HOW PROMINENCE AND PROSODIC PHRASING INTERACT

Uwe D. Reichel[1], *Katalin Mády*[1], *Felicitas Kleber*[2]

[1]*Research Institute for Linguistics, Hungarian Academy of Sciences*
[2]*Institute of Phonetics and Speech Processing, University of Munich*
{*uwe.reichel|katalin.mady*}*@nytud.mta.hu, kleber@phonetik.uni-muenchen.de*

Abstract: Prosodically left-headed languages like Hungarian show the tendency to locate relevant prosodic events at the start of prosodic phrases. In a production study we tested the strength of this principle by examining whether it is pertained even if a deviating prosodic structure is suggested by the stimulus setting. We measured duration and pitch-related features in front of prominent and non-prominent words in non phrase-initial position according to the stimulus suggestion. We found not only significantly higher prominence but also preceding boundary signals for the prominent words indicating that Hungarian speakers adjust prosodic phrasing to prominence requirements within an utterance in order to pertain left-headedness.

1 Introduction

Prosodic left- or right headed languages are characterized by a fixed prosodic structure. Prosodic structure is defined by prosodic phrasing and prominence, thus headedness refers to stable patterns of prominence relative to prosodic phrase boundaries. In prosodically strictly left-headed languages words are always stressed on the first syllable and relevant prosodic events tend to be located at the beginning of prosodic constituents. The classification of Hungarian as strictly left-headed [4, 7, 14, 9, 10] is based on various evidence: it has fixed word-initial lexical stress [14] defined as the potential location of phrase-level accents [15]. This leftmost location remains when combining words to prosodic phrases [7, 4, 14].

Prosodic structure is phonetically implemented by various boundary and prominence signals. Among the boundary signals are pauses [13], prefinal lengthening of phonetic segments preceding prosodic boundaries [16], and discontinuities in the fundamental frequency (F0) contour [2], generally pitch resets at the beginning of a new intonation phrase. Phonetic correlates of prominence are amongst others an increase in segmental duration and in intensity [5, 6] as well as prominence-lending F0 contours [9].

This study addresses the strength of the left-headedness principle for Hungarian by examining whether it is pertained even if a deviating prosodic structure is suggested by the content of the utterance. The actual prosodic structure is determined by measuring the above mentioned boundary and prominence signals.

2 Material and methods

2.1 Fruits in baskets production study

10 Hungarian native speakers (all females, age range from 18 to 33 years) took part in our production study, in which they had to name fruits of different size arranged in two baskets (see Figure 1). The recordings took place in an anechoic room of the Research Institute for Linguistics in Budapest. Prior to the training phase and the experiment, it was explained to the participants that we wanted to investigate emphasis. They were shown a fruit (e.g. a pear) on a paper, and they were asked to name it. Then they saw the same fruit as a large image, and they were asked to express the difference simply by uttering the same fruit name in a different way. They were not given any oral example by the experimenter in order to avoid any bias in their production.

The training phase and the experiment were implemented in the SpeechRecorder software [3]. Participants were first familiarized with the five fruits and their sequence that was fixed throughout all experimental images. They were explained that baskets can contain 2 or 3 fruits, and that each fruit can have normal or large size. The subjects were instructed to account for the grouping of the fruits within the two baskets and for the different sizes of the pieces.

At the end of the training phase, participants were told that during the experiment, they will always see two baskets containing fruits as shown in Figure 1. Within the first basket, fruits were either of the same size, or one of them was larger than the others. The task of the native Hungarian speakers was to name the fruits displayed on a screen from left to right. The first basket always contained the fruits *málna* (*raspberry*), *mango* (*mango*), and *alma* (*apple*), the second *mandula* (*almond*), and *mandarin* (*mandarin*). Thus this stimulus design enhances the production of a prosodic boundary located between the baskets and prosodic prominence on the large fruits. By this means we generated prosodic settings not compliant to the left-headed principle by placing prominent fruits X_2 in non-initial basket position (setting $x_1 - X_2 - x_3$, where each 'x' indicates a of fruit, the index gives its position in the basket, and capital letters indicate big-sized images). We compared the realization of $x_1 - X_2 - x_3$ with a neutral baseline $x_1 - x_2 - x_3$.

The prosodic realization of the contents of the second basket is not part of the present analysis, but in order to give a complete description of the setting, the fruits in the second basket were either of the form *x4 – X5* or *X4 – x5*, resulting in four combinations together with the first basket (large *X2* vs. baseline small *x2*). All sequences were named with 6 repetitions, resulting in a set of 240 utterances for 10 speakers.

Figure 1 - Fruits in baskets stimuli to trigger certain prosodic structures. Phrase boundaries are expected to be inserted between the two baskets. Large fruits are expected to receive greater prominence. **Left:** Baseline condition $x_1 - x_2 - x_3$ without prominent items in the first basket. **Right:** Condition $x_1 - X_2 - x_3$ with one non-basket-initial prominent item. Only the first basket is considered for the present study.

2.2 Hypothesis

We hypothesize, that in the deviant $x_1 - X_2 - x_3$ condition Hungarian speakers realize an additional boundary in front of the prominent fruit X_2 in order to pertain left-headedness.

2.3 Annotation and preprocessing

The data was manually text-transcribed and automatically signal-text aligned on the phone and the word level by WEBMAUS [12, 8]. F0 was extracted by autocorrelation (PRAAT 5.3.16 [1], sample rate 100 Hz). Voiceless utterance parts and F0 outliers were bridged by linear interpolation, and the contour was smoothed by moving median filtering with a window length of 6 samples and transformed to semitones relative to a base value. This base value was set to the F0 median below the 5th percentile of an utterance and served to normalize F0 with respect to its overall level.

2.4 Boundary and prominence features

To test our hypothesis of section 2.2 we measured duration and pitch-related features first on X_2 vs. x_2 to see whether there was a difference in the realization of prominence, and second between $x_1 - X_2$ and $x_1 - x_2$, respectively, in order to quantify whether the prominence of the second fruit in $x_1 - X_2$ is accompanied by a preceding boundary as opposed to $x_1 - x_2$. The features are summarized in Table 1.

Duration To capture prominence differences we compared the word duration as well as the duration of the first vowel between x_2 and X_2. To capture prosodic phrasing by pause insertion and by prefinal lengthening, the pause duration in front of x_2, resp. X_2, as well as the duration of the last vowel of x_1 were measured. No duration normalization was carried out, values are given in seconds.

Pitch discontinuity In order to quantify prosodic boundary strength for $x_1 - X_2$ and $x_1 - x_2$, respectively in terms of pitch discontinuity, we adopted the stylization method introduced in [11].

To capture F0 discontinuities with respect to level and range we fitted three base-, mid- and topline triplets to the F0 contours for x_1, for x_2 and for the window spanning both $x_1 - x_2$, respectively (for ease of reading x_2 represents both x_2 and X_2 in the subsequent stylization description, since stylization applies to both conditions).

The fitting of the lines is illustrated in Figure 2. Within each interpausal unit a window of length 50 ms is shifted along the F0 contour with a step size of 10 ms. Within each window F0 medians are calculated (1) from the values below the 10th percentile for the baseline, (2) from the values above the 90th percentile for the topline, and (3) from all values for the midline. Through each of these three median sequences a line was fitted by linear regression, yielding the base-, top- and midline, respectively.

The bottom right part of Figure 2 also shows the range stylization result which is simply derived by fitting a linear regression line through the point-wise distances between the base- and the topline. A negative slope means that base- and topline converge, whereas the positive slope in the illustrated example reflects line divergence.

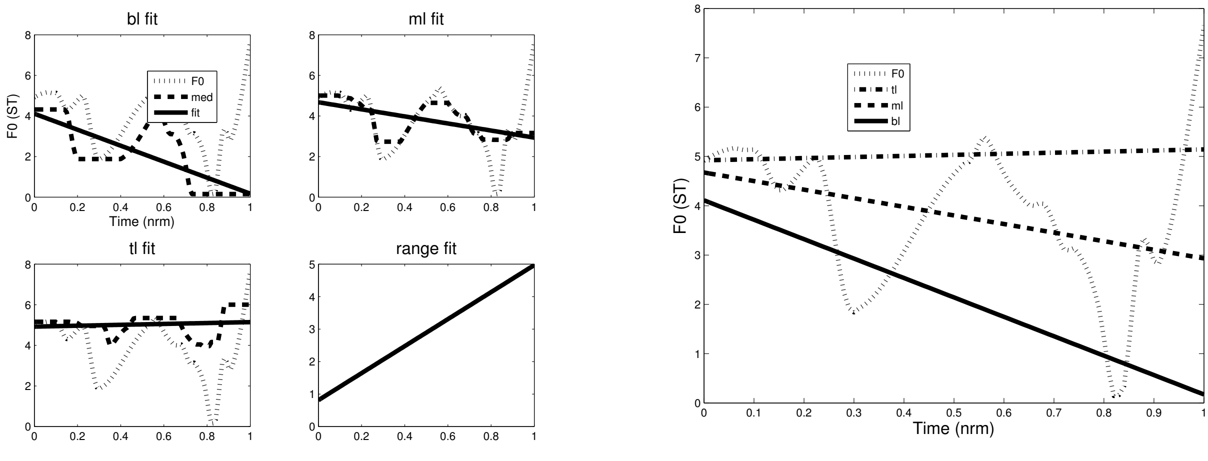

Figure 2 - **Left:** *Stylization of base- (bl, top left), mid- (ml, top right) and topline (tl, bottom left) based on F0 median sequences below the 10th percentile for the baseline, above the 90th percentile for the topline and for all values for the midline. The F0 range (bottom right) is represented by a regression line fitted through the pointwise distances between the base- and topline.* **Right:** *Resulting base-, mid and topline.*

From these line fits we derived level and range discontinuity measures as shown in Figure 3. Level refers to the midlines, range to the distance between base- and toplines. Discontinuity is measured between the two segments x_1 and x_2 in order to quantify reset, as well as between each of these segments and the joint segment $x_{1.2}$ to see in how far the segments deviate from a common faith. Only those pitch discontinuity features proposed in [11] were used, which we believe can be safely interpreted as boundary cues. This allows for disentangling the prominence- and phrasing-related differences between $x_1 - x_2$ and $x_1 - X_2$.

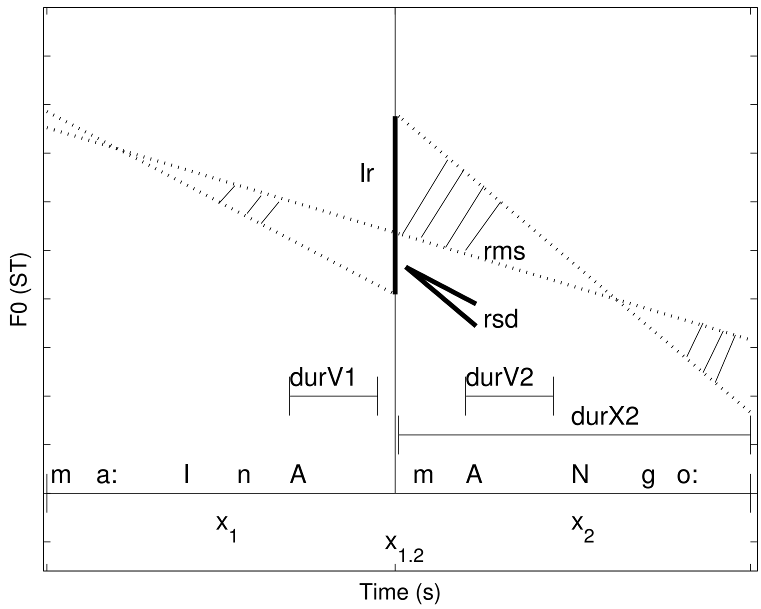

Figure 3 - *Discontinuity and duration features derived at the word boundary between x_1 'málna' and x_2 'mango'. Level reset LR gives the absolute distance between the end of the midline for x_1 and the start of the midline for x_2. RSD reflects the slope difference of the pitch range regression lines of x_1 and x_2. RMS gives the midline deviation of x_2 (rmsX2) and of both x_1 and x_2 (rmsX12) from a common faith, i.e. the midline through both words $x_{1.2}$. DUR subsumes the durations of the last vowel of x_1 (durV1), the first vowel of x_2 (durV2), and the total duration of x_2 (durX2), respectively.*

Table 1 - Prominence and boundary features. x_2 refers to both x_2 and X_2. Types *b* and *p* indicate prominence and boundary related features, respectively.

Name	Description	Type
durX2	word duration of x_2	*p*
durV2	duration of first vowel in x_2	*p*
rmsX2	root mean squared deviation between the midline in x_2 and the corresponding midline part of $x_{1.2}$	*p*
rmsX12	overall root mean squared deviation between the midlines in x_1 and x_2 compared to the midline in $x_{1.2}$	*b*
lr	pitch level reset, i.e. the absolute distance between the end point of the midline in x_1 and the start point of the midline in x_2	*b*
rsd	slope difference of pitch ranges in x_1 and x_2, i.e. the difference how ranges change over time	*b*
durV1	duration of last vowel in x_1	*b*
durP	duration of pause preceding x_2	*b*

3 Results

Linear mixed-effect models (`lme4` package in R) were used with each of the above mentioned acoustic measures as dependent variable and the size of the 2nd fruit (x_2 vs X_2) as fixed effect and speaker as random effect. Significance level was set to $\alpha = 0.05$.

It turned out that not only all prominence but also all examined boundary features showed a significant difference between the x_2 and the X_2 condition ($p < 0.001$ for features *durP, durV1, durV2, durX2,* *rmsX2*, and *rmsX12*; $p < 0.05$ for features *lr* and *rsd*). All these differences confirmed our hypothesis, that $x_1 - X_2 - x_3$ is not only marked by a stronger prominence but also by stronger boundary signals preceding X_2. The prominence and boundary feature values for both conditions are displayed in the boxplots in Figures 4 and 5.

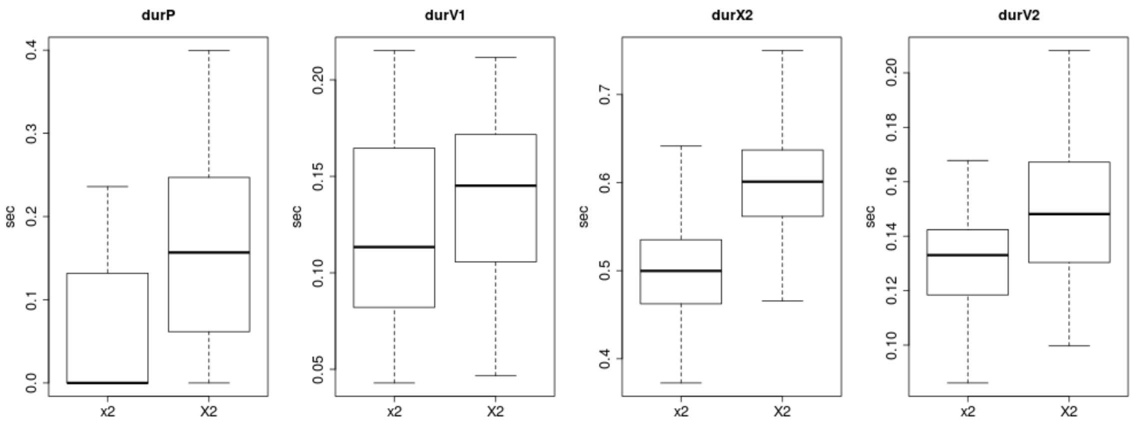

Figure 4 - *Duration and pause feature differences between the non-prominent x_2 (left boxplots) and the prominent X_2 (right boxplots) condition.*

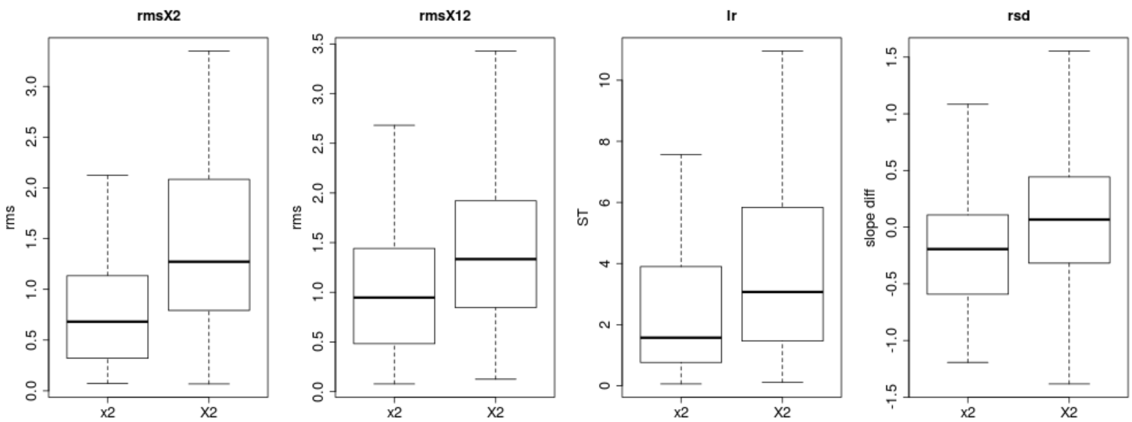

Figure 5 - *Pitch discontinuity differences between the non-prominent x_2 (left boxplots) and the prominent X_2 (right boxplots) condition.*

4 Discussion and conclusion

Our results indicate that in Hungarian prominence is not only marked by prominence-related phonetic cues as segment duration and F0 movements, but also can lead to a prosodic re-phrasing if the prominent word is not phrase-initial. This re-phrasing is realized by boundary cues like prefinal lengthening and pitch discontinuity left-adjacent to the prominent word and serves to maintain a left-headed prosodic structure.

In order to examine this re-phrasing we used a new 'Fruits in basket' stimulus design. This design allows for a controlled and fast elicitation of prosodic structure, since it does not require any context establishing information structure and contrast, that may contain additional latent and hard-to-control influence factors. This study focused on examining the content of the first basket.

Within this framework and for the examined duration- and pitch discontinuity based boundary features, it turned out that left-headedness is a very stable property of Hungarian prosody which overrides alternative prosodic structures in production.

5 Acknowledgments

The work of U.D. Reichel is financed by a grant of the Alexander von Humboldt society. The study was carried out within the project "Form and function of prosodic structure in Hungarian and German" supported by DAAD grant no. 57143585 to F. Kleber and MÖB grant no. 65032 to K. Mády.

References

[1] BOERSMA, P. and D. WEENINK: *PRAAT, a system for doing phonetics by computer*. Techn. Rep., Institute of Phonetic Sciences of the University of Amsterdam, 1999. 132–182.

[2] DE PIJPER, J. and A. SANDERMANN: *On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues*. Journal of the Acoustical Society of America, 96:2037–2047, 1994.

[3] DRAXLER, C. and K. JÄNSCH: *SpeechRecorder – a universal platform independent multi-channel audio recording software*. In *Proc. International Conference on Language Resources and Evaluation*, pp. 559–562, Lisbon, 2004.

[4] É. KISS, K.: *The Syntax of Hungarian*. Cambridge University Press, 2002.

[5] FRY, D. B.: *Duration and intensity as physical correlates of linguistic stress*. JASA, 27:765–768, 1955.

[6] FRY, D. B.: *Experiments in the perception of stress*. Language and speech, 1:126–152, 1958.

[7] HUNYADI, L.: *Hungarian sentence prosody and universal grammar: on the phonology-syntax interface*. Lang, Frankfurt am Main, 2002.

[8] KISLER, T., F. SCHIEL and H. SLOETJES: *Signal processing via web services: the use case WebMAUS*. In *Proc. Digital Humanities*, pp. 30–34, Hamburg, Germany, 2012.

[9] LADD, D.: *Intonational Phonology*. Cambridge University Press, Cambridge, 2 ed., 2008.

[10] MÁDY, K., A. SZALONTAI, A. DEME and B. SURÁNYI: *On the interdepencende of prosodic phrasing and prosodic prominence in Hungarian*. In *Proc. 11th International Conference on the Structure of Hungarian*, Piliscsaba, Hungary, 2013.

[11] REICHEL, U. and K. MÁDY: *Comparing parameterizations of pitch register and its discontinuities at prosodic boundaries for Hungarian*. In *Proc. Interspeech 2014*, pp. 111–115, Singapore, 2014.

[12] SCHIEL, F.: *Automatic Phonetic Transcription of Non-Prompted Speech*. In *Proc. ICPhS*, pp. 607–610, San Francisco, 1999.

[13] SWERTS, M. and R. GELUYKENS: *Prosody as a marker of information flow in spoken discourse*. Language and Speech, 37(1):21–43, 1994.

[14] VARGA, L.: *Intonation and Stress: Evidence from Hungarian*. Palgrave Macmillan, Hampshire, New York, 2002.

[15] VENNEMANN, T.: *Neuere Entwicklungen in der Phonologie*. Mouton de Gruyter, Berlin, 1986.

[16] WIGHTMAN, C., S. SHATTUCK-HUFNAGEL, M. OSTENDORF and P. PRICE: *Segmental Durations in the Vicinity of Prosodic Phrase Boundaries*. JASA, 91(3):1707–1717, 1992.