

PITCH AND NON-PITCH CUES TO WORD STRESS IN CZECH

Christine Bartels

University of Massachusetts, Amherst, U.S.A.

ABSTRACT

This study investigated the role of pitch, duration, amplitude, and vowel formants in word stress in Czech. It was found that stress and pitch peak tend to diverge, in that the latter occurs more often on the syllable following the stressed, first syllable than on the stressed syllable itself. Further, the placement of the pitch peak is dependent on vowel quantity and vowel height: both long and high vowels are relatively favored. Among the remaining cues, syllable duration outranked amplitude and vowel formant ratios.

INTRODUCTION

Czech is traditionally described as a stress-accent language having fixed word stress on the first syllable, independent of vowel quantity [1,2]. It shows high and low pitch accents and characteristic phrasal intonation contours but does not use tone for lexical contrast. Pitch accents generally fall on the initial, stressed syllable. This correlation is not absolute, however; the pitch accent is frequently placed later in the word, at times leading to the perception among non-native listeners that this later syllable is the more prominent one [1]. The production study reported here investigated the role of two contextual factors, vowel height and vowel quantity, in conditioning such divergent placement of stress and pitch peak in Czech. These factors were chosen because the informal observation is that high and long vowels seem particularly effective in attracting the pitch peak. The study also examined whether other acoustic measures serve as reliable correlates of word stress in Czech. In the non-tonal stress languages

English [3] and German [4], duration and amplitude may have equal or greater importance than F_0 in marking the stressed syllable, and spectral characteristics also contribute to conveying word stress in English [5]. Even within a single language, different levels of the prosodic hierarchy may put different weight on these acoustic cues in coding prominence [6,7]. By looking at a stress language that is untypically unreliable in its use of pitch cues, this study aimed to contribute to the growing experimental evidence for the phonetic diversity hiding behind the traditional notion of "stress."

METHOD

Nine native speakers of Czech, five women and four men, were asked to produce contextually embedded trisyllabic nonsense words of the form /nV(:).nV(:).na/. The vowels in the first and second syllable were either [i] or [a], and either short or long. Each speaker read these words twice in different random orders, with each word embedded in the frame sentence *Řekni — ještě jednou* ('Say — once more'), which places main sentence prominence on the test word. The recorded test words (288 tokens in all) were digitized at 10 kHz and the following acoustic measures taken from waveform, F_0 contours, or spectrograms of the first two syllables: syllable and vowel duration, average, peak, and total rms amplitudes, and average $F1$ and $F2$ frequencies from the central 50 ms of the vowel's steady state. Vowel quantity (short vs. long), vowel quality ([a] vs. [i]), and where appropriate, syllable position served as the independent

variables in separate repeated measures ANOVAs for each of the acoustic measures listed above, as well as for certain first/second syllable ratios. To ascertain that the absolute timing of the F_0 peak relative to word onset did vary and that its variable alignment was not just a reflection of variation in syllable length, a simple regression analysis was carried out with duration of the first syllable as the independent variable.

RESULTS

Pitch excursions were rather small throughout for these Czech speakers. Only five out of nine subjects, referred to as the H* group below, showed clear peaks in the test word; the others did not exhibit a consistent pattern. Even within the H* group, speakers varied considerably in the range of timing produced.

If a token showed a distinct pitch peak, it occurred no later than the second syllable. Within the first two syllables, the absolute timing of the F_0 maximum with respect to word onset and syllable boundaries varied widely, but peaks rarely occurred before the second half of the first vowel and were always preceded by a distinct, gradual rise.

Shift of the F_0 maximum to the second syllable was the rule rather than the exception in these data: In 58.3% of the tokens, the ratio of the F_0 maxima for first and second syllable was lower than predicted based on average peak values for the vowels and quantities involved in a given token; thus, only 41.7% of the time did the overall pitch peak correctly identify the stressed syllable. Averaging F_0 maxima across subjects for first and second syllable, no significant difference was found. Averaging across the H* group only, the mean values were 193.7 Hz and 197.4 Hz, respectively, a minor difference, but one that favors the second syllable.

When tested across all nine speakers, in the first syllable but not the second peak height correlated significantly with quantity [$F(3,51) = 8.736; p = .0001$] and vowel quality [$F(3,51) = 4.574; p = .0065$], being higher for long vowels and for [i]. For long but not short first syllables, peak height of either syllable correlated with the vowel height of the first vowel [$F(9,153) = 2.672; p = .0066$]. The same test carried out for the five H* speakers alone yielded an additional main effect for syllable position itself [$F(1,9) = 9.345; p = .0136$], with the F_0 maximum being slightly higher for the second syllable (see above), but the dependence of the second-syllable maximum on the height of a long first vowel was no longer significant.

The first/second syllable ratio of F_0 maxima also depended strongly on both vowel quantity and vowel quality. On the one hand, a long first vowel but not a short one attracted the overall pitch maximum [$F(3,51) = 7.365; p = .0003$]. On the other, a high second vowel following a low first vowel, i.e. the combination /na(:)ni(:)na/, shifted the peak to the second syllable [$F(3,51) = 3.797; p = .0237$]. In cases of conflict, neither effect dominated consistently. When the same analysis was carried out for the H* group only, similar main effects were found: a long first vowel attracted the pitch more than a short one [$F(3,27) = 17.593; p = .0001$], and a high first vowel did so more than a low one [$F(3,27) = 7.094; p = .0012$]. However, for this group the ratio was >1 , i.e., the overall pitch maximum was found to fall on the first syllable, only if that syllable contained a long [i].

The effects of vowel quantity and vowel quality on F_0 syllable averages were similar, but because of the word-initial rise from a lowpoint in H* peak contours, this measure yielded an even

lower score as a stress cue: only in 21.5% of tokens did the ratio of F_0 averages, adjusted for vowel properties, favor the first syllable.

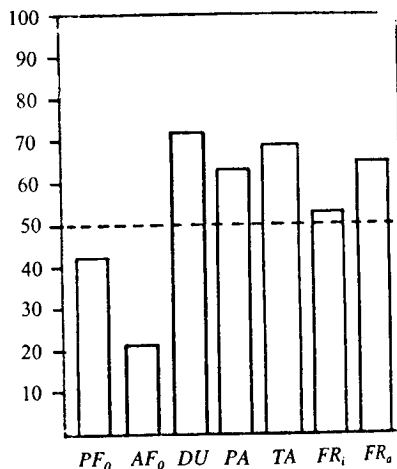
Among the other stress cues examined, syllable duration was the one most frequently employed. (Vowel duration was a much less reliable measure.) The average syllable durations, by vowel type, were: short [a], 187.4 ms; long [a], 330.1 ms; short [i], 174.7 ms; long [i], 278.1 ms. The ratio for first vs. second syllable for a given token based on these values was exceeded in 71.5% of all tokens.

Average syllable amplitude and peak syllable amplitude also played a role in cuing for stress. Adjusted for vowel height and length effects, they favored the first syllable in 59.0% and 63.9% of all tokens. Because of the bias in syllable durations, total syllable amplitude, at 68.1%, did even better.

Formant values contributed negligibly if the vowel involved was [i]: the ratio F_2/F_1 , expected to increase for more "hyperarticulated" forms of the vowel and in fact doing so in longer vowels here, exceeded average values for these speakers only 52.8% of the time in first-syllable [i]'s. Formant values contributed somewhat more in [a]'s: both F_1 and F_2 were significantly higher in first than in second syllables, while at the same time, in a stronger, somewhat opposing trend toward "hyperarticulation," the second formant decreased with vowel length. Thus both F_1 alone and the mean of the two formants exceeded average values 59.7% of the time in first-syllable [a]'s, F_2 alone exceeded the average 61.1% of the time, and the ratio F_2/F_1 stayed below the average in 63.9% of all tokens containing first-syllable [a]'s.

Figure 1 summarizes these results.

Figure 1. Percentage of tokens in which the acoustic cues measured correctly identified the stressed syllable. PF_0 , peak F_0 ; AF_0 , average F_0 ; DU , syllable duration; PA , peak amplitude; TA , total amplitude; FR_i , formant ratio [i]; FR_a , formant ratio [a].



DISCUSSION

This study confirms the listener's impression that pitch, in terms of F_0 maximum and average, is not a reliable cue to stress in Czech; in fact, in the type of data examined here, it shows a tendency to peak *after* the stressed first syllable. As suspected, this tendency turns out to be strongly dependent on segmental factors such as vowel height and vowel length, in that both long and high vowels attract pitch.

Of the typical non-pitch cues to stress, i.e. duration, amplitude, and spectral characteristics of vowels, only duration comes close in Czech to the role it plays in English, where it appears to be the least reliable cue [3]. Thus the results of this study are in agreement with the general impression that Czech has weak

stress accents only, compared to other Slavic languages. However, duration and amplitude outrank pitch; this fact matches the conclusions of an early perception study conducted with Czech listeners [8].

It appears plausible, though, that in the H^* speakers at least, the prolonged word-initial *rise* in F_0 has become relevant; this cue is perceptually salient but is not usually considered in quantitative studies. In a range of languages, for instance Swedish [9] and Japanese [10], dialects differ as to the temporal alignment of pitch accents and pitch movements with syllable structure, usually in the direction of relative delay of movement endpoints. Thus one might speculate that to the extent that Czech speakers use a distinctive pitch accent at the word level, once the strong force of segmental effects has been accounted for, the temporal placement of this peak presents a compromise between the need to align it with the word stress and the attempt to maximize the salience of the preceding rise.

ACKNOWLEDGMENTS

I gratefully acknowledge the technical assistance and helpful comments I received from John Kingston, as well as the discussions I had with Lisa Selkirk and John McCarthy. Thanks to members of the Institute of Formal and Applied Linguistics and the Institute of Theoretical and Computational Linguistics of Charles University, Prague, as well as to the members of the Institute of Experimental Biology of the Czech Academy of Science, Olomouc, for their willingness to serve as subjects, and to Julia Hirschberg, Lou Conover, and Yuzo Fujikura for their active support. This work was sponsored in part by the U.S.-Czech/Slovak Science and Technology Joint Fund, project no. 92058.

REFERENCES

- [1] Broch, O. (1911), *Slavische Phonetik, Heidelberg: Carl Winter's Universitätsbuchhandlung.*
- [2] Jakobson, R. (1962 [1926]), "Contributions to the study of Czech accent," in *Selected Writings I: Phonological Studies*, The Hague: Mouton, pp. 614-625.
- [3] Beckman, M. (1986), *Stress and Non-Stress Accent*, Dordrecht: Foris.
- [4] Kohler, K. (1991), "Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology, and semantics," in K. Kohler (ed.), *Studies in German Intonation*, Arbeitsbericht Nr. 25, Institut für Phonetik, Kiel University.
- [5] de Jong, K. (1995), "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *J. Acoust. Soc. Am.* 97, 491-504.
- [6] Beckman, M., and J. Edwards (1994), "Articulatory evidence for differentiating stress categories," in P. Keating (ed.) *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology 3*, Cambridge: Cambridge University Press, pp. 7-33.
- [7] Shattuck-Hufnagel, S., M. Ostendorf, and K. Ross (1994), "Stress shift and early pitch accent placement in lexical items in American English," *J. Phonetics* 22, 357-388.
- [8] Janota, P. (1967), "Perception of stress by Czech listeners," in *Proceedings of the Sixth International Congress of Phonetic Sciences*, pp. 457-461.
- [9] Bruce, G. (1983), "Accentuation and timing in Swedish," *Folia Linguistica* 17, 221-238.
- [10] Pierrehumbert, J., and M. Beckman (1988), *Japanese Tone Structure*, Cambridge, Mass.: MIT Press.