INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

An acoustic study of vowel undershoot in a system with several degrees of prominence

Janina Mołczanow[1], *Beata Łukaszewicz*[1], *Anna Łukaszewicz*[1]

[1]University of Warsaw, Poland

jmolczanow@uw.edu.pl, b.lukaszewicz@uw.edu.pl, a.b.lukaszewicz@student.uw.edu.pl

Abstract

The paper presents the results of a pilot study investigating the relationship between vowel quality and duration in Ukrainian. In this language, lexical stress is cued by increased duration; smaller but systematic differences in length occur between unstressed, rhythmic stress-bearing, and pretonic syllables. The presence of several degrees of lengthening within one word makes it possible to test the long-established theories of vowel reduction posing a direct link between decreased duration and vowel undershoot. Overall, the analysis of the aggregated data collected from four native speakers of Ukrainian points to a strong correlation between decreasing duration and the undershoot of F1 targets. However, in separate by-position and by-speaker analyses, no correlation between F1 and duration is observed in the positions of rhythmic and lexical stress. We thus conclude that the stability of the F1 target *vis-à-vis* temporal parameters may constitute another parameter expressing metrical prominence. In addition, our data suggests that formant undershoot may be affected by an articulatory effort.

Index Terms: vowel reduction, vowel undershoot, rhythmic stress, Ukrainian, acoustic analysis

1. Introduction

The relation between vowel duration and formant structure has been well documented. Numerous studies of phonetic reduction have demonstrated that decreased duration of unstressed vowels results in the overall compression of the vowel space, manifested by the formant undershoot (e.g. [1], [2], [3], [4], [5]). Previous research has also shown that F1 undershoot can occur in the absence of reduced duration, in which case it is usually attributed to the reduction in articulatory effort ([6], [7], [8]). Most of these studies investigated the relation between duration and F1 by changing the position of the stimulus within the utterance or by altering speech tempo (e.g. [8], [9]). The current study investigates vowel undershoot in Ukrainian, a language exhibiting several degrees of lengthening within one word ([10], [11], [12]). These durational differences render Ukrainian a perfect testing ground for the theories of phonetic reduction as a duration-dependent undershoot.

Ukrainian has contrastive lexical stress. According to standard descriptive sources as well as recent phonetic studies ([10], [11], [12], [13]), lexically stressed syllables are much longer than unstressed syllables; according to [10], [11] – by a factor of 1.5. In addition, vowels in Ukrainian are lengthened in pretonic positions (i.e. positions immediately preceding lexical stress), and in positions of rhythmic stress (secondary stress occurring in every second syllable, but blocked by the vicinity of lexical stress; [11], [14]). (We use here the terminology commonly adopted in the literature on Slavic prosody, e.g. [15], where 'tonic' refers to primary lexical stress, while 'pretonic' points to positions immediately preceding main stress.) The structure of a quadrisyllabic word with lexical stress located on the final syllable is shown in (1).

$\sigma_1\sigma_2\sigma_3\sigma_4$ ˌ*velosyˈped* 'bicycle' (1)

σ_1 - rhythmic stress
σ_2 - unstressed
σ_3 - pretonic lengthening
σ_4 - lexical stress

Comparisons carried out in segmentally identical pairs of words differing only in prosodic conditions, show that the pretonic vowel (σ_3) is on average only 7.5 ms longer than the initial vowel (σ_1) with rhythmic stress ([14]). Both are significantly longer than other (i.e. unstressed) syllables (σ_2). It should be noted that despite similar duration, pretonic lengthening and rhythmic stress are phonologically different phenomena – the former is associated with the primary (lexical) stress, while the latter is conditioned by the grammatical (rhythmic) stress (see [11] for discussion).

Similarly to Ukrainian, duration serves as the main exponent of stress in other Slavic languages with lexical stress, e.g. Russian. However, Ukrainian is different from Russian in that it does not exhibit phonological vowel reduction. It is noted in [16] that though the decrease in the duration of unstressed syllables is accompanied by qualitative reduction in Ukrainian, this process is non-categorical and does not interact with other phonological phenomena. As yet, the relationship between stress and formant frequencies in Ukrainian has not been systematically investigated. According to [16, p. 176], there is a qualitative difference between vowels in the first and the second pretonic positions – the former being usually centralised while the latter being closer in quality to the stressed vowel. However, given that pretonic syllables are longer than other unstressed syllables, the presence of centralisation in the former but not in the latter is problematic in view of the numerous studies demonstrating that formant undershoot is correlated with decreased duration of unstressed syllables. The current pilot study addresses this issue by investigation the correlation between vocalic formants and duration in different prosodic contexts.

2. Methods

2.1. Stimuli

The aim of the experiment has been to provide formant and duration measurements of the vowel /a/ in several prosodic positions in Ukrainian. The stimuli used in the study contained the vowel /a/ in four prosodic contexts, comprising one lexically stressed ('tonic') and three prosodically weaker

http://dx.doi.org/10.21437/Interspeech.2019-1806

positions, including the pretonic position, the initial (rhythmic stress) position, and the second (unstressed) position. All the words had the structure [σσσˈσ(σ$_{1\text{-}2}$)], with lexical stresses always falling on the fourth syllable. Although as indicated by (σ$_{1\text{-}2}$) the tonic syllable could be followed by 0-2 syllables, in most items, it was followed by 1 syllable. 12 words were recorded for each prosodic condition (initial - second - pretonic - tonic). In addition, 12 word tokens containing the stressed vowel /i/ and 12 tokens with the stressed vowel /u/ were included. These data were necessary to perform the normalisation of the acoustic space for each speaker (see below). The identity of consonants adjacent to the target vowel was controlled in terms of the place of articulation. That is, the vowel /a/ occurred in the following consonantal contexts in the same number of words in each condition: velar_coronal, labial_coronal, coronal_coronal, coronal_labial, coronal_velar. The structure of the stimuli is shown and exemplified below.

Table 1: *Structure of test items.*

Prosodic position		Example
initial	σ̲σσˈσ(σ)	*k̲atali ˈzator* ‘catalyst’
second	σσ̲σˈσ(σ)	*okat̲o ˈlyčyty* ‘convert to Catholicism’
pretonic	σσσ̲ˈσ(σ)	*delika̲ ˈtesy* ‘dainty’
tonic	σˈσ̲/ˈσ̲σ	*pro ˈka̲t* ‘lease’, *ˈkasa* ‘cash desk’

2.2. Data collection

The data was collected from four native speakers (2 females, 2 males) of standard Ukrainian, who are life-long residents of the Drohobyč area in Western Ukraine. Participants were asked to read target words embedded in a frame (*Skažete ... druhyj raz* ‘You (pl.) will say … for the second time’). Lexical stress was marked orthographically to facilitate the identification of words. The words put in a frame were presented on a computer monitor, and the list was randomised to avoid order effects. The recordings were performed using an H4 Zoom portable recorder, set at a sampling rate of 44.1 kHz, and an AT897 microphone. In sum, 864 word tokens were recorded (4 speakers × 72 words × 3 repetitions).

2.3. Measurements and normalization

All formant frequencies were measured at acoustic midpoints, to ensure that all data were measured in a consistent way ([17]). The data were balanced with respect to the places of articulation occurring in the preceding and following consonants to control the effect of asymmetrical consonantal contexts. One such context, labial_coronal, is illustrated in Fig. 1. (Arrows point to acoustic midpoints.)

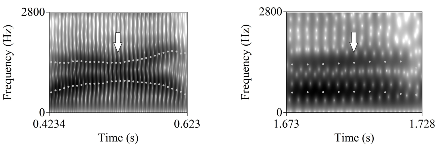

Figure 1: *The F1 and F2 trajectories of the vowel /a/ in the tonic (left panel) and unstressed (second) syllable (right panel) in the labial_coronal context. (Data from speaker f2.)*

The measurements were conducted in *Praat* (using the Burg LPC algorithm). The ceiling of the formant search range was 5000 Hz for male speakers, and 5500 Hz for female speakers. In order to establish connection with human perception, formant frequencies were expressed on the mel scale ([17], [18]). To ensure data comparability, all formant values were translated into a normalized acoustic space ([17]). The normalization procedure consisted in establishing the centre of the vowel space for each speaker and rescaling the data from this speaker by the average distance of each data point from the centre. This was possible because apart from tokens with the vowel /a/, we also collected data containing the peripheral vowels /i, u/ in exactly the same consonantal contexts. The normalization steps are illustrated in Fig. 2-4 with the data from one speaker (m1). First, analogously to [17], the centre of the vowel space was defined as the grand mean of the medians of the three corner vowels; for speaker m1 in Fig. 2, the centre falls at F1 = 595 mels and F2 = 1251 mels. Next, the median distance between each token of the /a/ vowel and the speaker’s vowel space centre was calculated; the unit length for the data points shown in Fig. 3 amounts to 102 mels. Finally, a normalized acoustic space was created by subtracting the grand mean from each data point and dividing the result by the unit length. The data from speaker m1, now translated into the normalized acoustic space, are shown in Fig. 4.

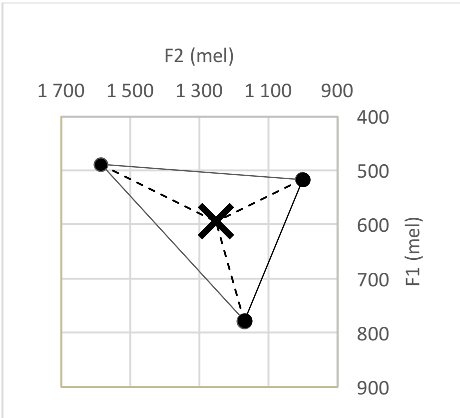

Figure 2: *The grand mean (marked as a cross). (Data from speaker m1.)*

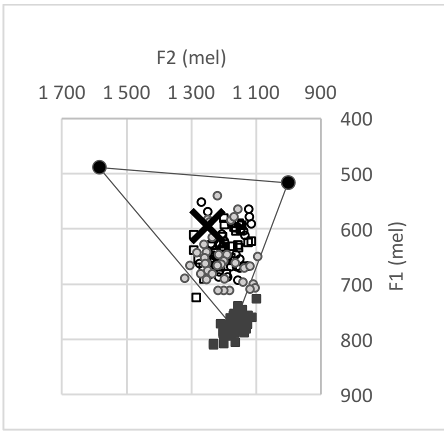

Figure 3: */a/ data points against the grand mean. (Data from speaker m1.)*

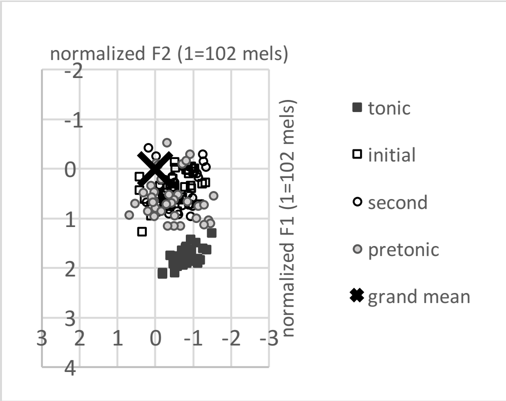

Figure 4: */a/ data points in the normalized acoustic space. (Data from speaker m1.)*

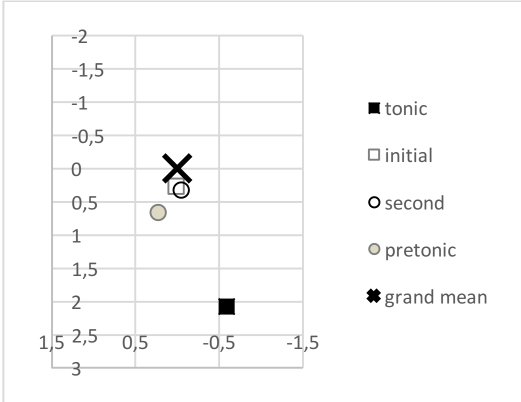

Figure 6: *Mean normalized F1-F2 values depending on prosodic position.*

3. Results

There was a strong positive correlation between F1 and duration; $r(576) = 0.85$, $p < .001$. A small negative correlation between F2 and duration was also found; $r(576) = 0.38$, $p < .001$. The relationship between duration and F1 points to the presence of vowel undershoot. However, as more detailed analyses presented below show, although the overall undershoot effect is robust, it follows from the concatenation of heteregenous subsets (speakers, positions), and not from a robust dependence of F1 on duration within particular prosodic positions, present in individual speakers. The impact of the subset analyses (cf. Fig. 5) is as follows. First, although initial (rhythmic stress) and second (unstressed) syllables exhibit the same ranges of vowel duration values, only the latter shows a weak, yet statistically significant, positive correlation between F1 and duration ($r(144) = 0.3$, $p < .001$). Second, although the tonic position shows the biggest correlation between F1 and duration of all positions when the data from all four speakers are concatenated ($r(144) = 0.51$, $p < .001$), the correlation effect disappears in separate by-speakers analyses; cf. the right bottom panel in Fig. 5. The duration-induced undershoot effect consistently appears in the pretonic position; it is present both in concatenated ($r(144) = 0.41$, $p < .001$) and individual by-speaker analyses.

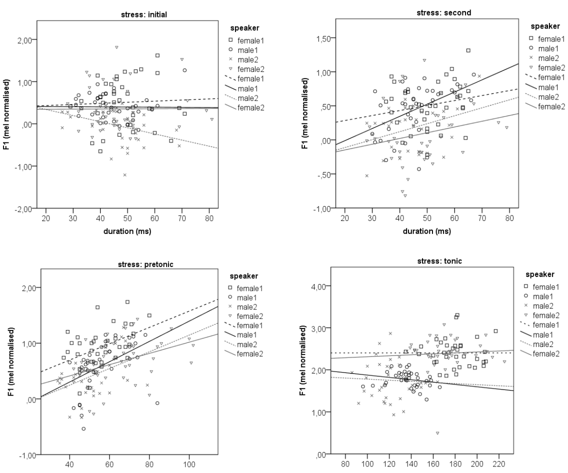

Figure 5: *The relationship between F1 and duration depending on position, divided by speakers.*

The effect of stress on vowel formants at acoustic midpoints was further tested. The analyses in terms of general linear models were conducted in SPSS (v. 25). To control for speaker-specific effects, apart from Stress, Speaker was also entered as a categorical predictor. Full factorial analyses were conducted (i.e. the Stress*Speaker interaction term was included).

For F1, all predictors turned out significant at the .001 level, with the model accounting for 79% of variance (R Squared = 0.785, Adjusted R Squared = 0.779). In post hoc tests (Tukey HSD), three significantly different homogeneous subsets emerged: (i) initial and second positions, (ii) pretonic, and (iii) tonic. Fig. 6 depicts the means for the four positions against the centre of the normalized vowel acoustic space. It is clear that although significant reduction marked by lower F1 values occurs in all non-tonic positions, the reduction is less conspicuous in the pretonic position than in initial and second positions.

For F2, all predictors were significant as well, but the model fit was smaller than in the case of F1, as the model accounted only for 35% of variance (R Squared = 0.348, Adjusted R Squared = 0.33). In post-hoc Tukey tests, the tonic position had significantly lower F2 values than the non-tonic positions; there was also a significant difference between the pretonic position and the second position ($p < .01$). The pretonic position stands out among other positions because it shows less centralisation than the initial and second positions, and, in general, is characterised by more fronting than all other positions (cf. Fig. 6).

By-speaker analyses revealed differences in F1 patterns in initial vs. second positions; see Fig. 7. Speakers f1 and m1 have relatively stable F1 lowering across the first two positions, pointing to a similar degree of reduction in these positions. Speakers f2 and m2 exhibited significant shifts in F1, rendering two different reduction patterns: speaker f2 had less reduction in the initial position than in the second position, speaker m2 had the opposite pattern. The existence of such differences in F1 patterns is interesting as for all four speakers there were no significant differences in duration between the initial and second positions. This tentatively shows that F1 lowering does not follow automatically from reduced duration and that factors such as articulatory effort might be involved (cf. [8]).

For F2, the interaction effect is mostly due to a bigger difference between the tonic position and the remaining positions in one speaker (f1), as well as the differences between speakers f2 and m2 as far as the initial and second positions are concerned; see Fig. 8. However, the overall reduction pattern is the same for all speakers. Vowels in the tonic position have significantly lower F2 compared to non-tonic vowels.

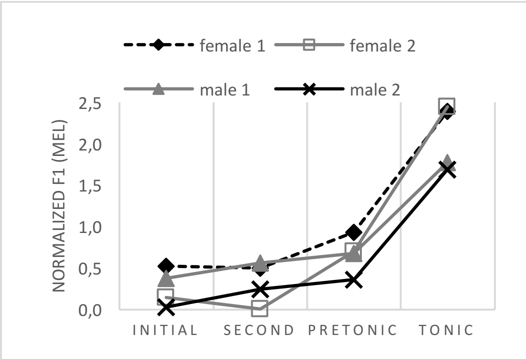

Figure 7: *By-speaker means of normalized F1 in four prosodic positions.*

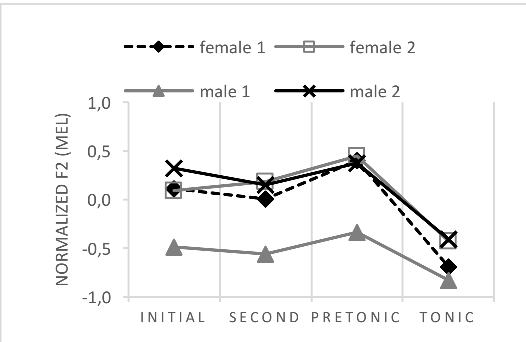

Figure 8: *By-speaker means of normalized F2 in four prosodic positions.*

4. Discussion

Previous studies investigated duration-dependent undershoot by placing the target vowel in different prosodic positions within a sentence ([1]), manipulating speech rate ([9]) or eliciting hyperarticulated speech ([8]). In Ukrainian, several levels of duration are present within one word, with tonic syllables being 1.5 times longer than their unstressed counterparts. Furthermore, differences in duration are attested among non-tonic syllables, where pretonic and initial positions are associated with increased length compared to the second syllable. This system is interesting because it does not have categorical vowel reduction, which often accompanies reduced length in systems displaying robust duration differences between stressed and unstressed syllables, e.g. Russian, Belarusian, Bulgarian (see [8] for further discussion).

The correlation between F1 and duration found in the present study points to the presence of vowel undershoot in Ukrainian. By-speaker and by-position analyses showed that the effect of duration on F1 is not automatic and depends on the prosodic structure. On the one hand, speakers who have faster renditions of the tonic vowels also have lower F1 values in these vowels. On the other hand, the tonic vowels are entirely immune to the undershoot effect in the sense that individual speakers show no duration-dependent F1 variability within their habitual duration ranges for the tonic position. The F1 independence of duration in the tonic vowel contrasts with the correlation between the two parameters in the immediately pretonic position, in which duration-dependent undershoot consistently appears for all speakers.

The present results do not align with earlier impressionistic descriptions of Ukrainian, which reported a qualitative difference between vowels in the first and the second pretonic positions [16]. Contrary to the claim in [16], we did not observe a greater degree of centralization of the vowel in the first pretonic position compared to the vowel in the second pretonic syllable. Quite the opposite, the first pretonic vowel exhibited less centralization than other unstressed vowels in the current study (see Fig. 6 for an illustration).

Previous research has demonstrated that increased duration cues rhythmic stress in Ukrainian ([10], [11], [14]). However, these studies concentrated on duration, intensity and pitch and did not investigate vowel quality as a potential variable contributing to the expression of rhythmic prominence. For all speakers in the current study, the initial and second positions do not differ in duration, but in terms of correlation between F1 and duration, which is non-significant in the initial (rhythmic stress) position, but significant in the second (unstressed) position. The lack of correlation between F1 and duration in the initial syllable renders this position similar to the tonic syllable, which also does not display duration-dependent F1 variability. Thus, besides duration, the stability of the F1 target may constitute another acoustic parameter employed to express prosodic prominence.

5. Conclusions

The findings of the current study demonstrate that the relation between duration and vowel reduction is more intricate than previously assumed. The analysis of Ukrainian data shows that not all durationally impoverished contexts warrant duration-dependent undershoot. No correlation between duration and formant undershoot is attested in metrically strong positions, suggesting that the stability of F1 targets in these positions serves to signal prosodic prominence. Needless to say, further study analysing data from a larger number of participants is needed to confirm these results.

6. Acknowledgements

Research was funded by National Science Centre (Poland), grant No. 2017/27/B/HS2/00780. We wish to thank the Ukrainian speakers for taking part in the experiment.

7. References

[1] B. Lindblom, “Spectrographic study of vowel reduction,” *The Journal of Acoustical Society of America*, vol. 35, no. 11, pp. 1773–1781, 1963.

[2] S.-J. Moon and B. Lindblom, “Interaction between duration, context, and speaking style in English stressed vowels,” *The Journal of the Acoustical Society of America*, vol. 96, no. 1, pp. 40–55, 1994.

[3] F. H. Guenther, “Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production,” *Psychological Review*, vol. 102, no. 3, pp. 594–621, 1995.

[4] M. Fourakis, "Tempo, stress, and vowel reduction in American English," *The Journal of the Acoustical Society of America*, vol. 90, pp. 1816–1827, 1991.

[5] F. J. Koopmans-van Beinum, *Vowel Contrast Reduction: An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions*. Amsterdam: Academische Pers, 1980.

[6] S.-J. Moon, *An acoustic and perceptual study of formant undershoot in clear and citation-form speech.* Ph.D. diss., University of Texas, Austin, 1991.

[7] E. Farnetani and Sh. Kori, "Rhythmic duration in Italian noun phrases: A study on vowel durations," *Phonetica*, vol. 47, no. 1-2, pp. 50–65, 1990.

[8] J. Barnes, "Phonetics and phonology in Russian unstressed vowel reduction: a study in hyperarticulation," unpublished manuscript, Boston University, 2007.

[9] J. Barnes, *Strength and Weakness at the Interface: Positional Neutralisation in Phonetics and Phonology*. Berlin: Mouton de Gruyter, 2006.

[10] B. Łukaszewicz and J. Mołczanow, "Leftward and rightward stress iteration in Ukrainian: Acoustic evidence and theoretical implications," in B. Czaplicki, B. Łukaszewicz, and M. Opalińska (eds.), *Phonology, Fieldwork and Generalisations*. Berlin: Peter Lang, 2018, pp. 259–275

[11] B. Łukaszewicz and J. Mołczanow, "Rhythmic stress in Ukrainian: Acoustic evidence of a bidirectional system," *Journal of Linguistics,* vol. 54, no. 2, pp. 367–388, 2018.

[12] B. Łukaszewicz and J. Mołczanow, "The role of vowel parameters in defining lexical and subsidiary stress in Ukrainian," *Poznań Studies in Contemporary Linguistics,* vol. 54, no. 3, pp. 355–375, 2018.

[13] M. F. Nakonečnyj, "Naholos," in I. K. Bilodid (ed.), *Sučasna ukrajins'ka literaturna mova*. *Vstup. Fonetyka.* Kyjiv: Naukova Dumka, 1969, pp. 358–369

[14] J. Mołczanow, B. Łukaszewicz, and A. Łukaszewicz, "Rhythmic stress or word-boundary effects? Comparison of primary and secondary stress correlates in segmentally identical word pairs," in *Proceedings of Speech Prosody 9*, 2018, pp. 676−679

[15] C. Y. Bethin, "Stress and tone in East Slavic dialects," *Phonology*, vol. 23, no. 2, pp. 125–156, 2006.

[16] N. I. Toc'ka, *Holosni fonemy ukrajins'koji literaturnoji movy.* Kyjiv: Vydavnyctvo Kyjivs'koho Universytetu, 1973.

[17] D. H. Whalen, W.-R. Chen, M. K. Tiede, and H. Nam, "Variability of articulator positions and formants across nine English vowels," *Journal of Phonetics*, vol. 68, pp. 1−14, 2018.

[18] S. S. Stevens, J. Volkmann, and E. B. Newman, "A scale for the measurement of the psychological magnitude pitch," *The Journal of the Acoustical Society of America*, vol. 8, pp. 185–190, 1937.