2025. Proc Ling Soc Amer 10(1). 5959. https://doi.org/10.3765/plsa.v10i1.5959.

Phonetic duration is more variable than phonological duration

Anya Hogoboom & Joseph Lorber*

Abstract. Duration may be phonologically meaningful, as with contrastive segment length or as a cue to stress, or it can be purely phonetic, as with final lengthening. This paper explores the amount of durational variation present in the articulation of vowels that have different sources of duration. We show evidence to support the hypothesis that phonological duration is more stable and phonetic duration is more variable. We present evidence from new analyses of two English production studies (Lunden 2016, 2017) with nonce words that show significantly greater durational variability due to final lengthening than durational differences due to vowel quality or stress. Variability in duration was calculated by looking at the residuals from generalized linear models of raw duration. Subsequent homogeneity of variance tests (Levene's tests) were performed as part of one-way ANOVAs on the saved residuals as the dependent variable. We suggest that the greater variability of phonetic duration is plausibly responsible for effects such as the avoidance of stress and avoidance of long vowels and geminates word-finally, as it is perceptually difficult to signal duration-based phonology in a position with highly-variable duration.

Keywords. phonology; phonetics; duration; variance

1. Introduction. This paper explores the hypothesis that the type of duration source affects how variable the duration will be. We assume two possible sources of duration: phonological and phonetic. Phonological sources are those where duration is a manifestation of some phonological property, for example, vowel quality (lower vowels being longer), vowel or consonant quantity (long vowels or geminates being longer), or the realization of stress in languages in which duration is a cue to stress (stressed syllables being longer). Phonetic sources of duration do not cue, or cash out, any phonological property. Final lengthening, which occurs at all prosodic levels (Oller 1973), is a type of phonetic duration. We propose that phonological duration is less variable, whereas phonetic duration results in greater variability. We explore this hypothesis through new analyses of two English production studies with nonce words that were previously analyzed for the effects of position and stress on duration and vowel quality by Hogoboom in Lunden (2016) and Lunden (2017).

Word-final syllables typically show effects of final lengthening, but, despite their increased duration they do not robustly act like a durationally-enhanced rime. For example, long vowels often draw stress (the Weight-to-Stress principle of Prince (1983)) but final syllables often eschew stress, an avoidance that is typically encoded as final-syllable extrametricality (Liberman & Prince 1977). Myers & Hansen (2007) discuss the fact that many languages that contrast vowel length do not do so word-finally (that is, there is a pressure toward what is often termed "final shortening"), and it is also less common to find consonant length contrasts word-finally (e.g., Dmitrieva 2017). Hogoboom & Lorber (2023) conducted a series of perception studies which show that listeners are less sensitive to duration changes word-finally than they are non-

* We thank the audience members at the 2025 Annual Meeting of the Linguistics Society of America for helpful questions and observations. We also thank Kim Love for advising us on a statistically-testable measure of variance and the participants of the Linguistics Research Meetings at William & Mary for helpful feedback. Authors: Anya Hogoboom, William & Mary (ahogoboom@wm.edu) & Joseph Lorber, William & Mary (jblorber9@gmail.com).

finally, offering further evidence that duration from final lengthening behaves differently. Hogoboom & Lorber (2023) suggest that finally-lengthened rimes do not behave like phonologically-lengthened rimes because of the phonetic source of the duration. In this paper we propose that the reason the type of duration source matters is because of a difference in how reliably consistent the duration is, and that it is more consistent when there is a phonological source and less consistent when it has a phonetic source.

In §2 we discuss the measure of durational variability used. The reanalysis of data from Lunden (2016) for durational variability is presented in §3 as Study 1, and the reanalysis of data from Lunden (2017) is presented in §4 as Study 2. The findings are summarized and discussed in §5 and §6 concludes.

2. Variation expectations. There is a general tendency for greater variability with longer intervals of any kind (tapping, drawing, etc). Turk & Shattuck-Hufnagel (2020) list both non-speech as well as speech timed behaviors that are known to increase in variability as they increase in duration. All else being equal, we would therefore expect a long vowel's duration to show greater variability than a short vowel's, and, without any pressure to constrain the variability, as which might be exhibited by the phonology, we expect to see a linear-like increase: as the target duration increases, the variability of the duration will increase.[1]

The measure of durational variation used here is the residuals of a linear regression of duration. The residuals are the absolute value of how far each data point is from the linear fit line, which tells us how variable the data is. Figure 1 shows illustrative examples of smaller residuals, as when happens when the linear regression is a better fit to the data, and larger residuals, as what results from when the linear regression is a less close fit. Red lines have been drawn between each data point and the linear fit line to illustrate the distance being measured.

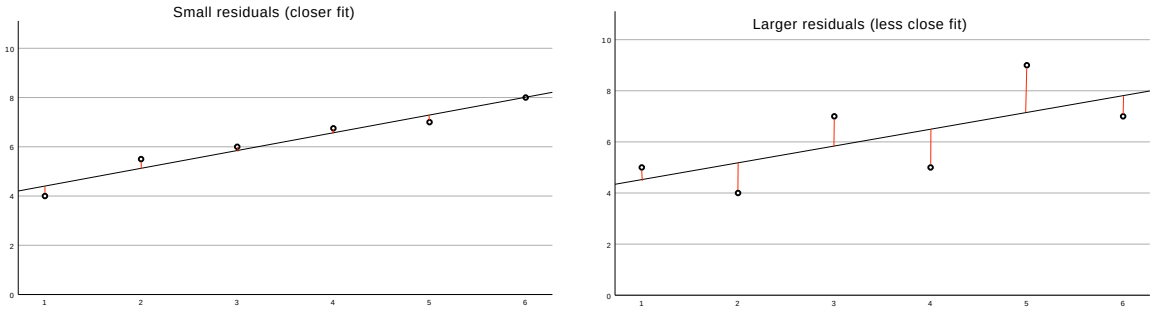

Figure 1. Examples of smaller and larger residuals from linear fit lines

We test the variability of duration by performing homogeneity of variance tests (Levene's tests) on the saved residuals, as part of one-way ANOVAs with residuals as the dependent variable and the relevant duration factor as the independent variable. We also visually inspect scatter plots of the standard deviation of the residuals plotted against the mean duration of various rime types to see which kinds of rimes exhibit the defaultly-expected linear increase and which do not.

[1] The speech examples cited and discussed by Turk & Shattuck-Hufnagel generally involve final domains, and so are consistent with the hypothesis put forward here.

3. English nonce word Study 1. Lunden (2016) reports on the duration and vowel quality of vowels in nonce words of the shape CV.CV́.CV.CV read by seventeen native English speakers. The nonce words were each written with one of <i>, <u>, or <a> in all syllables, with the antepenultimate syllables written in capitals to indicate stress on that syllable. The consonants in the words were always one each of [b], [d], [f], and [s]. The nonce words were read in question/answer pairs, as exemplified in (1).

(1) Example written stimulus
Which baDAfasa did her brother notice?
Her brother noticed the baDAfasa that smelled funny.

Speakers were given examples of the target stress pattern, and practiced it, with the words *America* and *asparagus*. The words in the answer portion of the question/answer pairs were subsequently delineated in Praat (Boersma & Weenink 2015). The beginning and end of each vowel was identified based on a combination of the onset/offset of regular waveform cycles and of the formants. Instances in which the speaker stumbled or hesitated were excluded, as were the few cases where a non-final unstressed vowel was produced without a regular waveform. Pronunciations where there was a clear phrase boundary after the target word were also excluded. For the current re-analysis, all of the sound files included in the 2016 analysis were re-listened to, and re-inspected, and further exclusions were made, such that all included measurements come from words that were clearly phrased together with the following relative clause.[2] At the time of the original measurements, the final vowel intensity was calculated as a percentage of stressed vowels' intensity, and was averaged by subject. The analysis excluded those subjects (=5) that had notably higher relative intensity on the final vowel as this was taken to be indicative of producing secondary stress on the final syllable, leaving the seventeen subjects whose data is used in both the 2016 and the current analyses.

The box plot in Figure 2 shows the range and medians of the durations of the vowels in each of the four positions. We see non-final unstressed vowels showing the expected shorter durations than the stressed antepenultimate vowels, and we also see the presence of word-level final lengthening. The ranges in the plot already indicate a wider range of word-final vowel durations than is found non-finally.

Figure 3 shows the standard deviation (SD) of the residuals of a linear regression[3] plotted against average duration for each of the three vowel qualities in each of the four positions. The colors indicate the orthographic vowel (shades of green =<i>, red =<u>, blue =<ɑ>) and shapes indicate position/stress (circles and rectangles representing non-final unstressed vowels, triangles representing stressed, diamonds representing (unstressed) word-final). Circles have been added to label the various sub-groupings: unstressed non-final vowels realized as [ə], stressed vowels, and word-final vowels. Both stressed and word-final vowels are realized with differing vowel qualities ([i], [u], [ɑ] in stressed syllables, [i], [u], [ə] in word-final syllables).

Looking first at vowel quality, we see that, among the stressed syllables, [ɑ] is longer on average (the blue triangle compared to the green and red ones). A homogeneity of variance test that

[2] Thank you to the attendee of the LSA who pointed out that relative clauses can often trigger a boundary that is indicated only through pitch and duration and not with a following pause.

[3] A generalized linear model (GLiM) run on *vowel_duration* as the dependent variable with independent variables *position*, *vowel*, and their interaction term *position*vowel*, and with *subject* as a blocking factor.

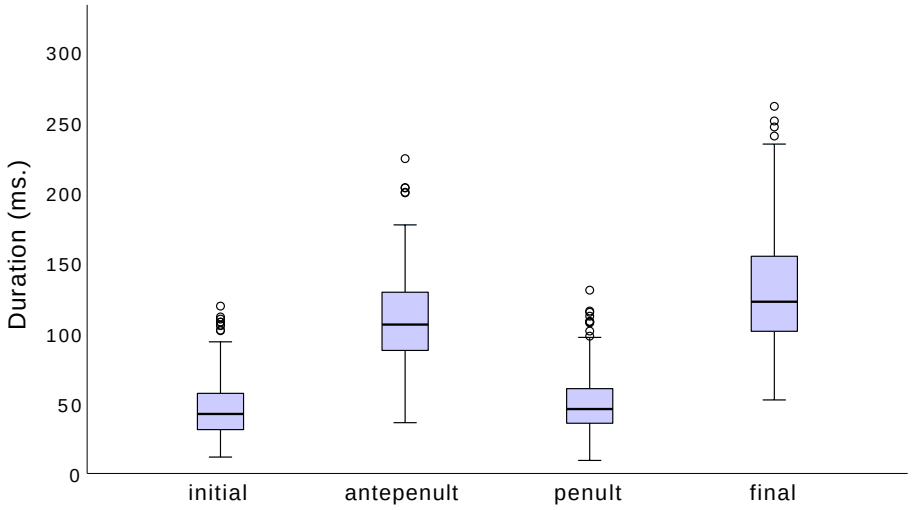

Figure 2. Durations of vowels in nonce words with <i>, <u>, and <a>

was part of a one-way ANOVA run on the three antepenultimate vowels shows no statistically significant difference in variability (Levene's test $= 0.188, df1 = 2, df2 = 347, p = 0.829$). The fact that durational variability does not increase with duration among the stressed vowels can be seen on the scatterplot, as the three stressed vowels are at the same level of the y-axis. Likewise, a homogeneity of variance test that was part of a one-way ANOVA run on the three final vowels finds no difference among the durational variability of the three vowel qualities in word-final position (Levene's test $= 0.429, df1 = 2, df2 = 347, p = 0.652$). Thus, while we see the expected tendency for lower vowels to be longer (through visual inspection of the graph in Figure 3), we do not see the expected-as-a-default linear increase in variability accompanying their longer duration. Rather, the variability seems to be suppressed when it accompanies the phonological property of vowel quality.

Figure 3 also shows non-final unstressed [ə]s in words with orthographic <u> as somewhat longer than those in words with orthographic <i> or <a>, and in fact we find a statistically significant difference in variability among the corresponding [ə]s (Levene's test $= 5.380, df1 = 2, df2 = 697, p = 0.005$, from one-way ANOVA run on the six non-final unstressed vowels, with IV *vowel*). While it is not clear what this durational difference is due to, it is clear that the difference between the <u>-source [ə] and the others must be phonetic, as no phonological distinctions are present in non-final unstressed vowel qualities.

Thus, among the vowel quality differences, we see the lack of an increase in variability when the duration increases due a phonological source (contrastive vowel quality) but we see a linear increase when the duration has a phonetic source (gradient vowel quality).

Turning to duration that is due to stress, we can see visually that the stressed vowels (the triangles) are on the same horizontal line as (the high end) of the [ə]s (circles and rectangles). Thus, the additional duration due to stress also does not result in notably increased variability. Comparing the variability between unstressed initial and stressed antepenultimate vowels, we

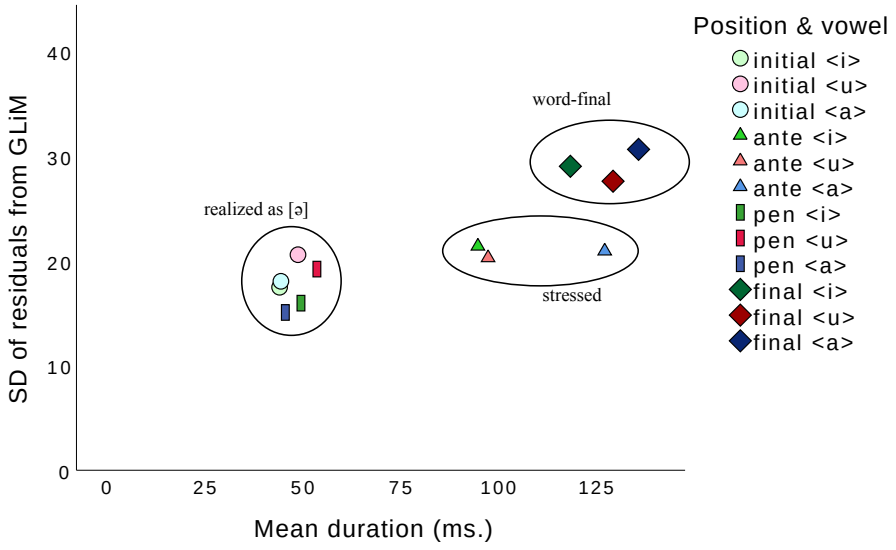

Figure 3. SD of residuals plotted against duration for 3 vowel qualities in 4 positions

find no to marginal statistical significance ($p \geq 0.041$).[4] This is especially noteworthy because stressed vowels are much longer than unstressed [ə]s and yet the duration of stressed rimes is not more variable than the duration of non-final unstressed vowels.[5]

The word-final vowels (the diamonds) however, do show a notable increase in variability along with their longer duration. Notice that the final vowels have a similar average duration to a stressed [ɑ] (∼125 ms.) but the higher SD of their residuals shows their greater durational variability. And indeed we find a statistically significant difference in variability between unstressed word-final and stressed antepenultimate vowels (Levene's test $= 28.418, df1 = 1, df2 = 698, p < 0.001$, from one-way ANOVA run on the six longer syllables, with IV *position*).

4. English nonce word Study 2. Lunden (2017) presents a similar study to the one summarized and re-analyzed in §3. In this study the data comes from eleven native English speakers who produced the same type of four-syllable nonce words, this time all with orthographic <a>, but under both antepenultimate and penultimate stress. In this version of the study, speakers first read all the question/answer pairs with the nonce words marked for one of the two stress patterns, and then transitioned to a second list marked for the other stress pattern, where the order of the stress pattern was reversed for each speaker, with real-word practice before each set. The vowels in the nonce word in the answer portion were subsequently delineated in Praat, using the same procedure as is reported for the first study in §3. Like the data in Study 1, all sound files were re-examined for the current analysis, and further exclusions were made, to ensure all included mea-

[4] <i>: Levene's test $= 3.745, df1 = 1, df2 = 190, p = 0.054$; <u>: Levene's test $= 0.005, df1 = 1, df2 = 246, p = 0.942$; <a>: Levene's test $= 4.226, df1 = 1, df2 = 258, p = 0.041$

[5] When comparing unstressed penultimate and stressed antepenultimate vowels there is a significant difference in variability for <i> ($p = 0.020$) and <a> ($p < 0.001$). However, homogeneity of variance tests find no difference for any vowel qualities between the two non-final unstressed positions ($p \geq 0.103$), supporting the grouping of the initial and penultimate positions. Further, the only robust difference is between penultimate and antepenultimate <a> and even here we do not see anything like a linear increase between the two in Figure 3.

surements come from words that were clearly phrased together with the following relative clause. At the time of the original measurements, the same procedure as was described for Study 1 was followed to identify and exclude speakers who may have placed a secondary stress on the final vowel of nonce words with antepenultimate stress (=4).

A box plot of the durations of the vowels in the four positions, under both stress conditions, is shown in Figure 4.

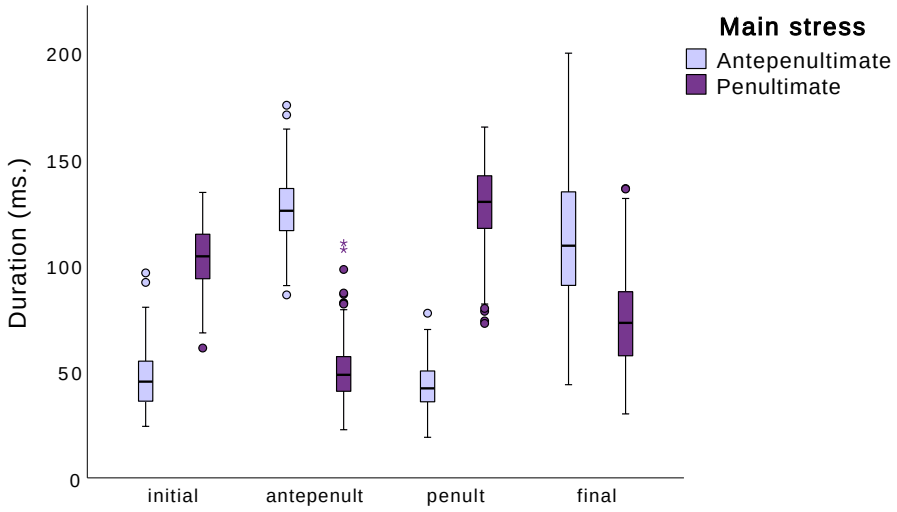

Figure 4. Durations of vowels in 4-syllable nonce words with <a>

We again see all non-final unstressed vowels showing the expected shorter durations compared to the stressed vowels, and continue to see the presence of word-level final lengthening, although to a more extreme degree in the antepenultimate stress pattern.[6]

Figure 5 shows the SD of the residuals of a linear regression[7] plotted against the mean duration of each of the four positions under each stress pattern. Shapes again indicate prominence level, where non-final unstressed vowels are circles, stressed vowels are triangles, and unstressed word-final vowels are diamonds. As we do not have the vowel quality differences (apart from [ɑ]∼[ə]) that we did in the previous study, color families merely enhance the prominence level groups (unstressed non-final are blues, stressed are neutral tones, and word-final are warm). Circles have been added to label these three sub-groupings.

Looking at duration due to stress, we again see the same pattern we did in the data from Study 1: While stressed vowels are longer, they show a similar level of duration variance as non-final unstressed vowels. This is supported by a homogeneity of variance test (from a one-way ANOVA run on the six non-final syllables, each as their own factor level) which did not find a statistically significant difference (Levene's Statistic $= \ 1.852, df1 \ = \ 5, df2 \ = \ 843, p \ =$

[6] Lunden (2017) leaves open the question of whether final lengthening is phonetically enhanced under antepenultimate stress to continue the alternating rhythm or whether this is in fact the 'normal' level of final lengthening, which is suppressed to some degree when adjacent to a penultimate stress.

[7] A GLiM run with DV *vowel_duration* and IVs position (4 levels), *main_stress* (2 levels), the interaction term *position*main_stress*, blocked by *subject*.

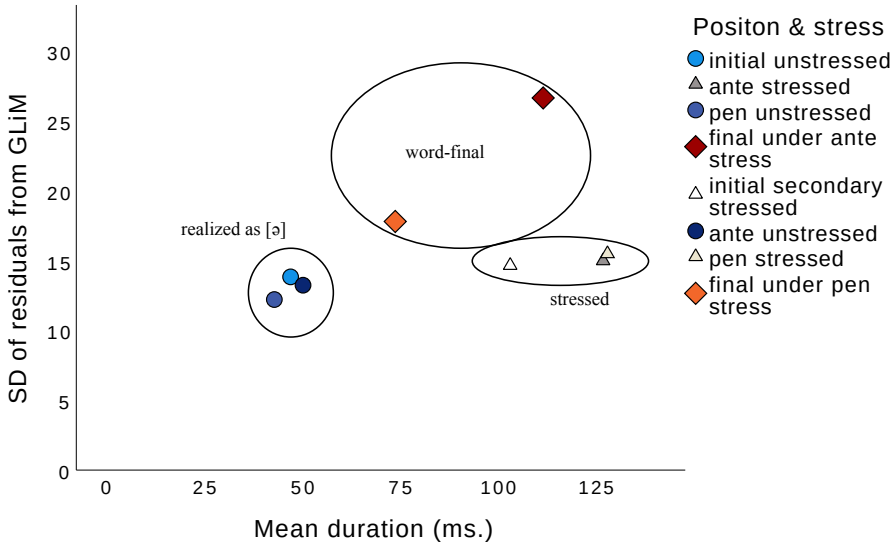

Figure 5. SD of residuals plotted against duration for 4 positions under 2 stress patterns

0.100). While vowels bearing primary stress (the beige and gray triangles in Figure 5) can be seen to be notably longer than vowels bearing secondary stress (the white triangle), we do not find significantly difference variance between these three stressed syllables (Levene's Statistic $= 0.282, df1 = 1, df2 = 459, p = 0.595$, from one-way ANOVA with IV *stress_type*, comparing the primary-stressed vowels to the secondary-stressed vowel).

Looking at the two types of word-final vowels, we again see the linear increase between duration and variability, which is even more pronounced due to having both a shorter type of final lengthening (final syllable under penultimate stress) and a longer type of final lengthening (final syllable under antepenultimate stress), both falling along a linear increase from the baseline non-final unstressed vowels. If we run a homogeneity of variance test on all eight types of syllables (each as their own factor level) we find a statistically significant difference (Levene's Statistic $= 16.935, df1 = 7, df2 = 1124, p < 0.001$; cf. the non-significant result when run on the six non-final syllables).

5. Discussion. The two sets of re-analyzed data give duration measurements for vowel qualities and stress/position conditions. The four-syllable nonce words with antepenultimate stress in Study 1 (Lunden 2016) allow the comparison of different vowel qualities, whereas the four-syllable nonce words with fixed orthographic <a> in Study 2 (Lunden 2017) allow comparison of each position under antepenultimate and penultimate stress. Notably, the degree of word-level final lengthening is different between the two stress patterns, which allows a further test of whether word-final lengthening, as a phonetic source of duration, results in a linear increase of duration and durational variability.

Figures 3 and 5 both show stressed vowels having notably more duration than non-final unstressed vowels but also show this additional duration not resulting in a corresponding increase in variability. Both also show word-final syllables, which are unstressed but show additional duration due to final lengthening, exhibiting increased variability along with their increased duration. That is, stressed syllables show only an increase in duration, whereas word-final syllables show

an increase in duration and variability. Additionally, Study 1 involved different vowel qualities and while we see that [ú] is, as expected, longer than [í] or [ú] it does not show significantly more variability.

Together, the studies show two types of durational increases due to phonological sources: vowel quality (Study 1) and stress (unstressed vs. stressed in Study 1; unstressed vs. secondary stress vs. primary stress in Study 2). None of these comparisons show a robust increase in variability to match the increase in duration. The hypothesis put forward here is that the otherwise-expected increase in variability is suppressed in cases of increased duration that has a phonological source. The examined source of phonetic duration was word-final lengthening, which was present in both studies and consistently showed a linear increase of duration and variability, including for both levels of word-final lengthening present in Study 2. An additional type of phonetic duration was found in Study 1: non-final unstressed [ə]s that were the pronunciation of an orthographic <u> were longer, both in initial and penultimate positions, than [ə]s from either <i> or <a>. The cause is unknown, but, as there is no contrast possible between the three in either position, the duration must have a phonetic, rather than phonological, source. These longer [ə]s in fact show a linear increase in variability along with their longer duration.

We therefore propose that while increased variability would naturally co-occur with increased duration, this variability is suppressed when the duration has a phonological source. It is not suppressed when the duration has a phonetic source.

If we accept that phonetic-source duration is highly variable, then we have an explanation for why word-final syllables behave differently in many respects. The phonological properties of stress, vowel length, and consonant length, all of which are known to have a pressure to avoid the final syllable (although all can occur word-finally cross-linguistically) need to create a contrast through duration. If a lengthened syllable is highly variable in its duration then it is not a reliable place for durationally-based phonological information. Hogoboom & Lorber (2023) note that contour tones are an exception to this pattern, which Yip (1989) noted are specifically drawn to word-final syllables. Zhang (2004) proposed it is specifically the longer duration of word-final syllables that attract the contour tones, as they need time to be realized. In this case, the phonological information is present in the pitch, and longer duration is just the host for it.

Further work needs to be done to expand our knowledge of the relevant behavior of different types of phonological and phonetic sources of duration. As part of this investigation we plan to undertake a typology of contrastive vowel length and consonant length variability. We would expect a language which has, for example, contrastive vowel length but not contrastive consonant length to show greater durational variability among consonant durations than among vowel durations, whereas this is should not be true in a language that contrasts both vowel and consonant length, where we would expect the durational variability of both to be controlled. We are also interested in identifying other phonetic sources of duration; one promising candidate would be the same rime in words of different lengths (i.e. polysyllabic shortening).

6. Conclusion. In examining the average duration and durational variability of different vowels in different positions under different stress patterns (Figures 3 and 5), we see a general tendency for a linear increase: as duration gets longer, variability gets higher. However, the increase in variability is instead constrained when the duration comes from a phonological source (vowel quality or stress) compared to when it comes from a phonetic source (final lengthening). Our finding that word-final duration is highly variable, and much more variable than duration that

encodes phonological information, offers an explanation for why final syllables are often dispreferred for signaling phonological information such as stress or segment length, as the greater durational variability makes them poor places to realize phonological contrasts.

References

Boersma, Paul & David Weenink. 2015. Praat: doing phonetics by computer. Computer program. Version 5.4.22, retrieved 8 October 2015. http://www.fon.hum.uva.nl/praat/.

Dmitrieva, Olga. 2017. Production of geminate consonants in Russian. In Haruo Kubozono (ed.), *The phonetics and phonology of geminate consonants*, 34–65. Oxford University Press.

Hogoboom, Anya & Joseph Lorber. 2023. Word-final strength and weakness. *Proceedings of the Linguistic Society of America (PLSA)* 8(1). 5562. https://doi.org/10.3765/plsa.v8i1.5562.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2). 249–336.

Lunden, Anya. 2016. Effects of position and stress on English and Norwegian vowel quality. M.s. William & Mary.

Lunden, Anya. 2017. Duration, vowel quality, and the rhythmic pattern of English. *Laboratory Phonology* 8(1). 1–20. https://doi.org/10.5334/labphon.37.

Myers, Scott & Ben Hansen. 2007. The origin of vowel length neutralization in final position: Evidence from Finnish speakers. *Natural Language and Linguistic Theory* 25(1). 157–193. https://doi.org/10.1007/s11049-006-0001-7.

Oller, D.K. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 54(5). 1235–1247.

Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14(1). 19–100.

Turk, Alice & Stefanie Shattuck-Hufnagel. 2020. Timing evidence for symbolic phonological representations and phonology-extrinsic timing in speech production. *Frontiers in Psychology* 10. Article 2952. https://doi.org/10.3389/fpsyg.2019.02952.

Yip, Moira. 1989. Contour tones. *Phonology* 6(1). 149–174.

Zhang, Jie. 2004. Contour tone licensing and contour tone representation. *Language and Linguistics* 5(4). 925–968.