THE EFFECT OF FOCAL ACCENT ON VOWELS IN HUNGARIAN: ARTICULATORY AND ACOUSTIC DATA

Alexandra Markó[1,2], Márton Bartók[1,2], Tamás Gábor Csapó[2,3], Andrea Deme[1,2], Tekla Etelka Gráczi[2,4]

[1]Eötvös Loránd University, [2]MTA-ELTE „Lendület" Lingual Articulation Research Group, [3]Budapest University of Technology and Economics, [4]Research Institute for Linguistics HAS
marko.alexandra@btk.elte.hu, bartokmarton@gmail.com, csapot@tmit.bme.hu, deme.andrea@btk.elte.hu, graczi.tekla.etelka@nytud.mta.hu

ABSTRACT

In the present study, utterance-initial vowels in preverbal focus vs. pre-focal topic positions were compared with respect to their acoustic and articulatory parameters. Parallel acoustic, and ultrasound recordings were made, and vowel duration; maximum f_0, F_1 and F_2 (measured at the midpoint of the vowel); tongue contours (and their variability measured by the NND method) were compared with respect to the prominence level.

We predicted higher prominence in the case of focus compared to the topic. Accordingly, longer vowel durations and earlier f_0-peaks were found in the focus condition, on the other hand, neither the maximum f_0 values nor the articulatory measures of vowel quality showed differences between the conditions. Although on Euclidean distance data we found no effect of condition, the variance of F_2 values differed significantly across the conditions, which might be attributed to better reach of the articulatory target. Therefore this parameter needs further analysis.

Keywords: focal accent, topic, Hungarian, vowel, articulatory analysis, acoustic analysis

1. INTRODUCTION

Hungarian is an obligatory syntactic focus marking language, as prominent units typically appear in certain syntactic positions [10]. Hungarian sentences can be divided into a topic and a predicate part, but topic is not an obligatory part of the Hungarian sentence [5]. In sentences with a narrow focus, the focused element is placed between the topic and the verb (in this order). In the case of narrow focus, the focused constituent shows the highest prosodic prominence within the predicate part, while the ensuing elements are deaccented [16]. Although topic may be accented (as well as unaccented), theoretical works claim that its prominence cannot exceed that of the focused constituent [6], which means that (due to the left-headed prosody of Hungarian) the phrase-initial topic might be as prominent as the focus. Although, the relationship between the phonetic characteristics of prominence of (phrase initial) topic and (phrase initial) focus is ambiguous, (according to the knowledge of the authors) experimental data on this issue are not available for Hungarian.

With respect to the acoustic correlates of prominence in Hungarian, several studies revealed the role of intensity and f_0 (see, e.g., [8]), as well as f_0-peak alignment [15]. Vowel duration, however, was not taken into account in these analyses as a possible cue of prominence in Hungarian, although it plays a role in several languages. The question did not arise even due to the consensual claim that since vowel quantity is phonologically distinctive in Hungarian, it cannot play a role in prominence marking. Nevertheless, recent studies have found that longer vowel duration has a role in the expression of prominence [17, 18, 23].

Besides, as for Hungarian, there is an apparent consensus in the literature that vowel quality does not vary as a function of the presence/absence of prominence (which is also a common pattern in several languages). However, apart from a few earlier studies (see a review in [24]), which were largely inexplicit about the details of their methods, and a recent pilot study on a not well-balanced material [18], acoustic correlates of vowel quality, i.e., formant structure, have not been analysed reliably. Moreover, linguo-articulatory correlates of vowel quality in focal accent have not been analysed with respect to Hungarian either. The question of prominence-dependency of vowel quality especially arises because several models (e.g., [13]) suggest that longer segment duration (which might be a possible correlate of prominence) may lead to more accurate articulatory movements, and thus the gestural target of the segment might be better reached. On this basis, we may also assume that longer vowels in the more prominent position may also be articulated with greater force. Furthermore, an acoustic study [9] also revealed that vowels show smaller variability, if they are in a (lexically) stressed syllable (vs. unstressed), and [7] confirmed that the above effect also exists for higher level (sentential) accent, as well.

In order to fill the above mentioned gaps of the phonetic literature on Hungarian prominence, the present study's first aim is to compare the appearance of some of the possible phonetic correlates (vowel duration, and characteristics of f_0) of prominence between focus and topic, when they both occur in the same (phrase-initial) position in the sentence. Our second aim is to analyse vowel quality as a possible cue of prominence, both in the acoustic and the articulatory domain.

In the present study, utterance-initial vowels (Hungarian lexical units bear fixed stress on the first syllable) in topic vs. focus position were analysed and compared with respect to both acoustic and articulatory measures. Vowel duration was measured, and ultrasound tongue images, F_1, and F_2 were obtained from the temporal midpoint of the vowel. The value of the peak of f_0 and its alignment were also analysed.

We predicted higher prominence in the focus condition which induces longer durations and higher and f_0-peaks compared to the topic. F_0-peak alignment was expected to show differences between the focus and topic conditions, as well. We also hypothesized that formant values and variability of tongue contours differ in the two conditions, due to the greater force in the articulation of the vowels in the focus position.

2. METHOD

2.1. Material and participants

Four members of the Hungarian vowel-inventory were chosen for the analysis: front and high /i/, back and high /u/, front and low /ɛ/ and back and low /ɒ/ (in these examples the feature backness co-varies with lip spreading). From these vowels, V_1pV_1 structured words (/ipi/, /upu/, /ɛpɛ/, /ɒpɒ/) were constructed, in which we analysed the word-initial vowel (we used symmetrical V-context to control for the coarticulatory effect of the second vowel).

The (pseudo-)words were embedded into meaningful sentences, which were presented to the participants as answers to a question in short dialogues. Since two V_1pV_1 words out of the four have a meaning in Hungarian, we constructed sentences in which the words functioned as proper names, and the filler sentences were also constructed with (other) similarly structured nonsense "proper names". We analysed the target words in two conditions: they were positioned in pre-focal topic and in focus positions, both occurring sentence-initially. All target words were repeated 5 times.

Examples of the short dialogues of the experiment can be seen in (1) and (2), where the target vowel is indicated by bold.

(1) Ki nevettette meg Zazát? ('Who made Zaza laugh?')
*I***pi*$_{\text{FOCUS}}$ *nevettette meg*. ('Ipi made her/him laugh.')

(2) Miért olyan szomorú Opo? ('Why is Opo so sad?'
A*pa*$_{\text{TOPIC}}$ *nem*$_{\text{FOCUS}}$ *beszélte meg vele a hétvégi terveit.* ('Apa didn't tell him/her his weekend plans.')

The dialogues were presented on a computer screen in a randomised order. 20 female native Hungarian speakers (aged from 19 to 28 years, reported no hearing or speech deficits) were asked to read the question silently, and then to read the answer (the target sentence) aloud. With each participant, 40 target utterances (5 repetitions per each vowel in each condition) and 80 filler utterances (with the same dialogue and sentence construction) were recorded.

2.2. Procedure

Parallel (and synchronized) ultrasound and acoustic recordings were made. The tongue movement was recorded in midsagittal orientation using the "Micro" ultrasound system (Articulate Instruments Ltd.) with a 2–4 MHz / 64 element 20mm radius convex ultrasound transducer at 83 fps. The speech signal was recorded with an omnidirectional condenser microphone at 44.1 kHz sampling rate.

The annotation of vowel boundaries was carried out by forced alignment [21] and corrected manually in Praat [3], on the basis of the F_2 trajectory. In the present analysis only the fully modal voiced occurrences (424) were included. The distribution of the vowels was the following: /ɒ/: 59 in topic, 48 in focus, /ɛ/: 35 in topic, 34 in focus, /i/: 62 in topic, 45 in focus, and /u/: 66 in topic, 75 in focus.

Vowel duration, f_0 and formant frequencies were automatically extracted from the acoustic signal. The f_0 was measured at the maximum, and the position of f_0-peak within the vowel time course was extracted and given in the percentage of the vowel duration. F_1 and F_2 values were detected at the temporal midpoint of the vowel in Praat. Formant frequencies were standardized within speakers using *z*-transformation [14] in the phonR package [20]. On the basis of F_1 and F_2 data, the Euclidean distance of the centroid of the vowel space and each token was also calculated [4].

The ultrasound frames were extracted from the temporal midpoint of the vowel as raw scan line data and converted to PNG images. Tongue contours were manually traced on the PNG files using the APIL tracing tool [1]. Variability of the tongue contours was measured by the Nearest Neighbour Distance (NND [25]) method.

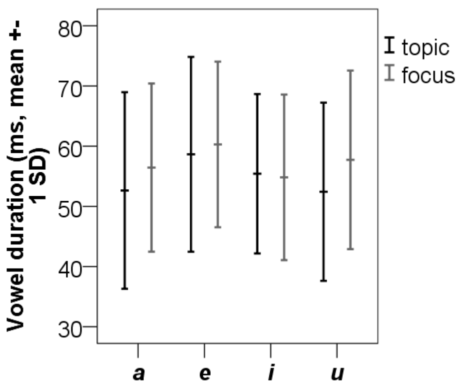
Duration, f_0, and NND data were analysed by linear mixed effect models (LMM) in R [22], using the lme4 package [2]. *p*-values were obtained via the Satterthwaite approximation available in lmerTest package [12]. We included random intercepts for speakers, and used vowel quality and condition (focus/topic) as fixed effects. Random slope models were also built and compared with the intercept (IC) model by anova in lmerTest package [12]. The two models did not show significant difference for any variables, therefore we introduce the results of the IC ones. Euclidean distances were compared using modified signed-likelihood ratio tests (MSLRTs) for equality of coefficient of variations [11, 19].

3. RESULTS

3.1. Vowel duration

In focus position, we found somewhat longer vowel realizations (/ɒ/ in topic: 53±16 ms, /ɒ/ in focus: 56±15 ms; /ɛ/ in topic: 59±16 ms, /ɛ/ in focus: 60±14 ms; /u/ in topic: 52±15 ms, /u/ in focus: 58±15 ms), except for /i/ (topic: 55±13 ms, focus: 53±15 ms) (the vowel duration data were checked, and proved to be valid, even if they seem to be very short) (Fig. 1). LMM confirmed a condition main effect on duration data ($F(1, 406) = 4.44$, $p = 0.036$) (while the effect of vowel quality was not significant).

Figure 1: Vowel duration as a function of condition and vowel quality (a = /ɒ/, e = /ɛ/, i = /i/, u = /u/) (mean ± 1 SD)

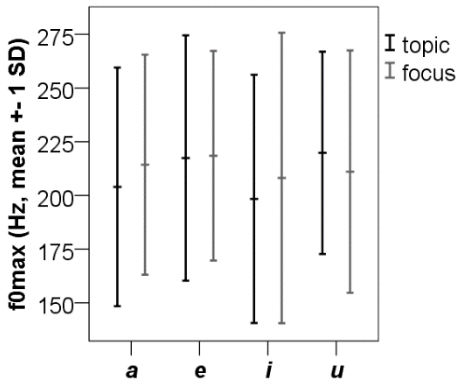


3.2. Fundamental frequency

3.2.1. Peak value of f_0

In focus position, f_0-peak was moderately higher in the focus condition in the case of /ɒ/ (topic: 204±56 Hz, focus: 214±51 Hz); /ɛ/ (topic: 217±57 Hz, focus: 218±49 Hz), and /i/ (topic: 198±58 Hz, focus: 208±68 Hz), while for /u/ we found higher f_0 in topic position (topic: 220±47 Hz, focus: 211±56 Hz) (Fig. 2). According to the LMM, however, on these maximal f_0 data, none of the tested factors had a significant effect.

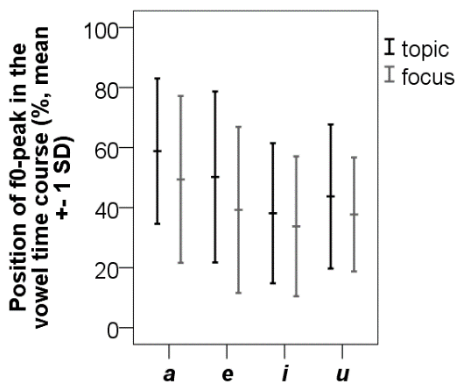
Figure 2: Maximum value of fundamental frequency as a function of condition and vowel quality (a = /ɒ/, e = /ɛ/, i = /i/, u = /u/) (mean ± 1 SD)



3.2.2. F_0-peak alignment

Although maximal values of f_0 did not differ with respect to the syntactical position, the alignment of the peak showed differences. In focus position, f_0-peak appeared earlier within the vowel time course than in topic position (Fig. 3). The data are expressed in the percentage of the vowel duration, i.e., the higher number represents a later f_0-peak: /ɒ/ in topic: 55±24%, /ɒ/ in focus: 47±26%; /ɛ/ in topic: 51±26%, /ɛ/ in focus: 36±26%; /i/ in topic: 41±24%, /i/ in focus: 35±22%, /u/ in topic: 42±23%, /u/ in focus: 40±21%.

Figure 3: F_0-peak alignment within the vowel as a function of condition and vowel quality



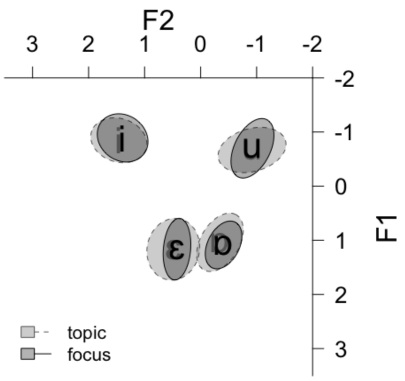
According to the linear mixed effects model, both the vowel quality and the condition played a significant role in the f_0-peak alignment (vowel: $F(3, 414.88) = 6.589$, $p < 0.001$; condition: $F(1, 414.54)$

= 10.978, p = 0.001). The effect of the interaction of the two factors was not significant.

3.3. $F_1 \times F_2$ space

Fig. 4 shows the standardized $F_1 \times F_2$ vowel space as a function of condition. Condition had no effect on Euclidean distances. The variance of F_1 values did not differ significantly across conditions either, but we found a significant difference in the variance of F_2 (MSLRT = 7.77, $p < 0.01$).

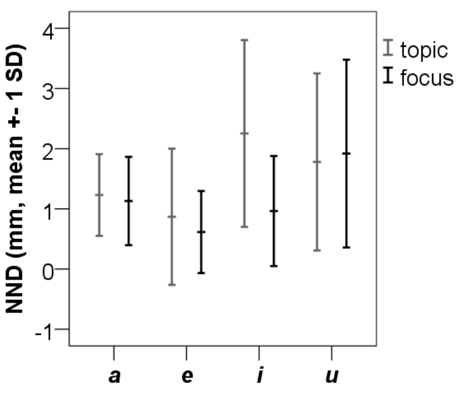
Figure 4: Standardised $F_1 \times F_2$ space of the analysed vowels as a function of condition



3.4. NND values

Distances of tongue contours were smaller in front vowels in focus position (/ɛ/ in topic: 1,6±1,1 mm, in focus: 1,2±0,5 mm; /i/ in topic: 2,7±1,2 mm, in focus: 1,5±0,7 mm), while back vowels did not differ in this respect as a function of the conditions (/ɒ/ in topic: 1,4±0,5 mm, in focus: 1,5±0,4 mm; /u/ in topic: 2,0±1,4 mm, in focus: 2,0±1,5 mm) (Fig. 5). LMM showed only the main effect of vowel quality ($F(3, 89) = 3.63$, p = 0.016).

Figure 5: NND values as a function of condition and vowel quality (a = /ɒ/, e = /ɛ/, i = /i/, u = /u/) (mean ± 1 SD)



4. DISCUSSION AND CONCLUSION

Hungarian is a left-headed prosody language. Although (according to the literature) topic cannot bear higher prominence than focus, utterance-initially both may be similarly prominent. In the present study, we expected that focus is more prominent than topic in the case of the same (phrase-initial) position. Evidence was found by the earlier literature that vowel duration and some of the f_0-parameters might be the acoustic correlates of prominence in Hungarian. Besides, vowel quality, which has so far not been analysed in a controlled fashion with respect to prominence in Hungarian, was also investigated. Not only acoustic (formant analysis) but also articulatory (analysis of ultrasound tongue images) measures were conducted, the latter one for the first time regarding Hungarian.

The results showed that focus position evoked longer vowel realizations than topic position, irrespective of the vowel quality. On the other hand, the maximum value of f_0 did not differ between the conditions, while the alignment of f_0-peak showed differences (in the focus condition f_0-peak occurred earlier).

Contrary to our predictions, longer duration did not evoke more peripheral articulation of vowels in the focus position; however, in this position, smaller variability was found (in the acoustic data). While Euclidean distances of the tokens from the vowel ellipse centroid did not differ as a function of condition, there was a significant difference in the variance of F_2, which might reflect less variable acoustics in the horizontal tongue position and/or lip-spreading dimension. However, NND values, which reflect tongue contour variability, did not confirm this effect in the articulatory domain.

Our study revealed that utterance-initial topic and focus show differences in some of the acoustic measures, which may be attributed to the higher prominence of focus. Vowel quality in general did not appear to differ between the conditions, however, the variance of F_2 was found to be smaller in the case of focus, which effect is needed to be analysed further.

Acknowledgements

We would like to thank Julianna Jankovics, Valéria Krepsz, Klaudia Kóródy, Bettina Száraz, and Zsófia Weidl for their help with conducting the experiments and analysing the data. Also, we are grateful to the anonymous reviewers for their valuable comments on earlier version of the paper.

5. REFERENCES

[1] APIL tracing tool https://github.com/arizona-phonological-imaging-lab/apil-web.

[2] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48.

[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43, http://www.praat.org/

[4] Bradlow, A. R., Torretta, G. M., Pisoni, D. B. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20, 255–272.

[5] É. Kiss, K. 2002. *The syntax of Hungarian.* Cambridge: Cambridge University Press.

[6] É. Kiss, K. 2006. Mondattan. In Kiefer, F. (ed), *Magyar nyelv*. Budapest: Akadémiai, 110–148.

[7] Farnetani E., Vagges, K., Magno-Caldognetto, E. 1985. Coarticulation in Italian /VtV/ sequences: A palatographic study. *Phonetica* 42, 78–99.

[8] Fónagy, I. 1958. *A hangsúlyról.* Budapest: Akadémiai Kiadó.

[9] Fowler, C. A. 1981. Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research* 46, 127–139.

[10] Genzel, S., Ishihara, S., Surányi, B. 2015. The prosodic expression of focus, contrast and givenness: A production study of Hungarian. *Lingua* 165, 183–204.

[11] Krishnamoorthy, K., Lee, M. 2014. Improved tests for the equality of normal coefficients of variation. *Computational Statistics* 29(1-2), 215–232. http://link.springer.com/article/10.1007/s00180-013-0445-2

[12] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13), 1–26. http://doi.org/10.18637/jss.v082.i13.

[13] Lindblom, B. 1963. Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 35, 1773–1781.

[14] Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am.* 49. (2), 606–608.

[15] Mády, K. 2015. Prosodic (non-)realisation of broad, narrow and contrastive focus in Hungarian: a production and a perception study. In *Proc. Interspeech* Dresden, 948–952.

[16] Mády, K., Kleber, F. 2010. Variation of pitch accent patterns in Hungarian. *Proc. Speech Prosody* Chicago, http://speechprosody2010.illinois.edu/papers/100924.pdf

[17] Mády, K., Reichel, U., Szalontai, Á. 2017. A prozódiai prominencia (nem)jelölése a németben és a magyarban. *Általános Nyelvészeti Tanulmányok* XXIX. 77–98.

[18] Markó, A., Bartók, M., Gráczi, T.E., Deme, A., & Csapó, T.G. 2018. Prominence effects on Hungarian vowels: A pilot study. *Proceedings of Speech Prosody 2018* (Poznan, Poland). https://www.isca-speech.org/archive/SpeechProsody_2018/pdfs/138.pdf

[19] Marwick, B., Krishnamoorthy, K. 2018. cvequality: Tests for the Equality of Coefficients of Variation from Multiple Groups. R software package version 0.1.3. https://github.com/benmarwick/cvequality

[20] McCloy, D. R. 2016. phonR: tools for phoneticians and phonologists. R package version 1.0-7.

[21] Mihajlik, P., Tüske, Z., Tarján, B., Németh, B., Fegyó, T. 2010. Improved recognition of spontaneous Hungarian speech: Morphological and acoustic modeling techniques for a less resourced task. *IEEE Transactions on Audio, Speech and Language Processing* 18(6), 1588–1600.

[22] R Core Team 2018. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

[23] Szalontai, Á., Wagner, P., Mády, K., Windmann, A. 2016. Teasing apart lexical stress and sentence accent in Hungarian and German. *Tagungsband 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum (P&P 12).* 216–219.

[24] Vértes O. A. 1982. A magyar beszédhangok akusztikai elemzésének kérdései. In: Bolla, K. (ed.), *Fejezetek a magyar leíró hangtanból.* Budapest: Akadémiai Kiadó, 71–113.

[25] Zharkova, N., Hewlett, N., Hardcastle, W. J. 2011. Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control* 15, 118–140.