2020. *Proc Ling Soc Amer* 5(1). 599–606. https://doi.org/10.3765/plsa.v5i1.4730.

The status of word-final phonetic phenomena

Megan Rouch & Anya Lunden*

Abstract. The right edge of the word is a known domain for processes like phonological devoicing. This has been argued to be the effect of analogy from higher prosodic domains, rather than an in situ motivated change (Hock 1999, Hualde and Eager 2016). Phonetic word-level phenomena of final lengthening and final devoicing have been found to occur natively word-finally (Lunden 2006, 2017, Nakai et al. 2009) despite claims that they have no natural phonetic pressure originating in this position (Hock 1999). We present the results of artificial language learning studies that seek to answer the question of whether phonetic-level cues to the word-final position can aid in language parsing. If they do, it provides evidence that listeners can make use of word-level phonetic phenomena, which, together with studies that have found them to be present, speaks to their inherent presence at the word level. We find that adult listeners are better able to recognize the words they heard in a speech stream, and better able to reject words that they did not hear, when final lengthening was present at the right edge of the word. Final devoicing was not found to give the same boost to parsing.

Keywords. phonological domain; artificial language learning; final devoicing; final lengthening

1. Word-final phenomena by analogy. Phonological final devoicing is known to occur as a phonological process word-finally in many languages, such as in German and Dutch (Hock 1999). This phonological devoicing is generally thought not to have a phonetic motivation word-finally. Instead, it is hypothesized that there is a phonetic motivation only utterance-finally, due to the approach to the following pause (Hock 1991, 1999; Hualde and Eager 2016). There is thus utterance-final breathy voice or partial, phonetically variable amounts of devoicing, which arise in what can be described as as near to assimilation to silence as possible (Hock 1991, 1999, Hualde and Eager 2016, Keating 1988, Lieberman 1967, Myers and Padgett 2014).

In the analogy theory, after this initial inherent phonetic push toward devoicing utterance-finally then the phenomenon can be phonologized utterance-finally, and analogized down to lower domain levels (Hock 1991, 1999, Hualde and Eager 2016). The analogy hypothesis provides a motivation and explanation for how phonological final devoicing is found at the right edges of words, when this would seemingly interfere with an inherent pressure to maintain voicing between voiced sounds, which phonological final devoicing interferes with in some cases. However, there may in fact be final phonetic amounts of devoicing, or breathy voice, in situ word-finally, which could lead to phonological devoicing word-finally.

* We thank members of the spring, summer, and fall 2018 - 2019 Computational and Experimental Linguistics Lab (CELL) at William & Mary and particularly Kate Harrigan for helpful feedback and discussion. Thank you to the audience of the 94th Annual Meeting of the Linguistic Society of America for helpful comments. Thanks to Kim Love of K. R. Love Quantitative Consulting and Collaboration for the statistical model. Any mistakes are of course our own. Authors: Megan Rouch, William & Mary (marouch@email.wm.edu) & Anya Lunden, William & Mary (lunden@wm.edu).

2. Word-final phonetic phenomena in situ. Another phonetic phenomenon, known to occur in situ word-finally, is word-final lengthening. Final lengthening is the phonetic phenomenon of the elongation of sounds at the right edge of words. Final lengthening, much like phonetic final devoicing, is found at all prosodic levels in varying amounts (Crystal and House 1988 and Lunden 2017 for English, Johnson and Martin 2001 for Creek, Lunden 2006 for Norweigan, Nakai et al. 2009 for Northern Finnish). In Figure 1, final lengthening and final devoicing are present on the final syllable of the nonce word ”basádafa” that was read in a sentence frame in which it was not in focus. (Data from Lunden 2011 recordings.)

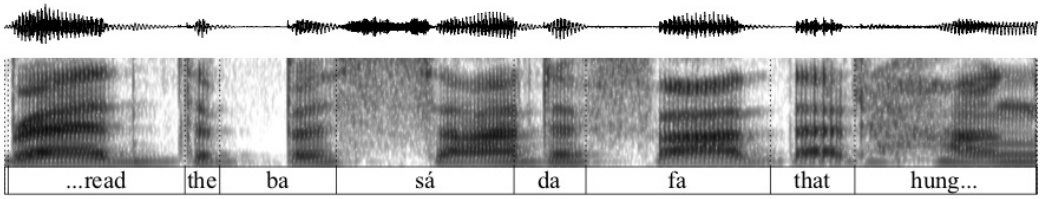

Figure 1. Nonce word “basadafa” read in a sentence frame in which it was not in focus. final lengthening and final devoicing evident on final, unstressed syllable “fa”

Final lengthening could theoretically be explained through the analogy theory as has been done for final devoicing; it occurs at all levels, but may have a main pressure to occur utterance-finally at the approach of a longer pause. Klatt (1976) compared to the “slowing down of a machine” as it prepares to stop. However, final lengthening, unlike final devoicing, does not have a phonological counterpart.[1] This would therefore imply either that analogy could also extend to phonetic phenomena, or that final lengthening must be natively present at all domain levels where it is found.

If final lengthening were present at all domain levels inherently, then it in turn could provide a motivation for final devoicing. Blevins (2004) suggests that final lengthening may provide a phonetic push for final devoicing. In this theory, the elongation of the final sound(s) within a word provides a phonetic push for breathy voice to occur, as it is harder to hold out the voicing of a sound, especially an obstruent, the longer it goes on. 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-internally 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).

Multiple studies have found phonetic amounts of lengthening and/or devoicing at the word-level; e.g. phonetic final devoicing found word-finally in Northern Finnish by Nakai et al. 2009. The fact that these phenomena occur at the right edge of words is a challenge to the analogy hypothesis, which requires phonologization to have occurred at a higher prosodic level before that phonological phenomenon is transferred to the word-level.[2] On the other hand, if phonetic final lengthening is native to the word-level, and phonetic final devoicing can follow

[1] The exception to this is cases where phonological final lengthening occurs on final vowels as opposed to the more-common final shortening, and which seems to be a phonologization of final lengthening, e.g. Tigre (Palmer 1962) and Scottish English (Wells 1982, Scobbie et al. 1999).

[2] One might also wonder how analogy is able to target the right edges of words if this is specifically taken to not be a phonological domain, as is sometimes stated in the analogy hypothesis (Hualde et al. 2015).

from it, then there is possible motivation for phonological devoicing natively at the word-level.

One way to probe the inherent existence of the right edge word boundary is by examining whether it is a salient area to listeners. If word-final phonetic cues are helpful to listeners, then this would be further evidence that phonetic-level phenomena are congruent with occurring in this position, and are not only the purview of higher prosodic domains. If phonetic-level lengthening and devoicing are truly part of the word-final environment then the analogical hypothesis is not needed to explain the presence of phonological alternations word-finally.

3. Artificial language learning. In artificial language learning (ALL) experiments, participants hear a small number of nonce words repeated as a speech stream (typically about two minutes for infant participants, and about seven minutes for adult participants. Participants are subsequently played the individual nonce words amongst other test words and asked to identify whether they were words in the speech stream they listened to, typically with a ranking on a scale for adult subjects and preferential head-turn procedure for infants. It has been shown that both infants and adults can complete this task fairly well with just the use of transitional probabilities (TP) for the unique syllables, using these statistics to parse words out of the speech stream of the artificial language (e.g. Saffran et al. 1996). However, it has been demonstrated that when the words are of different lengths, for example consisting of disyllabic and trisyllabic words, this statistical probability-taking skill on the part of infants breaks down as the probabilities become too complex to track (Johnson and Jusczyk 2003, Johnson and Tyler 2010). The same is not true of adults, as they have been found to be capable of this task with words of different lengths (Tyler and Cutler 2009). This incongruity between adults and infants is not understood, as infants are typically better at tasks related to language-learning than adults.

Since in real life infants do successfully parse words of different lengths from speech streams, one possibility as to why they fail to do so in experiments is the lack of prosodic cues that would be present in normal speech. We know that prosody helps with parsing, and in studies with adults the presence of final lengthening has been shown to help in ALL tasks (Kim et al. 2012 for Korean and Dutch speakers parsing trisyllabic AL words; Tyler and Cutler 2009 for English, Dutch, and French speakers on trisyllabic and quadrisyllabic AL words; Saffran et al. 1996 for English speakers). It has been noted that while final lengthening is a universally present phenomenon, there may be some effect of a listener's native language on the usefulness of in word-parsing; for example, Ordin et al. (2017) demonstrate that for Basque and German speakers final lengthening assists learners but the same is not true for Italian speakers. They suggest that this may have to do with the placement of stress in Italian, which may be a more useful word-boundary cue in that language.

While final lengthening has been demonstrated to provide a boost in parsing AL words out of a speech stream for adult speakers of various languages, final devoicing has not been tested in the same way thus far. The present set of studies seeks to determine whether phonetic word-final phenomena assist in word-parsing in ALL tasks, specifically if they can help above the level of TPs alone when the words are of different lengths, and how their simultaneous presence (which is more true to language produced naturally) affects listeners' ability to parse a speech stream.

4. Methods for word-final phonetic phenomena ALL studies.

4.1. THREE-SYLLABLE STIMULI STUDY. PARTICIPANTS. 100 native English-speaking undergraduate students from ages 18-23 years old (F=78; average age=18.8) at William & Mary participated in this study for participation pool credit.

STIMULI. To create stimuli for the three-syllable study's experiments, nonce artificial language (AL) words of three syllables were constructed from 18 unique CV syllables, made up of six consonants ([l, r, b, ʃ, z, k]) and six vowels ([ɑ, i, e, o, u, ə]). Each consonant was 90 ms. in length and each vowel had a baseline of 110 ms. in length. No C or V repeated within a word in the learning or the testing phase and each C and V was used in every position (initial, penult, final) among the words. Words were generated in MBROLA (Dutoit and Pagel 1995), using the us1 voice, and then modified as necessary in Praat (Boersma and Weenink 2019) for each of five conditions. The conditions were: (1) transitional probability (TP) alone, (2) final lengthening (FL) where the final vowel of the word was 150% the length of that same vowel in any other position in a word), (3) final devoicing (FD) where the final vowel of the word was given a breathy quality for the last 50% of its normal-length duration, (4) final lengthening (150%) with devoicing (25%), and (5) final lengthening (150%) with a larger amount of final devoicing (50%).

For conditions 2, 4, and 5 which had FL, the final vowel was given more length in MBROLA and then clipped to the exact correct length (165 ms.) in Praat. For conditions 3-5 which had FD, the devoicing effect was synthesized by first creating a V_ihV_i sequence in MBROLA for each word-final vowel, as the [h] then had the quality of that vowel. Subsequently in Praat, this [h] was run through the stop Hann band filter in Praat, set to filter out 0-500 Hertz. This [h] served as the devoiced portion of the vowel and was spliced into the word-final syllable. In order to achieve a more natural gradient change in intensity, the voiced portion of the vowel was split into thirds, where the second and third portions were progressively lowered in intensity.

To create the learning phase stimulus that participants would be listening to, a random list of 100 strings of the numbers one through six were generated. Each real word of the AL was assigned a number one through six, and the words were then concatenated in this order in Praat, controlling for repetition of the same word in a row across the strings. The same order was used for all five conditions. This string was copied and self-concatenated in Praat for a total of 6.5 minutes of stimulus. Because the stimulus length was kept consistent, the words were heard somewhat fewer times each in the conditions with final lengthening.

The 18 words for the testing phase consisted of the plain version of the six real words (i.e. without final lengthening or final devoicing), six part-words which were syllable strings that crossed word boundaries (and therefore participants heard at times in the speech stream), and six non-words which were made of the same syllables but in orders never heard in the speech stream. All were generated in MBROLA using the same us1 voice as for the stimulus words. The restriction that no C or V be duplicated within a word was also true within the part-words and non-words (or "not real" words to encompass both).

PROCEDURE. Participants were played a speech stream of an artificial language for 6.5 minutes. They listened through Sennheiser HD 280 pro headphones in a sound-attenuated booth, having been told that they would hear words from a made-up language strung together and not to do anything but passively listen.

During the test-phase, participants rated 18 test words in an MFC task. Participants rated each on a five-point scale for how sure they were or were not that the word was one they had or had not just heard (from "certain it was" (5) to "certain it wasn't" at (1)).

4.2. VARIABLE-LENGTH WORD STIMULI STUDY. PARTICIPANTS. 60 native English-speaking undergraduate students from ages 18-21 years old (F=36, gender nonconforming=1; average age=18.8) at William & Mary participated in this study for participation pool credit.

STIMULI. The stimuli creation process in the variable-length word study was nearly identical to that of the three-syllable word study. Here, nonce artificial language (AL) of two, three, and four syllables were constructed from 18 unique CV syllables, made up of the same six consonants ([l, r, b, ʃ, z, k]) and six vowels ([ɑ, i, e, o, u, ə]), all of the same baseline lengths as in the first study. No syllable was put in the same position within the words twice across the inventory. Words were again generated in MBROLA using the us1 voice and then modified as in three-syllable word stimuli study in Praat for the three conditions. The variable-length word conditions were: (1) TP, (2) FL, and (3) FD. The FL and FD qualities were created in the same way for this study as in the first study, using a combination of MBROLA and Praat, and again the order of words within the stimulus, which was once again 6.5 minutes, was kept consistent across conditions. Eighteen test words were constructed in the same way, where the sets of part-words and non-words had two two-syllable words, two three-syllable words, and two four-syllable words.

PROCEDURE. The procedure in the variable-length word study was identical to that of the three-syllable word study.

5. Results. Study 1, with three-syllable word stimuli, was run in order to provide a baseline for the results of Study 2, with variable-length word stimuli. The graph in Figure 2 shows the five conditions of Study 1. We see that regardless of condition, participants rated the words they heard in the stimulus more highly than either kind of not real word.

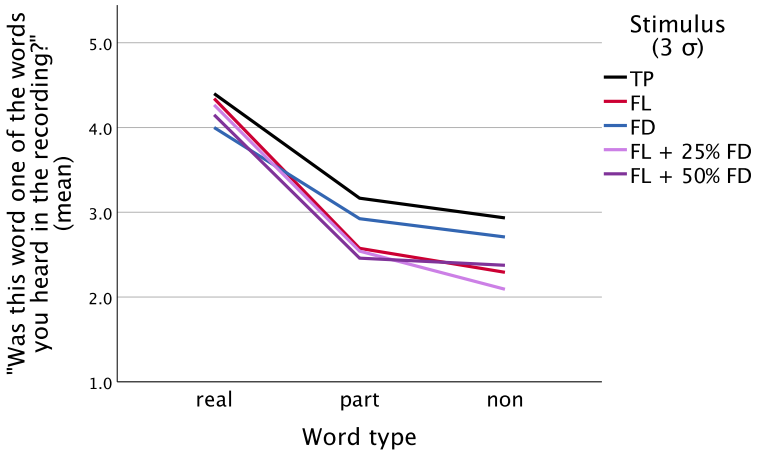

Figure 2. Results from Study 1: three-syllable word stimuli

Neither condition that combined final lengthening with final devoicing performed better than the final lengthening condition alone. Therefore Study 2 was run with only the independent TP, FL, and FD conditions. The graph in Figure 3 shows the three conditions (TP, FL, FD) across both studies.

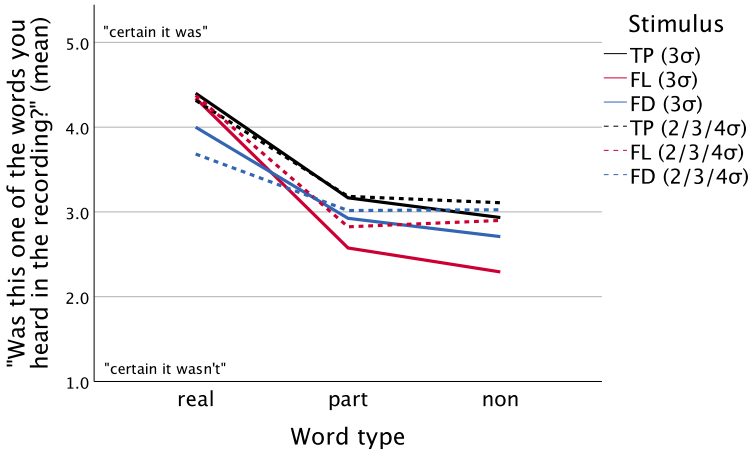

Figure 3. Results from TP, FL, FD conditions in Study 1 and Study 2

A generalized linear mixed model was fit on the ordinal response variable in SAS for these 6 iterations together, with the independent variables *study* (three-syllable words, variable-syllable words), *condition* (TP, FL, FD), and *type* (real, part, non), and was blocked by *subject*.

There was not a significant interaction between the three IVs ($F = 0.74; p = 0.5649$), nor between *study* and *condition* ($F = 1.23; p = 0.2968$). There was a significant interaction of *study* and *type* ($F = 9.23; p = 0.0001$), meaning that the difference between the mean responses to each type of word does differ by study. This is due specifically to the effect of *study* on responses to non-words ($F = 13.35; p = 0.0003$). There was no significant difference in the responses to real ($F = 2.58; p = 0.1090$) or part ($F = 1.40; p = 0.2378$) words between the two studies.

There is also a significant interaction between *condition* and *type* ($F = 11.15; p < 0.0001$). Because of the lack of a three-way interaction, we know that the differences between the different conditions hold for both studies. As a high mean response to real words is correct and a low mean response to part-words and non-words is correct, the greater the effect on *type* the more helpful the condition was to listeners. The effect on *type* is greatest in the FL condition ($F = 157.58; p < 0.0001$, cf. TP: $F = 85.25; p < 0.0001$, FD: $F = 49.32; p < 0.0001$).

We can see that overall, participants were worse at rejecting non-words in the variable-length words study than in the three-syllable word study. We see visually that in the FL condition of Study 1 that participants did the best at both accepting real words and rejecting not real words, which is consistent with the statistical finding that *type* varies the most in the FL condition across both studies. We do not see FD particularly enhancing performance, and participants in fact do worse at recognizing the real words of the study in the FD condition than they do with TPs alone.

6. Conclusion. The fact that we see word-level final lengthening significantly improving participant ability to parse words from the speech stream supports the theory that the right-edge of the word may inherently carry phonetic-level cues. Final devoicing, however, was not found to significantly improve parsing. We note that, as a dampening effect (i.e making softer already-present phonetic content), final devoicing may not be as helpful as final lengthening, which is an enhancing effect (i.e. providing more phonetic content). Final devoicing may still inherently exist at the word-level as a phonetic by-product of natively-present final lengthening, given

Blevin's (2004) hypothesis that final devoicing is coupled with final lengthening.

The next question to answer is whether final lenglthening present at the ends of words in an artificial language would help infants to parse above the level of transitional probabilities alone when all words were kept the same length, or if final lengthening would allow them to parse words of different lengths in an ALL task, which they cannot do with TPs alone. It has been demonstrated that prosodic phenomena can assist infants in the parsing of sentences (Morgan 1996); therefore it may be the case that a word-edge marker like final lengthening may assist them in word-parsing as well. If this were true, it would provide even stronger evidence that listeners are sensitive to the right edge of the word and make use of word-level prosodic phenomena. If such evidence were found, it would give further weight to the proposal that word-level phonetic phenomena can themselves turn into a phonological alternation.

Appendix. The following are the words that were used to make the stimuli used in each study, as well as the part-words and non-words used in conjunction with the real words in each testing phase.

real	kɑʃobu	biləre	ʃəruli	zukiʃɑ	rozekə	lebɑzo
part	bɑzoʃə	ʃobule	liroze	ʃɑbilə	rulikɑ	kəlebɑ
non	ʃobiru	kɑzuro	ləʃɑzo	bukile	zeʃəbɑ	relikə

Table 1. Testing stimuli in Study 1: three-syllable words

real	bɑru	ʃilo	rəkebi	zuliʃɑ	lerozɑkə	koʃəbuze
part	lobɑ	ruko	bɑruʃi	buzerə	bileroza	kəzuliʃɑ
non	rukə	zɑko	ʃɑlebi	loʃəbu	bɑzeʃiro	rəkezuli

Table 2. Testing stimuli in Study 2: variable-length words

References

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Crystal, Thomas & Arthur House. 1988. Segmental durations in connected-speech signals: Syllabic stress. *Journal of the Acoustical Society of America* 83(4). 1574–1585.

Hock, Hans Henrich. 1991. *Principles of historical linguistics*. Berlin: Walter de Gruyter.

Hock, Hans Henrich. 1999. Finality, prosody, and change. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.), *Proceedings of LP 98*, 15–30. Prague: Karolinum Press.

Hualde, José Ignacio, Christopher & Sarah Little. 2015. Final devoicing in Castilian Spanish. Talk given at MidPhon 20, University of Indiana.

Hualde, José Ignacio & Christopher Eager. 2016. Final devoicing and deletion of /-d/ in Castilian Spanish. *Studies in Hispanic and Lusophone Linguistics* 9(2). 329–353. https://doi.org/10.1515/shll-2016-0014.

Johnson, Elizabeth K. & Peter W. Jusczyk. 2003. Exploring statistical learning by 8-month-olds: The role of complexity and variation. Jusczyk Lab final report; 141–148.

Johnson, Elizabeth K. & Michael D. Tyler. 2010. Testing the limits of statistical learning for word segmentation. *Developmental Science* 13(2). 339–345. https://doi.org/10.1111/j.1467-7687.2009.00886.x.

Johnson, Keith & Jack Martin. 2001. Acoustic vowel reduction in Creek: Ef-

fects of distinctive length and position in the word. *Phonetica* 58. 81–102. https://doi.org/10.1159/000028489.
Keating, Patricia A. 1988. A survey of phonological features. Indiana University Linguistics Club.
Kim, Dahee, Joseph D.W. Stephens & Mark A. Pitt. 2012. How does context play a part in splitting words apart? production and perception of word boundaries in casual speech. *Journal of Memory and Language* 66(4). 509–529. https://doi.org/10.1016/j.jml.2011.12.007.
Klatt, Dennis. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59(5). 1208–1221.
Lieberman, Philip. 1967. Intonation, perception, and language. MIT Research Monograph.
Lunden, Anya. 2006. *Weight, final lengthening and stress: A phonetic and phonological case study of Norwegian*. Santa Cruz: University of California dissertation.
Lunden, Anya. 2017. Duration, vowel quality, and the rhythmic pattern of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1). 1–20. https://doi.org/10.5334/labphon.37.
Morgan, James L. 1996. Prosody and the roots of parsing. *Language and Cognitive Processes* 11(1-2). 69–106.
Myers, Scott & Jaye Padgett. 2014. Domain generalisation in artificial language learning. *Phonology* 31. 399–433. https://doi.org/10.1017/S0952675714000207.
Nakai, Satsuki, Sari Kunnari, Alice Turk, Kari Suomi, & Riikka Ylitalo. 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37(1). 29–45. https://doi.org/10.1016/j.wocn.2008.08.002.
Ordin, Mikhail, Leona Polyanskaya, Itziar Laka & Marina Nespor. 2017. Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory and Cognition* 45(5). 863–876. https://doi.org/10.3758/s13421-017-0700-9.
Palmer, Frank R. 1962. *The morphology of the Tigre noun*. London: Oxford University Press.
Saffran, Jenny R., Elissa L. Newport & Richard N. Aslin. 1996. Word segmentation: The role of distributional cues. *Journal of Memory and Language* 35(4). 606–621.
Scobbie, James M., Nigel Hewlett & Alice E. Turk. 1999. Standard English in Edinburgh and Glasgow: the Scottish vowel length rule revealed. In Paul Foulkes & Gerard Docherty (eds.), *Urban voices: Variation and change in British accents*, 230–245. Arnold.
Tyler, Michael D. & Anne Cutler. 2009. Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America* 126(1). 367–376. https://doi.org/10.1121/1.3129127.
Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.