonetically variably, but does not have a phonological relative ⁵.

The difference between phonetic and phonological phenomena becomes relevant in this paper, as the prosodic cues being discussed are at the phonetic level, however can have an influence on language in a phonological way. As mentioned before, phonological phenomena are a categorical “on” or “off” (such as of voiced or voiceless in phonological FD), while phonetic phenomena are chaotically variable in a continuous way (such as gradiential phonetic FD or FL). For example, from Nakai (2013)’s work she proposes that phonetic FL may have a perceptual phonological effect on misidentification of long vowels in Japanese as short when short vowels are finally-lengthened at the word-level, prompting a neutralization of vowel length distinction. This may also be assisted by FD in that final long vowels, even with FL, may be perceived as short due to their voiced sections being within the realm of finally-lengthened short vowels following FD. Myers and Hansen (2007) contend, and support their assertion with data from Finnish, that though FL is a dominant universal pressure cross-linguistically, FD can, similarly

⁵ Except for the aforementioned word-final vowel lengthening in footnote 1.
to in Nakai (2013) and Nakai et al. (2009), shorten the perceptible or salient portion of a vowel word-finally (the voiced part with vowel formants intact), utterance-finally to the point that in a language even with phonemic vowel length distinction, vowel shortening and subsequent neutralization may occur.

The work done in this thesis seeks to ascertain whether phonetic gradientiality or “chaos” is relevant for perception, supporting the presence of these phenomena through evidence of perceptual salience and usefulness of those phonetic phenomena, showing why they might be present at all or that their presence is used, or if they are there “accidently” or incidentally by analogy.

3.3 The present study

We have seen the gap between ideas of analogy and the existence of phonetic, prosodic phenomena such as FL and FD at various prosodic domain levels lower than utterance-finally. Therefore this study seeks to provide evidence for the inherent existence of FL and FD at the word level in the form of perceptual usefulness for novel word segmentation and language learning. This investigation will be carried out through a very similar AL study to Tyler and Cutler (2009), with the addition of FD as another condition, assessing whether prosodic cues improve word segmentation above the level of just TPs, and whether FL and FD might behave similarly or differently. If the cues assist this segmentation above the level of TPs alone, and if FD behaves in a similar or different way to FL which has already been tested, this may provide evidence for these phenomena to exist organically at the word level. This would be for no other reason than that they have this perceptual purpose and have been programmed in there because of that, or they are there, perhaps they are linked or caused by a similar root, no matter if that root is intrinsic existence or analogy, and used because of their inherent existence.

4. Methods

4.1 Participants

All participants were native English speakers from ages 18-24 years old, (N=106; average=19.58 years). Participants in the study were either from the William and Mary psychology participant pool or psychology or linguistic classes and received class credit for participation, while others were not part of the participant pool.
4.2 Stimuli

Learning phase stimuli were created using MBROLA (Dutoit and Pagel 1995) and Praat (Boersma and Weenink 2005) software. Nonsense artificial language (AL) words of one, two, and three syllables in a CV(CV(CV)) pattern of the segments [m, dʒ, j, z, k] for consonants and [a, i, e, o, u] for vowels were produced in MBROLA, using the US1 voice. MRBOLA was used to create the speech, then modified in Praat to fit the three conditions. The first condition was (i) transitional probability (TP) or the likelihood that syllables will occur in sequence together only, wherein this likelihood is higher within words, lower across word boundaries, and nonexistent when a sequence has never occurred. The second was (ii) final lengthening (FL) with final lengthening added to the ends of the words. And the third was (iii) final devoicing (FD) with final devoicing added to the ends of the words.

These segments were chosen for their distinguishability from one another as well as intelligibility coming out of MBROLA. Twelve unique syllables were created with each segment being used at least twice, and two consonants and two vowels being used three times. Finally, syllables were created such that actual words of English were attempted to be avoided, although some names (e.g. “Moe” and “Joe”), one noun (e.g. “jaw”), two onomatopoeic words (e.g. “gee” and “moo”), and one abbreviation of a noun (e.g. “za” for “pizza”) were present. In the all three conditions consonants were 90 milliseconds long in all word-internal positions. In the TP and FD conditions, high vowels were 100, mid vowels 110, and the low vowel 120 milliseconds long. In the FL condition, these amounts were increased to 150 for high vowels, 165 for mid vowels, and 180 milliseconds for low vowels when the vowel was in word-final position (150% of the original base length of the vowels). Finally, in the FD condition, 33 ms, 36 ms, and 40 ms of the end of the vowel respectively for high, mid, and low vowels was devoiced (33% of the final vowel).

The devoicing effect was created by synthesizing a VhV sequence in MBROLA for each word-final vowel. Subsequently, in Praat, the word was extracted with 66% of the normal, voiced, final vowel, and 33% from the voiceless, vowel-like [h]. This extracted [h] was run through the stop Hann band filter in Praat, set to filter out 0-500 Hertz, then concatenated with the beginning of the word to form the full length of the word with 33% of the final vowel devoiced.

To create each stimulus for the three conditions, the words were concatenated in randomly-ordered, six-word groupings generated by a random number generator, and hand-controlled only to make sure that no word was ever immediately repeated directly after
itself, repeating this randomization of the six words in 51 groupings until about 2 minutes of stimulus had been created for the FD and TP conditions, and 2 minutes and 18 seconds for the FL condition. These sections of the 51 strings were then copied to create almost exactly 7 minutes of stimulus for each condition. The order of the words in the 2 or 2.3 minute concatenation was identical in each condition; so, note that although the words were in identical orders in each condition, the FL stimulus simply had less of this order repeated, because of the fact that the words were longer and so took up more, proportionally, of that 7 minutes.

For the testing phase, the six real words were presented along with six each of two non-real words, “part-” and “non-.” The six part-words consisted of sequences of syllables which occurred across word boundaries in the speech stream, thus their TP was lower than that of the syllables which were always in sequence together in one word. The non-word category consisted of six words made of sequences of syllables which never occurred in the speech stream (e.g. a three syllable word made of the middle of one word, a single syllable word, and the end of a two syllable word). However, for both non-real word types, all the syllables used were present in the speech stream. Additionally, for words of the type “part-,” the word-internal location the parts were taken from was varied so that this could not have been learned as a regular pattern or made it obvious these were crossing word boundaries without actual encoding of where prosodic cues had occurred (e.g. some began in the middle of trisyllabic words, some at the beginning or end of disyllabic words, and some ending in monosyllabic words which would create the critical condition of having the cue present in final position and thus be a potential lure).

The real, part-, and non- (nonce, AL) words were as follows:

<table>
<thead>
<tr>
<th>Real</th>
<th>µdʒafī</th>
<th>fokuze</th>
<th>dʒimo</th>
<th>kedʒo</th>
<th>za</th>
<th>je</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part-</td>
<td>dʒafike</td>
<td>zafoku</td>
<td>dʒomu</td>
<td>ñedʒi</td>
<td>ze</td>
<td>mo</td>
</tr>
<tr>
<td>Non-</td>
<td>dʒajoze</td>
<td>mukeza</td>
<td>moji</td>
<td>ñedʒu</td>
<td>dʒi</td>
<td>ku</td>
</tr>
</tbody>
</table>

4.3 Procedure

The experiment was carried out in a sound-attenuated booth, one or two participants running at a time. Each sat in front of their own computer monitor with their own set of Sennheiser HD 280 Pro headphones and in the listening phase were instructed that they were going to listen to about 7 minutes of stimulus, explained as words from a made-up language strung together. They were told they did not have to memorize anything, and to just let the
words “wash over them.” They were not told what they were listening for or that they should be trying to segment words, and likewise were told nothing about the AL (whether the words had any meaning or the language any grammar).

Each participant was assigned randomly to one of the three conditions (TP N=35, FL N=35, FD N=36). Once the seven minutes was over, the experimenter returned for the testing phase. A short study in Praat, referred to as a “study” so as to not pressure the participant, was initiated wherein the participant was played all 18, six each, of the real, part-, and non- nonce words in a randomized order, and asked to rate them 1-5 on how certain they were they had heard this word previously in the words in the made-up language they listened to\(^6\). 1 represented that they were certain they heard it, 2 that they thought they heard it, 3 meant “I don’t know,” 4 indicated they thought they did not hear it, and 5 certain that they did not hear it. There was a 500 millisecond pause before each stimulus was played, and there was a replay button present on the screen that was usable once per stimulus, which participants were instructed was meant to be used if they had missed hearing the word entirely, and not to ponder more about whether they had heard it before or not.

5. Results
5.1 Data analysis

Data analysis was done with SAS. A Linear Mixed Model was run on response as the dependent variable with factors condition, word_type, condition\*word_type, with the random effect of subject nested within condition in SAS; reported p-values come from pairwise comparisons of the interaction effect between condition and word_type. Average word acceptance rating results are shown in Figure 13.

\(^6\) There is some reason to think that the phrasing of this question or make-up of the testing phase sequences could make a meaningful difference in task-success rate, as in Wang et al. (in prep), participants were asked “Do you think that you heard this sequence in the previous section?” and were tested on only the disyllabic words from the learning phase, though two disyllabic and two trisyllabic words made up the learning-phase. In the study done in this thesis, the word-nature of the AL words was emphasized in both the learning phase (e.g. “You will be listening to words from a made-up language”) and in the test phase, where the on-screen instructions read: “For each word you hear, please select whether it was one of the words you just heard or not.” and “Was this one of the words you heard in the recording?” Additionally, participants were tested on all six learning-phase words in full, in addition to the six part- and six non-words.
Overall, aggregated averages indicate a word-acceptance bias in all three conditions, whose averages never exceed 3 (the level of “I don’t know” in the scalar response). In all three conditions, subjects had a significantly different response between real words and part-words and real words and non-words ($p<.001$).

Participants in the TP condition had a statistically significant difference between real and part-words ($p<.001$), however there was no statistically significant difference in rejection of not-real words between part- and non-words ($p=.203$).

In the FL condition, there was a statistically significant difference between real and part-words ($p<.001$), however there was no significant difference between part- and non-words ($p=1.0$).

FD participants significantly differentiated between real and part-words ($p<.001$), but much like the other two conditions, there was no statistical difference between part- and non-words in this condition ($p=1.0$).

Participants in the FL condition were statistically significantly better than TP at accepting real words as ones they actually heard in the speech stream, although not statistically significantly better than FD (compared to TP $p<.001$, FD $p=.094$). FD did not differ from TP in
this part of the task ($p=.453$). This suggests that FL as a cue improves real-word identification over the level of TP in a way that FD does not, however that FD may be on its way to doing so.

Relatedly, the results seem to indicate that FD is not a very salient cue for word segmentation. The range in the FD condition of acceptance of real words versus rejection of either not-real word type is more compressed than in either other condition (total range of average data for each condition showing FD has the most compressed range: TP: .65, FL: .74, FD: .49), indicating that FD has the least robust effect on participants’ responses.

The FL and FD conditions’ treatments of the not-real word types, part- and non-, were as the same (not statistically significant at $p=1.0$ for both conditions). In fact, on the surface, looking at the averages participants in the FL condition actually seem to get worse at rejecting non-words compared to part-words (part-word average versus non-word average acceptance rate respectively in the FL condition is 2.46 and 2.37). However this is again not a statistically significant difference. This is not different from what we saw of TPs before, in that TP also has no statistically significant difference between part- and non-words ($p=.203$).

FL responds to non-words with a level similar to FD (FL compared to FD for non-words $p=1.0$, average word acceptance is 2.37 and 2.36 respectively). TP approaches significance or is marginally significantly better than FL or FD at this part of the task, rejection of non-words ($p=.053$ and $p=.054$ for FD and FL respectively). However, TP shows no strength over either other condition at rejecting part-words ($p=.337$ and $p=1.0$ compared to FD and FL respectively).

While there was not a statistically significant difference between TP and either other conditions’ rejection of part-words, the TP-only condition was the best of the three at rejecting not-real words on the whole, excelling at the non-word rejection (statistically significant, and averages for TP and FL respectively are 2.7 and 2.37). The TP condition’s success at this portion of the task addition to FL and FD’s struggle to reject non-words provokes further thought about why these cues may be helpful for accepting words, but fail to assist in the rejection of non-words. These results will be postulated about in the discussion which follows.

5.2 Discussion

The data of the study indicates that all three conditions are capable of accepting real words above either not-real word type with statistical significance. Additionally, both prosodic cues, FL and FD to varying degrees, facilitate AL word-learning in the form of real word acceptance. This study’s results replicate the finding that adults are capable of AL word segmentation on words of different syllable-lengths using TPs, or statistical information, alone.
This also is the first experiment, to my knowledge, which tests participants’ ability to segment monosyllabic words with other length words at all, as well as multiple monosyllabic, disyllabic, and trisyllabic words in one speech stream (trisyllabic and 4-syllable words, as well as disyllabic, trisyllabic, and 4-syllable words have been tested before). Additionally, this study replicates the finding that an FL cue on the ends of AL words helps participants to parse the words at a higher level than just TPs.

The FL condition is the best at accepting real words, and is capable of separating out real words from either not-real word type (part- or non-) in a more definitive way than and above either other condition. Presumably this is because of the extra information the FL cue is providing in the speech stream. The FD condition does not appear to be getting the same boost, despite FDs presence in the speech stream at various prosodic levels within the AL here and in natural speech. TP is the best at rejecting non-words, statistically significantly rating these lower than either other condition did.

Overall, we see a different treatment of real versus not-real word types depending on the condition. The lack of statistically significant difference between part- and non-words in the TP condition is a much less robust effect than those in the FL and FD conditions despite the values from the pairwise comparison, as there is a total .22 difference between part- and non-words for TP, where there is only a .09 and .11 difference for FL and FD respectively. By the average acceptance rates, TP seems to be varied basically by exposure to the words, which decreases by word type from lots of exposure to real words, some exposure to part-words, and no exposure to non-words (real, part-, and non-word acceptance rate averages respectively: 2.05, 2.48, 2.7). Presumably FL and FD are using both TP and the prosodic information present uniquely in each condition to rule out the not-real word types as potential words. They may or may not have ever heard the syllable sequence (occasionally for part-, never for non-), and never would have had a prosodic cue in the speech stream to help them encode either not-real word type; in fact, prosodic information, much like TP information, would stand against either not-real word type; therefore with double the information, they are treating the not-real types as definitively the same, much less close to having a difference. This may suggest a different treatment, gradential based on exposure versus categorically distinguished based on a yes/no acceptance or rejection of real versus not-real words, in the different conditions which have only domain general

---

7 While experiments 1 and 2 from Wang et al. (in prep) use AL sequences of four monosyllabic nonce words, participants were only tested on their ability to recognize or deny familiar/cued sequences, not the individual monosyllabic words.
statistical information available to them (TP) or that plus linguistically-categorical prosodic cues (FL and FD).

Surprisingly, we note that both prosodic cue conditions are approaching being significantly (at marginally) worse at out-right rejecting non-words compared to the TP condition. Both FL and FD are statistically significantly equal at rejecting part-words to TP, and we see no improvement in this portion of the test phase like we might expect. FD is also worse than TP at rejecting this word type, based on the averages (TP: 2.48, FD: 2.25). Generally, it is logical to think that the prosodic cue conditions should outperform the TP condition at both subtasks in the test phase because they have more information to work with, subtasks being more often accepting real words after having properly segmented them out of the speech stream with the boost of cues above TPs, and also more often rejecting both not-real word type. This should be possible with prosodic cues indicating part-words were bridging word boundaries and were therefore not words, or even simply rejecting non-words either for never hearing them, or encoding that the pieces had never been properly cued to be words with the prosodic cues.

One hypothesis to explain why we do not see better rejection of not-real words with the addition of prosodic cues from the data is that the prosodic cues have been generalized, or have been learned more linguistically-specifically to be a marker of what is a word in the language. Though participants were asked in the test phase if the word presented to them auditorily was a word they heard prior, they may have encoded that the FL and FD cues being present at the end of a word meant that it was in fact a word of the language. This is somewhat semblant to the results of Myers and Padgett (2014) in that a cue which is present at the end of a word potentially may be generalized to denote whether something is a word in a language or not. However, it does not seem to indicate anything about analogy but rather that a cue can in fact be encoded as a marker of a word boundary. It seems that having the cues at the ends of the test words in the test phase might have hindered the experiment in discovering whether participants were actually better able to segment the six real AL words out of the AL speech stream, as their statistic-taking TPs and prosodic cue word-boundary demarcating capabilities were actually interrupted or overridden in the test phase when, by all given indications, all the words were likely to be real words in that particular AL.

FD and FL are both marginally significantly worse at rejecting non-words than TP. Additionally, there is no statistically significant difference between FD’s treatment of part- and non-words, much like FL. However, again, FD is not statistically significantly better at accepting real words than TP, unlike FL which is, though there is also no statistical difference between FL
and FD at this portion of the task. In this way, FD seems at first glance to be approaching the pattern that FL has, of specific language-learning as compared to TP gradientiality. However, the aforementioned limited range of FD, accompanied with the fact that it performed worse on either end of the spectrum than either FL or TP seems to indicate that overall, for English speakers, it is not as robust a cue as FL may be to word boundarries or segmentation. If the data from the full experiment were combined with data of a pilot study\(^6\), the FD condition appears to be a copy of the shape of the TP condition’s curve about .3 lower (.3 better at accepting real words and worse at rejecting not-real words) than TP. This is shown in figure 14, a graph of the combined data across the pilot and full study.

![Graph showing mean word acceptance](image)

Figure 14.

Additionally, preliminary results from the pilot study nearly replicated the results of the full study, minus this change in statistical significance in the FD not-real word types and movement of TP to be better at rejecting not-real words than other conditions numerically. Because of these facts and FD’s averages compared to that of FL from the spring study only

---

\(^6\) The pilot study was carried out in the fall of 2018, and was rerun as the full study due to stop closures being absent in the stimulus, as well as changes to the stimulus such as amount of final lengthening in the FL condition, changing of phoneme [d] to [d\(\text{\textbar}\)] for its greater distinguishability, and changing of lengths of vowel and consonant phonemes.
real, part-, and non- in FD: 1.87, 2.25, and 2.36 showing a tightly gradiential stair-stepping effect vs. real, part-, and non- in FL: 1.63, 2.46, 2.37 with a lot more range comparatively and gap between TP and both conditions uniformly, respectively), I speculate that FD is actually more gradiential and controlled by amount of exposure to the word types like TP than categorically-differentiating FL. While showing the same trajectory, the fact that FD regularly shows a lower response than TP leads me to suspect that FD does not have an impact the magnitude of response. I postulate that the shapes of TP and FD in are different enough from FL, and the averages gradiential enough, that TP and FD are having a different type of effect on word-learning than FL. TP and FDs’ treatment of the words seem to be more gradiential, based on exposure.

This study supports the hypothesis that prosodic phonetic phenomena at word boundaries can be perceptually helpful for word segmentation out of an AL speech stream. This gives some support to their inherent presence at the word boundary, and therefore the importance of the word boundary as a domain where relevant phonetic-push phenomena happen in situ. FL certainly improves segmentation above the level of TPs, and FD approaches it. FL and FD also seem to be behaving differently as perceptual cues, however, in that FL provokes a linguistically-categorical response of real versus not-real words, whereas FD, like TP, I conclude, is limited to a gradiential effect, or responsive based on exposure amount.

6. Conclusion

The results in this experiment provide perceptual evidence for the inherent existence of prosodic cues word-finally at all domain levels in the speech stream, and for the importance of the right edge of words phonologically and perceptually. The improvement of word segmentation in the prosodic cue conditions indicates that the cues are useful in accepting the real words from the AL. This demonstrates that they may assist in word segmentation, which provides a reason for them to exist inherently. Therefore, this is also evidence against the theory of analogy from higher domain levels, and of the natural importance of the word boundary as a domain which has phonetic items providing a push towards both phonological categorization as well as linguistically-constrained phenomena and learning.

The fact that an FL cue at the right edge of the word has improved performance over TPs alone in past studies (Tyler and Cutler 2009; Ordin et al. 2017) and in the current study suggests that this is a cross-linguistically, in more than one language, useful language learning-specific cue. This study has also shown that adults can segment monosyllabic words with other length
words, in addition to multiple monosyllabic, disyllabic, and trisyllabic words from one AL speech stream using solely TPs, but improving with FL added, and approaching improvement with FD added. The benefit seen in novel word segmentation in adults suggests that testing this with infants would be the next step; if infants are not capable of segmenting novel words which differ in syllable length using solely TPs (Johnson and Jusczyk 2003; Johnson and Tyler 2010), they may become capable of segmenting them when these cues are added.

These findings have further supported past results of research done about adult performance in ALL word segmentation using just TPs, in that adults are capable of this at a greater-than-chance level. It also shows that adults perform better generally with the addition of cues to the word boundary, such as the prosodic cues mentioned here. While neither FL nor FD were shown to distinguish between the two not-real word types, this does not seem to be as robust a finding for FD as it is for FL. This is because overall, FD has a very compressed range compared to either FL or TP, so statistical significances and lack thereof from the FD condition are less likely to be as significant. Additionally, as mentioned before, if the data from the pilot study if combined with the current study’s, FD followed TP in showing a more gradiential increase in response. FL is notably different from either, in that it has a robust, more linguistically-constrained learning effect, treating both not-real word types as equally worse than real words statistically and numerically. This might suggest two different kinds of learning mechanisms and cues. TPs are a gradiential, general learning cue, and statistical learning is a broad learning mechanism which can be applied to language learning. FD does not seem to add on to this statistical learning result, at least from the results in this study. However FL allows for a better distinction between real words and either other not-real word type, part- or non-. This is a linguistically-driven language-learning effect, provoked by a language-specifically applied, though cross-linguistically present, learning cue: lengthening on the ends of words.

It is important to note that while this study replicates the finding that FL assists word segmentation beyond TP capabilities, the motivation of the present study is one of a few lenses to look at FL through, of which some studies focus on more than others. While Tyler and Cutler (2009) do show the facilitatory effect of FL, their focus on FL is more of a happenstance from their focus on more generally rhythmic phenomena in the speech stream such as stress, accent, or iambic feet, which lend to lengthening in general. This is a different, though strongly related, lens to look at FL through than in my study, which is more akin to the experiment done by Ordin et al. (2017). In the approach taken in my study and in Ordin et al (2017)’s, FL is regarded as not only something which can be tied to rhythmic phonological phenomena, but also as its own
phonetic phenomenon that exists inherently and universally in the speech stream, and may be treated differently language-specifically. These different approaches to FL in particular demonstrate that prosodic cues, or word-demarcating phenomena, if that is indeed what they are, may be different in each language, though all the phenomena may exist universally in the speech stream of a given language, and language may only need to “pick” a few in order to have a functional and load-bearing structure for language-acquisition that is learnable with repeated exposure to those prominent happenings by an infant.

There may be a lack of robust facilitation on the part of FD above the level of TP because the FD present in this study has been “watered down” so to speak, or not given the full potential or power it has in real language. Though the 33% threshold of devoicing word-finaly used here is a realistic amount, the actual amount of devoicing present is less than natural speech due to the fact that there was no FL in the FD condition, therefore there was actually less FD present than there would have been in real speech where the vowel would have been naturally longer, contributing more duration to that 33% in the first place. This might have contributed to the lack of distinct pattern or results we see in FD compared to either TP or FL. It is also worth considering however that FD may not have the same stand-alone effect in general that FL does because of this; this could be a preliminary indication that FD, as a potentially perceptually useful cue, might be tied inherently to FL, thus prompting the theory that FL provides phonetic push for FD. Overall, I would expect that if more data were gathered in this iteration of the experiment, or if the experiment were rerun either with a higher proportion of devoicing present or devoicing present at 33% of the final vowel with FL added in, we may then begin to see the kind of effect that FL is having for language-parsing. However, if no effect was found, this would not be because FD is a domain-general cue that behaves gradientially, but rather that it does not change performance when added in with TPs; it may not be a salient language cue at all, though it is inherently tied to language and is not domain general. It may behave as effectively another iteration of the TP condition.

The results of this study thus are suggestive of a difference between two patterns of learning, one which is domain general or gradiantial and the other which is linguistically-driven and -constrained. Further investigation could lend to the continued discussion within the field of language acquisition about what parts of human language are learned through which mechanisms, domain general or linguistically-constrained and -specific. Additionally, if the plateau-like FL curve versus gradiential curve of TP could be replicated and instantiated as a regular pattern, this difference in curvature could then be used as a sort of diagnostic tool for
determining whether a linguistic phenomenon acts as a language-learning cue in a specific language or not.

This study acts as a piece of evidence towards the inherent importance of phonetic cues at the word boundary. The fact that listeners are able to make use of final lengthening to segment words suggests that it is a word-level cue they are familiar with. We also see that it is not the case that any marking of the word-final boundary perceptually facilitates word parsing, as cued FD (in the FD condition) was a consistent cue to word-boundaries and yet did not result in the better identification of real words. The uniqueness of the word boundary would be consistent with either the hypothesis that they are acoustic by-products of sound and speech production that listeners can make use of, or the new, suggested hypothesis, that these phonetic phenomena are perceptually driven and surface as a result.

One potential reinterpretation of these results may be that participants were interpreting the AL words as utterances, and thus any segmentation was actually presumed to be of utterances. This would mean that what participants were picking up on the FL present as utterance-final FL, and assumed every word was spoken alone, thus its own utterance which would have FL there where it has been attested to occur, and be segmentable. While this could be an interpretation, there are no intonational or pitch cues in the stimuli of the present study which utterances in English would be packaged with, as opposed to Langus et al. (2012) in which two different types of prosodic cues which are markers of different things to Italian speakers were treated differently. Additionally, the durations of the syllables and segments are based on reasonable word-level durations, not as if they were an entire utterance, where we would expect notably longer durations. Langus et al. (2012)’s work also gives evidence that people treat words and/or phrases differently than utterances, with the aforementioned differences of packaging in English in particular of prosodic cues versus intonational pitch arcs respectively.

One of the questions explored in this paper is the motivation for these phenomena to exist word-finally. As suggested in 2.3, Blevins (2004)’s explanation that inherently present FL can cause FD could be extended to the word-level. If this is indeed the phonetic motivation, manipulation of proportions of devoicing in a finally lengthened segment in a further perceptual ALL study may be helpful to address if this relationship exists. If it is shown that a natural proportional relationship between the two is most facilitative perceptually, this would support that natural linkage. Further, if more skewed proportions hinders participants’ segmentation of nonce words, this would be additional evidence of a connection between the two. If this relationship between FL and FD can be established it would be a phonetic motivation for the
phenomena word-finally. This would render the analogy hypothesis moot, due to the fact that it is meant to explain these phenomena word-finally given the assumed lack of a phonetic-push in word-finally. Additionally, this might provide a more realistic test of whether more natural levels of FD would have a helpfulness effect on word parsing.

Future perceptual research following the conclusion of this project might also be to investigate whether these prosodic cues allow infants to segment AL words of different syllable lengths, which prior research has shown disrupt the statistical-probability taking skills they possess (Johnson and Jusczyk 2003; Johnson and Tyler 2010). Additionally, experiments to research whether FL behaves in the same language-specific way for infants who might not have a fully formed template of their native language yet, as they do for adults who do, would also be helpful for assessing the cross-linguistic prevalence of the cues for segmentation.

Finally, given the mixed results of this experiment which indicate useful new information about language-learning capabilities and mechanisms of generalization of cues for word boundaries, a follow-up directly getting at whether word-boundary cues which exist in a speaker’s native language or universally present speech stream cues will facilitate word segmentation from exposure to a novel AL speech stream would be pertinent. This experiment would be set up identically to the one described in this thesis, however in the test phase none of the test words would have prosodic cues on the ends in any condition. This would avoid testing whether they had generalized a cue as a word-boundary marker, and instead focus on whether the given word-boundary cue made learning the specific words easier (which may be a language-specifically affected outcome, based on the given cues in the learner’s native language). A different version of this could be to present participants in the test phase with test words they never heard in the learning phase, some of which have the prosodic cues on the end and others which do not, asking whether these are words from the language they just heard. This could more firmly assess whether generalization of a word-boundary cue is possible and helpful for identifying even novel words from a language which one has never been exposed to, much like infants might when learning their first language.
References


*Intonational phrasing in Romance: The role of syntactic and prosodic structure*.  


Fougeron, Cécile. 1998. *Articulatory variations at the beginning of prosodic constituents of different levels in French*. Diss. ANRT University of Lille III.


International Journal sponsored by the Foundation “Foundations of Language” 34.1: 131-166.
Hualde, José Ignacio, Christopher Eager, and Sara Little. 2015. Final devoicing in Castilian Spanish. Talk given at MidPhon 20, University of Indiana Bloomington, September 13, 2015.


Nickels, Stefanie, Bertram Opitz, and Karsten Steinhauer. 2013. "ERPs show that classroom-instructed late second language learners rely on the same prosodic cues in syntactic parsing as native speakers." *Neuroscience Letters* 557: 107-111.


