ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2006
Dresden, Germany
May 2-5, 2006

Intensity as a macroprosodic variable in Czech

Tomáš Duběda

Institute of Phonetics
Charles University in Prague
dubeda@ff.cuni.cz

Abstract

The present paper provides an acoustic description of macrointensity patterns of stress units (prosodic words) in read Czech, as reflected by the intensity of syllable nuclei. Normalized intensity values show that there is a gradual macrodynamic decrease over the inter-pause group, followed typically by a significant intensity reset. Local intensity drops occur between the last two syllables of the stress unit; in addition, there is a major intensity drop before the pause. Syllables bearing perceived accents do not show intensity peaks.

1. Introduction

Intensity as the acoustic counterpart of perceived loudness is far from being a frequent object of prosodic research. Its variations may be affected by external influences (such as head movements during the recording), and its behaviour is believed to be only moderately systematic and to have limited perceptual relevance. Unlike fundamental frequency or duration, it has hardly been used in speech synthesis [5]. More specifically, the last two editions of *Speech Prosody* did not include any paper which would contain either of the words *intensity* or *loudness* in its title.

We can distinguish three levels of dynamic variability: a) intrinsic aspects, reposing on the production mechanism of individual segments [7]; b) contextual aspects, resulting from segment coarticulation [10]; c) macroprosodic aspects, belonging to the suprasegmental level. The term "microintensity" has been put forth as a useful label covering dynamic patterns within segments, such as the position of the intensity peak in a vowel, or the dynamic steepness of the pre-peak part [3]. Apart from full-spectrum intensity studies, a finer approach has been proposed by Sluijter et al. [11]: intensity increase in English stressed syllables is reported not to be uniform, but associated mainly with higher frequencies.

Bartkova et al. [1] identify pitch, duration, position in the word and position with respect to pause as the most important factors determining segment intensity in a French speech sample. The data described by Jokisch & Kühne [4] show that syllables have higher intensity at sentence beginning than at its end. In the same vein, Maddieson [8] considers overall amplitude decline throughout the utterance to be universal. Ladd [6] mentions intensity as one of the parameters of English stress. Finally, Duběda [2] suggests two major dynamic tendencies of Czech stress units: unit-final intensity drop and absence of intensity peak on the accented syllable.

2. Aim, speech material and method

The present paper provides an acoustic description of macrointensity patterns of stress units (prosodic words) in read Czech, as reflected by normalized intensity of syllable nuclei. The nuclei (mostly vowels, sometimes sonorants in Czech) were chosen as highly representative points of dynamic reference for their constant voicing, mutual comparability (in terms of intensity, the vowel class is more coherent than the consonant class), and regular distribution (one nucleus per syllable). Intensity is only considered in its relation to the structure of stress units, independently of its interaction with fundamental frequency or duration. Spectral balance is not considered either. Specific points of interest, suggested by previous research on Czech and other languages, are a) unit-final intensity patterns; b) intensity patterns associated with the accented syllable; c) across-unit intensity patterns.

The speech sample is a text passage counting approx. 900 syllables, read by three speakers (men, average age 29 years, standard pronunciation with no marks of regional or social varieties) and recorded in studio conditions. To exclude external factors, the speakers were trained to keep their distance from the microphone as constant as possible during the recording.

The recordings were saved and manually annotated in the Praat software. Intensity was extracted with a time step of 2 ms, and the minimum pitch was set to 70 Hz. The length of the window was optimized automatically by the Praat program.

Different representative intensity values have been used so far: the intensity mean over the segment [4], the intensity maximum of the segment, the value in or near the middle of the segment [1], or a variable point determined by the f_0 pattern [10]. The first option was adopted for the present analysis.

Since the concern of the present paper is macrointensity, the raw mean intensity values had to be normalized in order to exclude intrinsic factors. The method is as follows (example of an occurrence of the vowel [ɛ]):

$$i_{norm} \text{ of a given } [\varepsilon] = i_{raw} \text{ of this } [\varepsilon] \frac{i_{aver} \text{ of all segments}}{i_{aver} \text{ of all } [\varepsilon]}$$

where:

i_{norm} – normalized intensity [dB]
i_{raw} – raw intensity [dB]
i_{aver} – average intensity [dB]

All values mentioned henceforth correspond to i_{norm} of the formula above.

The basic unit within which intensity is studied is the stress unit (prosodic word). Since Czech stress has a word-initial position, it mostly coincides with the first syllable of the stress unit:

Vinice se postupně likvidovaly.
[ˈvɪɲɪt͡sɛsɛ | ˈpɔstupɲɛ | ˈlɪkvɪdɔvalɪ]
'The vineyards were gradually liquidated.'

10.21437/SpeechProsody.2006-59

In some cases, the stress unit may have one or several pre-stress syllables (anacruses):

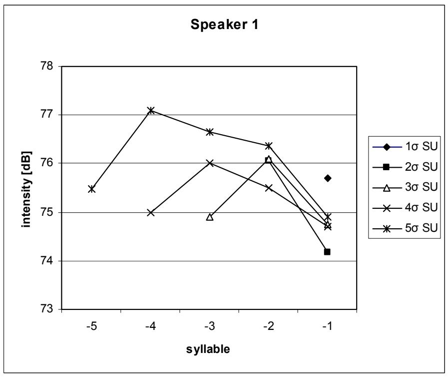
asi tři tisíce keřů
[ʔasɪˈtr̝̊ɪcɪsiːt͡sɛ | ˈkɛr̝uː]
‘around three thousand bushes’

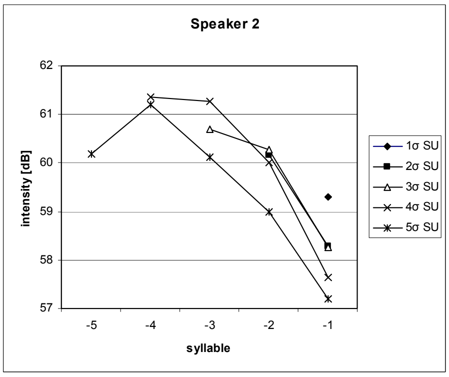
This type of unit, making up only 4,4 % of all stress units in the sample, was excluded from the analysis for its presumably specific rhythmic patterns.
The assignment of accents and the segmentation into stress units were made by the author, on a perceptual basis. The boundaries of intonation units (or phrases), delimited by intonational coherence and mostly followed by a pause, were set as well.

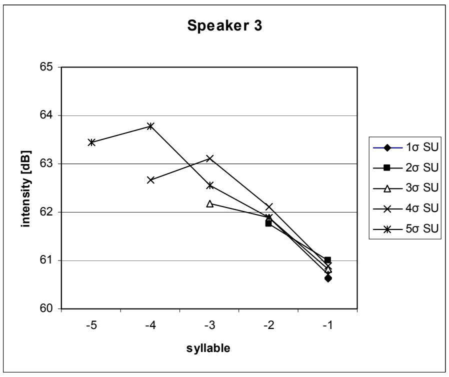
3. Results

3.1 Intensity patterns within stress units

Figures 1–3 show intensity patterns for stress units containing up to 5 syllables; longer stress units were excluded for their low frequency. Due to different levels of gain during the recordings, the dB ranges differ across speakers. Despite this fact, we take the dimensions of the three graphs for comparable; each graph displays an intensity range of 5 dB.
For all speakers, the graphs show a downward tendency which is supplemented by an initial rise in longer units; the presence of this rise seems to be speaker-dependent: in Speaker 1, it concerns units longer than 2 syllables, in Speaker 2, units longer than 4 syllable, and in Speaker 3, units longer than 3 syllables.
For stress units containing 2–5 syllables, the significance of the final intensity drop was tested by means of a t test for paired sets of values. The results, given in detail in the upper part of Table 1, show that all values are significant for Speaker 2, and all but one for Speakers 1 and 3. Both non-significant differences concern longer units (4 and 5 syllables).
Unit-initial intensity behaviour was tested as well. The lower part of Table 1 shows that for stress units of 3–5 syllables, intensity rises between the first and second syllable prevail over falls and that only two rises and no falls are significant.







Figures 1–3: *Average intensity in stress units (SU) of different length.*

Table 1: *Intensity falls and rises at the ends and at the beginnings of stress groups (SU), as shown by Figs. 1–3. For each dB difference, the table displays the number of observations (N) and the p value of a no-difference hypothesis (t test for paired sets of values). Non-significant differences are in italics.*

	Speaker 1			Speaker 2			Speaker 3		
	Diff.	N	p	Diff.	N	p	Diff.	N	p
2σ SU: difference penultimate-last	fall 1,87 dB	92	0,032	fall 1,88 dB	90	< 0,001	fall 0,75 dB	94	0,002
3σ SU: difference penultimate-last	fall 1,36 dB	131	< 0,001	fall 2,02 dB	101	< 0,001	fall 1,11 dB	107	< 0,001
4σ SU: difference penultimate-last	*fall 0,79 dB*	*54*	*0,136*	fall 2,37 dB	44	< 0,001	fall 1,12 dB	46	0,005
5σ SU: difference penultimate-last	fall 1,47 dB	28	0,031	fall 1,79 dB	16	0,010	*fall 1,18 dB*	*20*	*0,086*
3σ SU: difference first-second	rise 1,19 dB	131	< 0,001	*fall 0,42 dB*	*101*	*0,115*	*fall 0,30 dB*	*107*	*0,352*
4σ SU: difference first-second	*rise 1,02 dB*	*54*	*0,056*	*fall 0,08 dB*	*44*	*0,861*	*rise 0,39 dB*	*46*	*0,298*
5σ SU: difference first-second	rise 1,63 dB	28	0,007	*rise 1,03 dB*	*16*	*0,139*	*rise 0,34 dB*	*20*	*0,537*

The overall intensity range of stress units (difference between the highest and the lowest value) generally rises as a function of unit length: the correlation between the range in dB and the length of the unit in syllables is 0,28 for Speaker 1; 0,98 for speaker 2; and 1,00 for Speaker 3.
A specific category of stress units is made up of those which are situated at the end of intonation units. 83 % of these units are delimited by a pause in the studied recordings. The intensity drop at the end of these units is given in Table 2 (one-syllable stress units were not taken into account).

Table 2: *Intensity falls and rises at the end of stress units (SU) located at the end of intonation units (IU). For each dB difference, the table displays the number of observations (N) and the p value of a no-difference hypothesis (t test for paired sets of values). Non-significant differences are in italics.*

	Speaker 1	Speaker 2	Speaker 3
difference penultimate-last in IU-final, pre-pause SU	fall 5,42 dB; N = 66; p < 0,001	fall 3,84 dB; N = 69; p < 0,001	fall 1,64 dB; N = 64; p < 0,001
difference penultimate-last in IU-final, non-pre-pause SU	*rise 0,12 dB; N = 11; p = 0,884*	*fall 0,91 dB; N = 23; p = 0,096*	*rise 0,87 dB; N = 5; p = 0,516*

According to Table 2, intensity decrease before pause (which also implies the end of an intonation unit) is highly significant in all speakers, and it is greater than any of the "fall" values mentioned in Table 1. On the other hand, intonation-unit-final stress units not followed by a pause may end in an intensity rise or fall, both without statistical significance.

3.2 Intensity patterns across stress units

Intensity patterns across stress unit boundaries, as expressed by the intensity difference between syllables left and right of the stress unit boundary, were quantified separately for immediately adjacent stress units (without pause) and those separated by a pause (Table 3).

Table 3: *Intensity falls and rises across stress units. For each dB difference, the table displays the number of observations (N) and the p value of a no-difference hypothesis (t test for paired sets of values). Non-significant differences are in italics.*

	Speaker 1	Speaker 2	Speaker 3
difference last-first except when separated by a pause	fall 1,24 dB N = 263 p < 0,001	*rise 0,11 dB N = 228 p = 0,528*	*rise 0,06 dB N = 228 p = 0,730*
difference last-first when separated by a pause	rise 7,41 dB 70 < 0,001	rise 9,22 dB 66 < 0,001	rise 1,81 dB 64 < 0,001

Immediately adjacent stress units show no significant intensity contrast, except for Speaker 1 whose between-unit falls correlate with his frequent initial rises (see Figure 1).

4. Discussion

4.1 Significance and perceptual discriminability

The data presented in section 3 should be considered with regard both to their statistical significance and their perceptual relevance. Pols [9] gives a just-noticeable difference of 0,5–1 dB for acoustic events comparable with syllable nuclei. Out of the 19 significant differences included in Tables 1–3, all values are greater than the upper limit of this interval, except for one, which lies in it. This observation speaks in favour of a very good match between statistical significance and perceptual discriminability.

4.2 Speaker-independent variability

The overall dynamic tendency of stress units in the studied material may be characterized as decreasing, with a major fall between the last two syllables. The first syllable, bearing the accent, does not correspond to a dynamic peak: in some cases, its average intensity is the highest in the unit, but the difference from the second syllable is never significant in these cases. On the contrary, for one of the speakers, there is significant intensity increase between the first and the second syllable. The most significant component of the general decreasing tendency is the unit-final intensity drop, not an initial peak.
The stress-unit-final intensity fall is greater before pause for all speakers. By contrast, intonation units other then followed by a pause show no clear-cut intensity patterns at their ends. This might be explained by the correlation between intensity and f_0: the units in question mostly correspond to prosodic continuation, where the f_0 increase would not allow for an intensity drop.
Intensity differences between immediately adjacent syllables left and right of a stress unit boundary are relatively small; there seems to be no tendency to compensate for the general downward intensity slope within the unit by a systematic between-unit intensity reset.

4.3 Speaker-specific variability

We can observe several instances of speaker-specific behaviour in our data: Speaker 1 has significant preference for stress-unit-initial rises, the other two speakers do not. Between stress units, Speaker 1 realizes significant falls, though with small magnitude, whereas the other two speakers do not. Speaker 3 has the "flattest" intensity contours of all three speakers in terms of final fall, pre-pause fall and post-pause resetting.

5. Conclusion

Macrointensity was studied within stress units and across their boundaries, with the objective to examine its rhythmic, accentual and phonosyntactic potential. All the conclusions are limited to the type of speech studied.
Putting the different pieces of analysis together, we can conclude that there is a gradual macrodynamic decrease over the inter-pause group, followed typically by a significant intensity reset. Within this general decrease, which we can call "intensity downtrend", there are systematic local intensity drops occurring between the last two syllables of stress units; in addition, there is a major intensity drop before the pause.

The accented syllable, which is leftmost in all stress groups included in the analysis, does not coincide with any systematic intensity pattern. The hypothesis of a positive intensity correlate of stress in Czech can be refuted once again.

Dynamic behaviour of speech is described at two prosodic levels: that of the stress unit and that of the intonation unit, both being units with other prosodic and linguistic correlates. In both cases, the identified tendencies are strongly right-boundary-related.

Intensity was presented in this paper in an autonomistic view, although its variations are believed to be in interaction with other prosodic parameters, such as final lengthening or intonation contours. In the first case, our data support the hypothesis of a trade-off relation between pre-boundary lengthening and dynamic decrease [12]. In the second case, intensity is sometimes described as being positively correlated with f_0 [12] [4], which has been proved at least for its "downtrend": a tempting hypothesis in this respect is that the distinction between declination, downstep/downdrift and final lowering, useful in the study of intonation [13], may be applied to the domain of intensity as well.

We have good reasons to believe that the described patterns, statistically significant and lying above the discriminability threshold, though limited to syllable nuclei, have perceptual existence. Further investigation of their perceptual impact, however, would be welcome.

6. Acknowledgement

This research was supported by the GAČR 405/04/P238 grant.

7. References

[1] Bartkova, K.; Haffner, P.; Larreur, D., 1993. Intensity Prediction for Speech Synthesis in French. *Proceedings of ESCA Workshop on Prosody*. Lund, 280–283.

[2] Duběda, T., 2002. Structural and quantitative properties of stress units in Czech and French. In *Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday*, A. Braun; H. R. Masthoff (eds.). Stuttgart: Steiner, 338–350.

[3] Duběda, T.; Keller, E., 2005. Microprosodic Aspects of Vowel Dynamics – an Acoustic Study of French, English and Czech. *Journal of Phonetics 33* (4), 447–464.

[4] Jokisch, O.; Kühne, M., 2003. An Investigation of Intensity Patterns for German. *Eurospeech 2003*, Geneva, 165–168.

[5] Keller, E.; Bailly, G.; Monaghan, A.; Terken, J.; Huckvale, M. (eds.), 2001. *Improvements in Speech Synthesis*. Chichester: Wiley and Sons.

[6] Ladd, D. R., 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

[7] Lehiste, I.; Peterson, G. E., 1967. Vowel Amplitude and Phonemic Stress in American English. In *Readings in Acoustic Phonetics*, I. Lehiste (ed.). The M. I. T. Press, 183–190.

[8] Maddieson, I., 1997. Phonetic Universals. In *The Handbook of Phonetic Sciences*, W. J. Hardcastle; J. Laver (eds.). Blackwell Publishers, 619–639.

[9] Pols, L. C. W., 1999. Flexible, robust, and efficient human speech processing versus present-day speech technology. *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, Vol. I., 9–16.

[10] Rossi, M.; Di Cristo, A.; Hirst, D.; Martin, Ph.; Nishinuma, Y., 1981. *L'intonation: de l'acoustique à la sémantique*. Paris: Klincksieck.

[11] Sluijter, A. M. C.; Heuven, V. J. van; Pacilly, J. J. A., 1997. Spectral balance as a cue in the perception of linguistic stress. *Journal of Acoustical Society of America* 101 (1), 503–513.

[12] Vaissière, J., 1983. Language-Independent Prosodic Features. In *Prosody. Models and Measurements,* A. Cutler (ed.), 53–66.

[13] Volín, J., 2004. Linear Regression as a Method of Detecting F0 Declination. *Speech Processing, 13th Czech-German Workshop,* R. Vích (ed.). Prague, 69–72.