

Tonal representation of Hungarian yes/no questions and acoustic correlates

Katalin Mády, Uwe D. Reichel, Anna Kohári, Ádám Szalontai

ELTE Research Centre for Linguistics, Budapest

Introduction. Sentence intonation can be described in various ways. A well-known approach is the contour-based model containing dynamic categories such as rise, fall or level. An extensive description of Hungarian intonation in this framework was provided by Varga [6]. Another approach was to decompose the tonal movements into High and Low (H, L) tonal events which are aligned with prominence-bearing syllables and phrase boundaries, resulting in the ToBI annotation system (**T**ones and **B**reak **I**ndices) [5]. Since systems like ToBI focus on specific points of interests in prosodic structure. They integrate structural analysis across different levels like syntax and phonology. Furthermore, they are quantifiable both in terms of labels associated with tonal events as well as acoustic features like fundamental frequency (f0) and intensity (energy). The ToBI system has been adapted to a wide range of languages, which facilitates the comparison between prosodic forms and functions and the prosodic typology across languages. However, annotations rely heavily on previous assumptions on the intonational phonology of the given language.

Intonation annotation. In this paper, we suggest a basis for a systematic and comprehensive description of Hungarian intonation in the tone-based approach. To avoid circularity when developing an annotation system and a phonological description at the same time, the German DIMA (*Deutsche Intonation, Modellierung und Annotation*) system is applied. This pre-phonological, surface-related, theory-neutral annotation system aims at the prosodic representation of the actually produced pitch contour [1]. As with ToBI, annotators are trained to rely on their perceptual impression rather than simply to set marks based on f0 movements. Annotations in DIMA are carried out on three layers. (1) **Phrase** boundaries are strong (%) or weak (-). (2) **Prominence** strength is weak (1), strong (2) or emphatic (3). (3) High (H) and low (L) **tones** are associated with phrase boundaries and prominent syllables. DIMA provides the possibility to mark tonal events (e.g. turning points) without linking them to a phrase boundary or prominence. This is useful when describing variation in the pitch contours that are perceivable but do not necessarily denote different grammatical or pragmatic functions. One such example is given in Fig. 1 showing three possible forms of neutral information-seeking yes/no questions. In ToBI, all three contours are described as L* H L%. In DIMA, the mid (red) contour contains an additional L tone to mark the turning point from level to rise before the high penultimate syllable, whereas the bottom (green) contour contains a high plateau indicated by H on the syllable following the L* pitch accent.

The downstep (!) and upstep (ˆ) symbols indicate that a pitch accent is considerably lower or higher than a previous identical pitch accent within the same utterance. These symbols are also used to mark lower or higher register or larger or smaller pitch range of a phrase compared to the previous one. This feature of DIMA is generalised in our approach: ! and ˆ are used to indicate the relative height of otherwise identical pitch accents of the same

speaker even across utterances. This distinction is useful when we explore intonation patterns of non-neutral pragmatic contents. For example, Hungarian incredulous yes/no questions are characterised by lower f0 and a smaller f0 range [2]. This is captured by the lowered pitch accent !L* as shown in Fig. 2, right panel.

Materials. 146 yes/no questions of 12 native Hungarian speakers (10 f., 2 m.) in the spontaneous Akaka Maptask Corpus [3] were labelled by three trained annotators. A common decision was achieved on consent, if necessary. A total number of 993 tone labels were available for analysis. Label counts were unbalanced: L tones for initial and final phrase boundaries and turning points were the most frequent (33%), followed by H (21%) which is the penultimate f0 peak between the rise and the fall. Out of 251 pitch accents, 196 (78%) were !L, L or ^L. The prevalence of low pitch accents is not surprising given the rising-falling pattern of yes/no questions.

Prosodic feature extraction. In order to identify acoustic correlates of the manually annotated tone labels, f0 was extracted by autocorrelation with Praat version 6.4.06 and postprocessed with CoPaSul version 1.5.3 [4]. We then extracted standard f0 and energy features (22 and 26, respectively) like median and IQR, as well as 76 local f0 contour features with CoPaSul. For local contour feature extraction, f0 was decomposed into a global and a local component based on the phrase boundary and tone time stamps. All features were extracted within analysis windows of length 200 ms (standard features), respectively 300 ms (contour features), centered on the tone label time stamps.

Analyses. In an initial exploratory analysis we identified all prosodic features that (1) turned out to differ significantly between at least one tone label pair with (2) at least medium effect size. For this purpose we applied Kruskal-Wallis tests with α set to 0.05, and we measured the effect size in terms of Cohen’s D, “medium” defined as $D \geq 0.5$. Given the exploratory nature of this analysis and the large amount of features, we did not apply any type-1 error correction. Our results show that among the standard f0 features 15 out of 22 fulfilled both above-mentioned criteria: among the standard energy features 18 out of 26, and among the local f0 contour features 69 out of 76. For two of these identified features, namely f0 and energy median, we show the distribution plots for all twelve tone labels in Fig. 3. Furthermore, the parabolic stylisation coefficient distributions reveal a predominantly falling-rising shape for L tone variants (except of ^L), and a rising-falling shape for H variants.

Discussion. Our prosodic annotations of spontaneous Hungarian yes/no questions with a slightly modified version of the DIMA system rely on perceptual labelling rather than identifying f0 minima and maxima in the signal. The analysis of various acoustic features support the reliability of these impressionistic labels showing that the six tone heights are indeed distinguishable, and they follow the expected order from low to high. Besides, prominent syllables are characterised by higher energy than their non-prominent equivalents (boundaries and turning points). These results are promising for the future development of an automatic tone prediction tool.

Acknowledgements. This work was funded by the National Research, Development and Innovation Office, grant NKFIH K 135038.

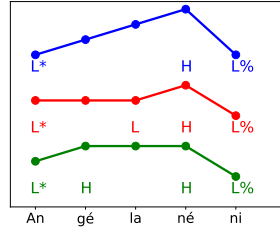


Figure 1: Three contours for the neutral yes/no question ‘Aunt Angéla?’ (Varga 2002:480).

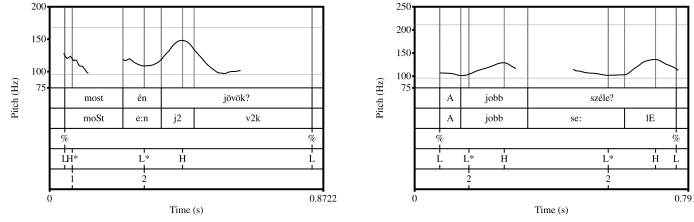


Figure 2: F0 contours and the speaker’s f0 range. Left: neutral yes/no question ‘Is it my turn?’. Right: incredulous question with lowered pitch and smaller pitch range ‘At the right edge?’.

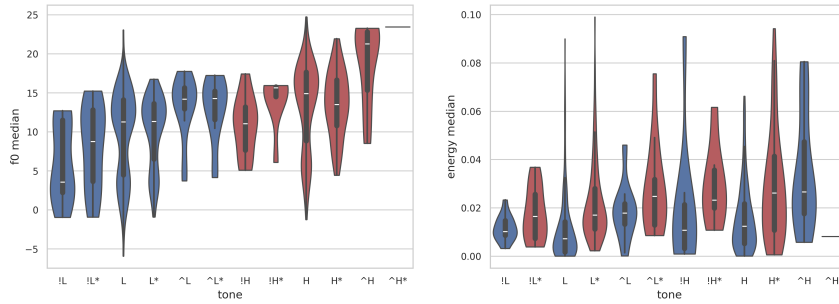


Figure 3: Violin plots for median values for lowered (!) and raised ^L(ow) and H(igh) tones. The * symbol indicates prominence. Left: f0, right: energy.

References

- [1] Kügler, F. et al. (2015). ‘DIMA – annotation guidelines for German intonation’. In: *Proc. 18. International Congress on Phonetic Sciences*. Glasgow.
- [2] Mády, K. et al. (2023). ‘Prosodic cues of distinguishing neutral and non-neutral yes/no questions in Hungarian: the acoustics of surprise’. In: *Proc. 16th ICPHS, Prague*, pp. 1608–1612.
- [3] Molnár, C. S. et al. (2023). ‘The Akaka Maptask Corpus’. In: *Beszéd kutatás – Speech Research Conference*. Budapest, pp. 81–83.
- [4] Reichel, U. D. (2016). *CoPaSul Manual – Contour-based parametric and superpositional intonation stylization*. Research Institute for Linguistics, Hungarian Academy of Sciences. Budapest, Hungary.
- [5] Silverman, K. et al. (1992). ‘ToBI: a standard for labeling English prosody’. In: *Proc. 2nd International Conference on Spoken Language Processing, Banff, Alberta*. Vol. 2, pp. 867–870.
- [6] Varga, L. (2002). *Intonation and stress: evidence from Hungarian*. Basingstoke & New York: Palgrave Macmillan.