Perceptual sensitivity to prenuclear and nuclear intonational patterns

Tomáš Duběda

Institute of Translation Studies, Charles University in Prague

dubeda@ff.cuni.cz

Abstract

We describe a perceptual experiment whose goal is to compare perceptual sensitivity to pitch accent contrasts in nuclear and prenuclear positions. The material consists of Czech sentences which have been resynthesized with controlled intonation. The results show that changes in nuclear pitch accents are perceived more sharply than changes in prenuclear pitch accents, and that the H* accent is perceptually more salient than the other accent types (L*H, L* and S*). The effect of constituent edge on the perception of intonational contrasts has not been confirmed.

Index Terms: pitch accents, nuclear and prenuclear intonation, perceptual sensitivity, Czech.

10.21437/Interspeech.2011-451

1. Introduction

Perceptual aspects of intonation in connected speech have been tackled from different points of view. For instance, [1] claims that the just-noticeable-difference threshold lies around 1 semitone for linguistically relevant melodic events, and that languages tend to preserve discriminability by implementing an optimal difference of 3 semitones. [2] showed that intonation is not perceived as a continuous variable, but as a sequence of level tones or glissandos associated with syllable nuclei. [3] highlights the fact that well-established psychoacoustic methods are not applicable in the domain of intonation, which is a fairly complex perceptual object.

Many contemporary prosodic models accept the *intonation phrase* as the central unit of intonational analysis; within this unit, they usually contrast the *intonation nucleus* (or *sentence stress*) – a perceptually salient place with particular phonological exposure, often located near the end – with the rest of the phrase. Given their linguistic relevance, nuclear configurations have received far more attention in prosodic theory than prenuclear intonational events, which are often treated as secondary, and may be even totally neglected in prosodic descriptions. Recent theories, however, try to show that there are regular patterns even in prenuclear intonation. Furthermore, it is generally agreed that nuclear patterns have a more complex structure, and their stylization requires more formal components: for example, in J. Pierrehumbert's model of English intonation [4], the prenuclear part corresponds to a simple sequence of pitch accents (introduced by an initial boundary tone), whereas the nuclear contour includes a pitch accent, a phrase tone and a boundary tone.

A relevant issue in intonational grammars is whether the choice of pitch accents (PAs) is the same for prenuclear and nuclear parts. [4] assigns one and the same inventory of PAs to both positions, while [5] puts forward 5 prenuclear PAs and 8 nuclear PAs in his analysis of English. Yet another question is whether the range of phonological contrasts induced by different PAs is the same in nuclear and prenuclear positions. Although this point does not receive enough attention in intonational grammars, many of them admit, at least implicitly, that this is not the case (e.g. [5] for English, [6] for French, and [7] for Czech). Indeed, it is hardly imaginable that a prenuclear pitch event would be the primary mark of modality, continuation/finality or expressivity in the above-mentioned languages, as these functions concentrate in the nucleus. Of course, prenuclear material is not entirely insensitive to functions expressed by the nucleus, and cases where marks of these functions are distributed over the whole phrase are frequent (e.g. [8] for yes/no-questions in French, or [9] for expressivity in Czech), but it is still sound to think that the role of the nuclear pattern is primary in most cases.

Thus, in the case of a single inventory of PAs applicable to all positions in the phrase, the PAs may contrast with each other less clearly in prenuclear positions than in nuclear positions. In other words, there may be neutralization of paradigmatic contrasts between prenuclear PAs despite their formal resemblance with (fully contrasting) nuclear PAs. This kind of situation is not unusual in language, and has been formalized e.g. by J. R. Firth [10] in his "polysystemic" approach.

This view of prenuclear accents would imply that their main function is culminative and/or demarcative (marking of stress and, in languages with fixed stress, of word boundaries), and that this function is rather syntagmatic than paradigmatic. Furthermore, on the perceptual level, the lack of paradigmatic contrast would mean that prenuclear accents may be perceived less sharply than the nuclear. This does not exclude, however, that some accent types are perceptually more salient than others in all positions.

2. Aim and hypotheses

The assumptions arrived at in the previous paragraph allow us to formulate the following three hypotheses:

- Hypothesis 1: A change in a prenuclear PA will produce less perceptual effect than a change in a nuclear PA.
- Hypothesis 2: Replacing a PA by a similar PA will produce less perceptual effect than replacing it by a dissimilar PA (see section 4.2 for a more thorough motivation of this hypothesis).
- Hypothesis 3: By virtue of general perceptual mechanisms, a change near the edge of the intonation phrase will produce more perceptual effect than a change in the middle of it.

The language tested is Czech, a Western Slavonic language with fixed stress (generally described as weak) on the first syllable of stressable words. Vowel length is phonological and there is no pronounced reduction of unstressed syllables. Prosodic nuclei usually take the final position in a phrase.

3. Methodology

The experiment consists of the following steps:

- Recording a set of structurally comparable sentences;
- Changing the intonation of these sentences by means of resynthesis;
- For all relevant positions, elaborating pairs of sentences with one contrasting PA, the rest being identical;

- Presenting these pairs of sentences to listeners who assess the perceptual difference between them.

The following sentences were used:

1. *Navzdory vadnému náhonu vyhráli.* "They won despite of a defective drive."
2. *Nemají bohužel měděné barely.* "Unfortunately, they have no copper barrels."
3. *Dorazil na závěr druhého závodu*. "He arrived by the end of the second race."
4. *Názory vedení bývaly rozumné*. "The management's opinions used to be reasonable."
5. *Na břehu viděli moderní budovy*. "They saw modern buildings on the shore."

All sentences have the same structure: they are declarative and contain a single intonation phrase (despite the English translation of sentence 2) consisting of four 3-syllable stress units with initial accent. The last stress unit is nuclear. Successive PAs are separated by 2 syllables so as to avoid tonal crowding. All segments are voiced.

These sentences were recorded by one female speaker.

We used the following set of PAs for the purposes of our study, based on [7] and [11]:

Table 1. *Czech pitch accents. Bold dots correspond to stressed syllables, and small dots correspond to adjacent unstressed syllables.*

Group	Pitch accent	Most typical structure
"Low"	L*H	
	L*	
"High"	H*	
	H*L	
"Level"	S*	

The "low" accents make up approximately 2/3 of all PAs in Czech; L*H is typical in all positions, L* does not occur phrase-initially. The "high" accents are rare and generally induce some degree of focus, with the exception of phrase-initial H*. To keep the test shorter, we decided only to study the more frequent of these two accents, namely H*, in initial, nuclear and immediately pre-nuclear positions. The S* accent corresponds to no or almost no intonational change around the syllable perceived as stressed. Since it does not occur phrase-initially and is very rare nuclearly, we decided to study it only in the 2nd and 3rd positions.

It is evident that the inventory of Czech pitch accents differs from what is the norm in many other languages, especially in the absence of accents corresponding to simple peaks. This is probably one of the reasons why speakers of other languages perceive Czech as "monotonous".

The range of contrasts in PAs that we studied for each of the four positions in the phrase is given in Table 2.

Table 2. *Pitch accents studied*

Position	Pitch accents				Nr. of different PAs	Nr. of contrasts
1st	L*H		H*		2	1
2nd	L*H	L*		S*	3	3
3rd	L*H	L*	H*	S*	4	6
4th	L*H	L*	H*		3	3

This means that for the phrase-initial position, we shall study a pair of sentences which contrast L*H and H* in the first stress group, the rest being equal; for the second position, we shall study pairs with contrasting L*H/L*, L*H/S* and L*/S* in the second stress group etc. The total number of contrasts studied is 13.

One obvious way of preparing pairs of sentences contrasting in a single PA is taking a natural variant containing one PA, and resynthesizing another variant with the other PA, which would possibly be copied from another sentence. However, this approach has at least three disadvantages: the recording must be assisted so as to obtain all required PAs, applying a different PA within a prosodic phrase may cause discontinuities in the overall f_0 course, and it is not easy to achieve a comparable scaling (i.e. pitch excursion magnitude) of the two contrasting PAs.

To avoid these problems, we decided to use artificial intonation throughout the material, i.e. to create a single tonal infrastructure with controlled tonal intervals, and to implement the studied PAs at relevant positions in this basic pattern. The structure L*H L*H L* L* L% was chosen as this basic pattern, as it was the most typical one in the recorded sentences. Fig. 1 gives an example of this procedure: two different modifications of the third PA are shown.

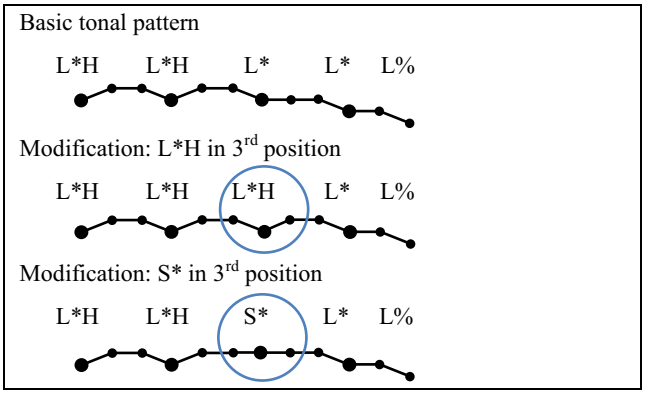

Figure 1. *The basic tonal pattern and two examples of its modification*

Next, the tonal configurations for all 13 relevant contrasts were applied on the recorded sentences (the 5 sentences were used in roughly equal proportion). The procedure was as follows:

- The magnitude of each elementary rise or fall was fixed at 2 semitones. This value is close to the average span of elementary pitch excursions in the recorded sentences, and lies between the two reference values given by [1] (cf. Introduction). This interval applies iteratively to all successive falls or rises; e.g. the basic tonal pattern as presented in Fig. 1 has a total pitch span of 6 ST.
- A declination factor of –0,38 ST/syllable (average value obtained by measuring the declination of the recorded sentences) was applied on the tonal configuration.
- The tonal configuration was aligned with the syllables of the natural sentence.
- The tonal configuration was scaled on the f_0 axis by incrementing/decrementing all points so as to obtain the best fit with the f_0 course of the natural sentence (least square method). This procedure guarantees a natural pitch register and minimum quality loss during resynthesis.
- The natural sentence was resynthesized with the new intonation by means of the PSOLA method [12]. For the

sake of simplicity, only one stationary f_0 value per syllable was applied: this approach is not only in line with the findings of [2], but it also leads to an output which is fully natural as far as intonation is concerned.

Figure 2 shows the described procedure schematically.

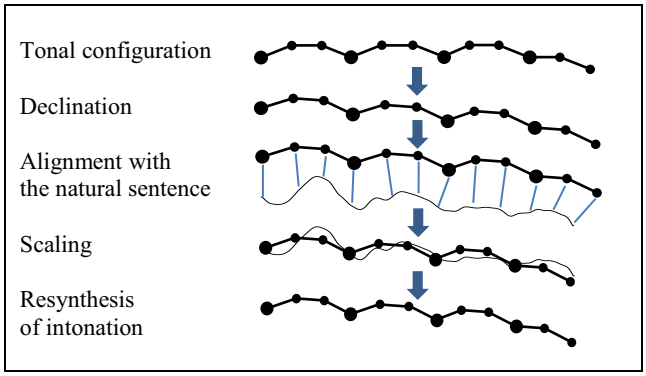

Figure 2. *Application of tonal configurations on the recorded sentences.*

The output of the mapping-and-resynthesis procedure is thus a set of 13 pairs of sentences with artificial, but naturally sounding intonation, each pair contrasting two different PAs in a given position. The artificial intonation respects declination and fits the natural sentence both horizontally (syllable-by-syllable mapping on the time axis) and vertically (best fit on the f_0 axis).

Next we prepared the listening test, which includes the following items:

- 13 sentence pairs in both orders (i.e. 26 items);
- 6 pairs of identical sentences as a means of assessing the listeners' perceptual reliability;
- 10 pairs of distractors (sentences taken from a different recording by the same speaker, modified intonationally).

The 42 items were randomized into a listening test; each pair of contrasting sentences was repeated twice. For each item, the listeners were choosing between three possible answers:

- no or almost no difference between the two sentences;
- perceivable difference, but no change of meaning;
- difference which could have an impact on meaning.

43 listeners (B.A. and M.A. students of modern languages) participated in the assessment, which was carried out in a classroom.

4. Results and discussion

The answers obtained were converted into scores (no difference: 0; perceivable difference: 1; difference in meaning: 2) and analysed statistically.

4.1. Hypothesis 1

We tested Hypothesis 1 (*A change in a prenuclear PA will produce less perceptual effect than a change in a nuclear PA*) by calculating the mean score for all contrasts in nuclear positions, and comparing it to the mean score for all contrasts in all prenuclear positions. The difference, shown in Fig. 3, is highly significant (unpaired t- test, $p < 0{,}001$).

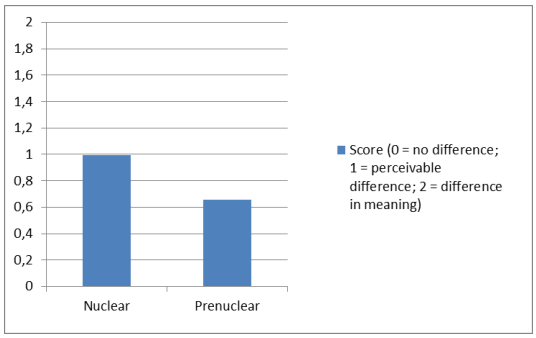

Figure 3. *Average perceptual score for nuclear and prenuclear contrasts.*

This means that when listening to pairs of sentences with one contrasting pitch accent, the listeners were significantly more aware of this contrast if it occurred in the nuclear (i.e. final) position than if it occurred elsewhere.

However, it should be noted that the range of contrasts is not the same in nuclear and prenuclear positions (cf. Table 2). To confirm the result presented above, we made a separate comparison of all three nuclear contrasts (L*H vs. L*, L*H vs. H*, L* vs. H*) with identical contrasts occurring in prenuclear positions. The average perceptual scores are the following:

- L*H vs. L* in nuclear positions: 0,535
  in prenuclear positions: 0,390
  $p = 0{,}086$
- L*H vs. H* in nuclear positions: 1,023
  in prenuclear positions: 0,762
  $p = 0{,}002$
- L* vs. H* in nuclear positions: 1,430
  in prenuclear positions: 1,244
  $p = 0{,}019$

In all three cases, the contrasts were perceived more sharply in nuclear positions; two differences out of three are significant. The extremely high significance of the results presented in Fig. 3 (which include the three contrasts above) can be explained by the presence of contrasts with the S* accent in prenuclear positions (cf. again Table 2).

4.2. Hypothesis 2

To test Hypothesis 2 (*Replacing a PA by a similar PA will produce less perceptual effect than replacing it by a dissimilar PA.*), we retained the three categories displayed in Table 1 ("low accents" – containing an L* tone; "high accents" – containing an H* tone; and the S* accent). One of the observations put forward by [7] is that the low accents are unmarked, while the high accents induce some degree of emphasis; the S* accent can be seen as reduced form of the low accents.

Figure 4 displays perceptual scores obtained for each of the between-group contrasts. Additionally, it contains the score for the group-internal contrast L*H vs. L*.

The four columns of the graph clearly fall into two categories: contrasts including an H* accent correspond to significantly higher perceptual scores than the other types of contrasts. The difference between the two categories was found to be highly significant, while the difference between members of the same category is never significant (unpaired t-test: 1st column with the 2nd: $p < 0{,}001$; 1st with 3rd: $p = 0{,}377$; 1st with 4th: $p < 0{,}001$; 2nd with 3rd: $p < 0{,}001$; 2nd with 4th: $p = 0{,}143$; 3rd with 4th: $p < 0{,}001$).

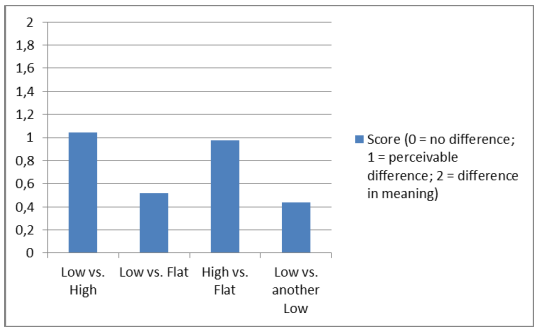

Figure 4. *Average perceptual score for contrasts between PA groups.*

The hypothesis of the markedness of H* is thus confirmed (except for the phrase initial position – cf. below). On the other hand, the S* accent is not categorically different from the low accents.

4.3. Hypothesis 3

Hypothesis 3 (*By virtue of general perceptual mechanisms, a change near the edge of the intonation phrase will produce more perceptual effect than a change in the middle of it*) partly overlaps with Hypothesis 1, because the nucleus (cf. 4.1) has always a final position in our material.

When testing the first position, we face the problem that only one contrast (L*H vs. H*) occurs there (cf. Table 2). The perceptual scores for this contrast, also occurring in the 3rd and the 4th positions, are the following:

- 1st position 0,337
- 3rd position 1,186
- 4th position 1,023

The score in the 1st position is significantly different from the two other positions (unpaired t-test: in both cases $p < 0{,}001$), but, contrary to the hypothesis, it is lower rather than higher. This indicates that the left edge of the prosodic phrase is not a location where users of Czech expect contrasts.

4.4. Listeners' reliability

The reliability of the listeners in our experiment can be verified in the following ways:

- An ideal listener should assign the "0" score to pairs of identical stimuli. The average score for these items across listeners is 0.10, which can be seen as a satisfying result.
- An ideal listener should provide the same answer for identical stimulus pairs in which only the order has been varied. In our material, the answer was different in 49% of the cases, with 15% of difference for the "best listener", and 92% of difference for the "worst listener". However, this finding appears less dramatic if we consider that in 98.7% the difference was between neighbouring categories (0 vs. 1; 1 vs. 2), and only in 1.3% between the two extreme categories (0 vs. 2).
- Listeners which are, statistically speaking, "outliers", i.e. listeners whose behaviour has a very low correlation with the average, should also be seen as risky.

To verify whether the results change if we exclude "bad listeners", we calculated again all values after removing listeners who had an average score of 0.20 and more for identical stimuli, those who had a difference of 0.50 and more for stimulus pairs with different order, and those who had a correlation below 0.50 with the average rating. 20 listeners were excluded by virtue of these restrictions, but none of the results presented above changed significantly, i.e. all categories as well as corresponding significance levels were maintained.

5. Conclusion

The use of sentences with controlled intonation allowed us to study the perceptual impact of contrasts between individual pitch accents in different positions, with the other conditions being kept constant (including the magnitude of the pitch excursion).

The hypothesis that a change in the nuclear PA is perceived more sharply than a change in a prenuclear PA has been confirmed with high significance. This is partly due to differences in PA distribution, but even for strictly comparable contrasts, a significant difference has mostly been shown.

The hypothesis that the H* accent is perceptually the most salient has been confirmed as well: the contrast between this PA and the other accent types turned out to be highly significant according to the listeners' ratings. The S* accent is not categorically different from the low accents (L*H and L*).

No positive effect of constituent edge on the perception of melodic contrasts could be observed, unless we consider that the final position of the nucleus is itself an edge phenomenon.

In the light of the results obtained, we have good reasons to believe that paradigmatic contrasts between Czech pitch accents are reduced in prenuclear positions, and that their function in these positions is rather syntagmatic (demarcative), with the notable exception of the H* accent, whose perceptual markedness was observed both phrase-finally and phrase-internally.

6. Acknowledgements

This article was prepared with the support of the GAČR P406/10/0101 grant.

7. References

[1] Hart, J. 't, "Differential sensitivity to pitch distance, particularly in speech", JASA, 69(3): 811–821, 1981.

[2] Mertens, P., "The Prosogram: semi-automatic transcription of prosody based on a tonal perception model", in B. Bel and I. Marlien [Eds.], Speech Prosody 2004, 23–26, 2004.

[3] Vaissière, J., "Perception of intonation", in D. B. Pisoni and R. E. Remez [Eds.], The Handbook of Speech Perception, Blackwell Publishing, 236–263, 2005.

[4] Pierrehumbert, J., The phonology and phonetics of English intonation, Ph.D. thesis, MIT, 1980.

[5] Gussenhoven, C., The Phonology of Tone and Intonation, CUP, 2004.

[6] Post, B., Tonal and phrasal structures in French intonation, The Hague: Holland Academic Graphics, 2000.

[7] Duběda, T., "Towards an inventory of pitch accents in Czech", Slovo a slovesnost 72, 3–12, 2011.

[8] Di Cristo, A., "Intonation in French", in D. Hirst and A. Di Cristo [Eds.], Intonation Systems. A Survey of Twenty Languages, Cambridge University Press, 195–218, 1998.

[9] Duběda, T., "The role of nuclear and extra-nuclear contours in the perception of expressivity in Czech", in R. Vích [Ed.], 19th Czech-German Workshop „Speech Processing", Institute of Photonics and Electronics, 75–79, 2009.

[10] Firth, J. R., Papers in Linguistics 1934–1951, London: Oxford University Press, 1957.

[11] Daneš, F., Intonace a věta ve spisovné češtině, Prague: ČSAV, 1957. [Sentence intonation in standard Czech]

[12] Boersma, P., "Praat, a system for doing phonetics by computer." Glot International 5-9/10: 341–345, 2001.