Remarks on production and perception of standard Serbo-Croatian tonal accents: A reply
Keijsper, C.E.

Published in:
Studies in Slavic and General Linguistics

Citation for published version (APA):
REMARKS ON PRODUCTION AND PERCEPTION OF STANDARD
SERBO-CROATIAN TONAL ACCENTS: A REPLY

C.E. KEIJSER

1. In section 3 of her reaction to my paper "Studying Neoštokavian Serbocroatian prosody", Gvozdanović (henceforth: G) states that on the basis of Ivvič and Lehiste's measurements (henceforth I&L), two hypotheses about the acoustic correlates of accent can be formulated. One of these hypotheses assumes that it is a change in the fundamental frequency pattern which is the relevant cue to accent; it is said to be the one defended in G 1980. ¹

G's formulation of this hypothesis in her Remarks differs, however, from the version defended in G 1980, as well as from the other versions of the hypothesis she proposed at one time or another (see the following sections). All versions of the hypothesis are claimed to be based on I&L's data. However, I&L's data are incompatible with G's hypothesis. Possibly, G has now become aware of this fact, because her present version is full of "systematically"s and "normally"s. Unfortunately, she fails to specify how she would account for the "unsystematic" and "abnormal" cases, in other words for the counterexamples to her proposal, say those in I&L's data. Also, she now seems to be aware of the fact that her application of the term "peak" needs modification, because she now distinguishes between syllables with a distinct peak and syllables without a distinct peak, and she even recognizes syllables where the peak "is absent in any systematic way". Unfortunately, she fails to specify which peaks are distinct, which peaks are not distinct, which peaks are present, and which peaks are absent. In this way, she tries to uphold her hypothesis while denouncing it at the same time.

In order to explain my understanding of G's Remarks, I shall first list three earlier versions of G's theory, viz. the ones presented in her first article (G 1972 (= Santen 1972))(section 2
below), in her dissertation (G 1980)(section 3 below), and in G 1984-1985 (sections 4-5 below). For every version I shall indicate types of pitch configuration represented in I&L's data which remain unaccounted for in the given proposal. In order not to complicate matters unduly, I will confine myself to the following 14 pages in I&L's work: I&L 1963: 38-39, 44-45, 50-51, 55; I&L 1965: 81-87. These pages contain data on the pronunciation of the four accents in "neutral" environments, i.e. when the influence of sentence intonation is regarded to be minimal (see Chapter Three of my paper Studying). Further, I shall adopt G's point of view on what is a peak, i.e. we will be talking about what I have called "peak3" in section 3.5 of Studying (the highest point reached during any vowel). In sections 6-7 below I shall return to G's Remarks.

2. G (1972: 101) divides vowels into [+tonal] and [-tonal] vowels, rather than into accented and unaccented vowels. The category of [+tonal] vowels includes accented and preaccentual vowels; they have their peak3 later than at 25% of their duration (ibid.). The category of [-tonal] vowels includes postaccentual vowels; they have their peak, in the first 20% of their duration (ibid.). The statement is said to be based on I&L's data (ibid.). From the references (op.cit.: 100-101) it appears that G has consulted, inter alia, I&L 1963, I&L 1963; 1965; 1967; 1969.

Taking only the 14 pages in I&L 1963 and 1965 specified above, we also find the following possibilities, however:
- Preaccentual vowels having peak3 in the first 25%. For example: the first syllable of type "---" as spoken by Ivič (I&L 1963: 45): here peak3 occurs at the starting point of the vowel (0% or 14.3% - see I&L 1963: 35-36); the first syllable of type "~" as spoken by Ivič (I&L 1963: 51): here peak3 occurs at the starting point of the vowel (0%/12.8%).
- Accented vowels having peak3 in the first 25%. For example: the second syllable of type "----" as spoken by Ivič (I&L 1963: 51) has peak3 at 13.5%; the first syllable of type "~" as spoken by D1 (I&L 1965: 81) has peak3 at the starting point of the vowel (0%/8%).
- Postaccentual vowels having peak3 later than at 20%. For example: the second syllable of type "~" as spoken by D1 (I&L 1965: 81): peak3
occurs at 86.7%; the second syllable of type "" as spoken by D6 (I&L 1965: 83): peak\textsubscript{3} occurs at 86.7%.

In effect, G's (1972: 101) claim that in a sequence 'vowel having peak\textsubscript{3} later than at 25% - vowel having peak\textsubscript{3} earlier than at 20%' the first vowel is accented, can hardly be correct: it seems to predict that in, for example, D6's "" (25.7% - 86.7% - 0%/10.9%) the second syllable is accented, and it seems to predict that in, for example, Ivić's "" (49.4% - 13.5% - 30% - 14.0%) the first and the third syllables are accented.

In addition, G 1972 fails to supply a correct summary of I&L's findings concerning the difference between a so-called falling and a so-called rising accent. On page 103 we learn that [+tonal] vowels (peak\textsubscript{3} later than at 25%) can be divided into [+rising] and [-rising] (and into [+long] and [-long]). [+Tonal] [+rising] vowels have a peak\textsubscript{3} later than at 50%, [+tonal][-rising] vowels have a peak\textsubscript{3} between 25% and 50%. The last [+tonal] vowel in an "accentual word" (prosodic word) is [+rising] or [-rising]. If such a [+tonal] vowel is also the first vowel, the features [+rising] or [-rising] are distinctive (i.e. there is a tone contrast in word-initial syllables). If such a [+tonal] vowel is not the first vowel, the vowel is automatically [+rising] (i.e. non-initial syllables cannot have a falling accent). If such a [+tonal] vowel is the only vowel in the word, the vowel is automatically [-rising] (i.e. monosyllabic words can only have a falling accent). All [+tonal] vowels preceding the last [+tonal] vowel are automatically [-rising] (i.e. there is no tone contrast in preaccentual vowels, and such vowels have their peak\textsubscript{3} between 25% and 50%) (ibid.). Here, I have indicated in brackets the information which the author probably wishes to convey. However, if we start from I&L's data, we arrive at the following possibilities:

1. Preaccentual vowels are
   - [-tonal] if they have a peak\textsubscript{3} earlier than at 20% (for examples see above);
   - [+tonal][-rising] if they have a peak\textsubscript{3} between 25% and 50% (for example: the first syllable of type "" as spoken by Ivić (I&L 1963: 55) has peak\textsubscript{3} at 44.9%);
   - [+tonal][+rising] if they have a peak\textsubscript{3} later than at 50% (for example: the first syllable of type "" as spoken by Ivić (I&L 1963:
55) has peak\(_3\) at 54.5\%).

2. Accented vowels are
- [-tonal] if they have a peak\(_3\) earlier than at 20\% (for examples see above);
- [+tonal][-rising] if they have a peak\(_3\) between 25\% and 50\% (for example: the first syllable of type " as spoken by D7 (I&L 1965: 84) has peak\(_3\) at 31.3\%; the first syllable of type " as spoken by E11 (I&L 1965: 85) has peak\(_3\) at 38.4\%);
- [+tonal][+rising] if they have a peak\(_3\) later than at 50\% (for example: the first syllable of type " as spoken by D7 (I&L 1965: 84) has peak\(_3\) at 79.6\%; the first syllable of type " as spoken by E11 (I&L 1965: 85) has peak\(_3\) at 84.1\%).

3. Postaccentual vowels are
- [-tonal] if they have a peak\(_3\) earlier than at 20\% (for example: the second syllable of type " as spoken by E14 (I&L 1965: 87) has peak\(_3\) at the beginning of the vowel (0\%/12.2\%); the second syllable of type " as spoken by E14 (I&L 1965: 87) has peak\(_3\) at the beginning of the vowel (0\%/12.8\%));
- [+tonal][-rising] if they have a peak\(_3\) between 25\% and 50\% (for example: the second syllable of type " as spoken by E11 (I&L 1965: 85) has peak\(_3\) at 29.9\%; the second syllable of type " as spoken by E11 (I&L 1965: 85) has peak\(_3\) at 45.5\%);
- [+tonal][+rising] if they have a peak\(_3\) later than at 50\% (for examples see above).

We have to conclude that G 1972 fails to distinguish between accented and unaccented syllables, between preaccentual and postaccentual syllables, and between falling and rising accents. It seems to me (cf. also L&I 1986: 161-162) that the paper is based on a few tables listing average data per type of vowel for a single speaker: I&L 1963: 36-37, 42-43, 49-50, 53-54. If one looks only at these tables, G 1972 still contains a number of mistakes, but in subsequent versions (see below) G eliminates most of these mistakes, by replacing [+tonal] and [+rising] by other categories. But she persists in disregarding the rest of I&L's data.

3. In her dissertation (G 1980), G takes a fresh look at the
subject. As I mentioned in 3.5 of *Studying*, here G distinguishes between falling fundamental frequency (peak₃ in the first 25%), nonfalling nonrising fundamental frequency (peak₃ between 25% and 75%), and rising fundamental frequency (peak₃ later than at 75%). Falling fundamental frequency can be high (i.e. higher than that of the preceding syllable nucleus (vowel)), low (i.e. lower than that of the preceding syllable nucleus), or indifferent (i.e. neither higher nor lower than that of the preceding syllable nucleus); it occurs in postaccentual syllable nuclei. Nonfalling nonrising fundamental frequency can be high or non-high; it occurs in preaccentual and accented syllable nuclei, where it is followed by nonfalling and falling fundamental frequency, respectively. Rising fundamental frequency occurs in rising accented syllable nuclei (i.e. with a so-called rising accent); it is followed by high falling fundamental frequency. Rising accented syllable nuclei are either rising or nonfalling nonrising; the next syllable nucleus is high (G 1980: 35-36).

It will be clear that all possible types of counterexample to this statement have already been given in section 2 above, except for the following:
- Preaccentual syllables having a peak₃ later than at 75%; for example: the first syllable of type '----' as spoken by Ivic (I&L 1963: 45) has peak₃ at the end of the vowel (100%). Further, it appears from G’s discussion that she has overlooked the possibility of an accented syllable with a so-called falling accent and a peak₃ later than at 75%. An example is the first syllable of D8’s ' ' (I&L 1965: 84), which has peak₃ at 82.5%. As to G’s categories of high, non-high, low (and non-low), here no counterexamples can be given, as the author fails to specify in her text what meaning must be assigned to these labels: since, for example, a syllable with a so-called falling accent may contain a fall of, say, 6 semitones, it is not unimportant to specify the point with respect to which the next syllable is higher or lower. As appears from the schematic representations reproduced in 3.5 of *Studying*, G has unfortunately chosen the wrong point, at least for postaccentual syllables following falling accents.

Next, the author proceeds to assume that prosodic word boundaries are signalled by a change from falling to nonfalling fundamen-
tal frequency or from nonfalling to high or low nonfalling fundamental frequency, and that accent (prominence) is signalled by a change from nonfalling to falling fundamental frequency (G 1980: 58). In effect, "within each prosodic word, at least one non-falling syllable nucleus occurs. Preceding it, only non-falling syllable nuclei occur, and following it, only falling syllable nuclei occur. A non-falling syllable nucleus which is followed by a syllable containing a falling syllable nucleus is perceived as accented" (G 1980: 99). As we saw in the foregoing, every part of this statement is wrong. To take a fresh set of examples,

- Type "" as spoken by D3 (I&L 1965: 82) is a prosodic word without nonfalling syllables: peaks occur at 18.3% and at 21.0%;
- Type "" as spoken by Ivčić (I&L 1963: 45) is a prosodic word in which a preaccentual syllable has a falling fundamental frequency: peaks occur at 20%, 45%, 100%, 55.6%, 0%/14.3%;
- Type "" as spoken by D2 (I&L 1965: 82) is a prosodic word in which a postaccentual syllable has a nonfalling fundamental frequency: peaks occur at 0%/9.7% and 45.8%;
- Type "" as spoken by Ivčić (I&L 1963: 39) has an accent on the second syllable, not on the first syllable; peaks occur at 38.5% and 23.4%;
- Type "" as spoken by Ivčić (I&L 1963: 45) has an accent only on the first syllable, not on the first, third (and fifth) syllable; peaks occur at 92.3%, 8.6%, 33.3%, 0%/15.4%, 41.2%.

Since the author based her reasoning on a non-existent set of data, it was excluded beforehand that the perception test reported in G 1980 could make any sense: G starts from the assumption (1980: 64) that, if she combines the first syllable of u kūču with the second and third syllables of ū kuču, she is combining a syllable having nonfalling fundamental frequency with syllables having falling fundamental frequency. However, this need not be the case. If G supplies the raw data of the pairs out of which she has constructed combinations of (originally) unaccented syllables, say data concerning intensity, vowel duration, and F0 at the beginning, peak and end of every vowel, the crucial part of her results (cf. L&I 1986: 166-167) can possibly be explained.
4. Before we proceed, it may be useful to make explicit that, even if one forgets about G's peak, prosodic word, high/low fundamental frequency, etc., her hypothesis about prominence perception is still disproved by I&L's data. In that case the crucial counterexamples are a subset of the counterexamples listed in the foregoing, namely the subset in which peak equals peak, i.e. where the highest point reached during a given vowel happens to be the highest point of the entire configuration. This subset includes the following types:
- Falling accents having peak in the first 25% of the vowel in the accented syllable (here G predicts an accent on the preaccentual syllable);
- Rising accents having peak later than at 25% of the vowel in the postaccentual syllable (here G predicts an accent on the postaccentual syllable).

As we will see in the following sections, in G 1984-1985 and Remarks G pays special attention to, at least, the first peak case mentioned here. The special attention is justified, because G's hypothesis about prominence perception is incompatible with the existence of the cases mentioned here. However, instead of denying the existence of the data or giving up her hypothesis, which would be the logical possibilities, G tries both to account for the data and to save her hypothesis, which is impossible, as we will see.

5. After G 1980, G seems to have become aware of the fact that syllables with a so-called falling accent may have a peak in the first 25% of the duration of the vowel (I&I 1986: 162), and/or that postaccentual vowels may have a peak later than at 25%. At any rate, in G 1984-1985 the author distinguishes between [+rising tone], [-rising tone], and nonrising contours. The nonrising contours are further divided into nonrising nonfalling and falling contours (G 1984-1985: 177). G 1984-1985 does not give the percentages to which these terms are intended to apply; I take them to be the same as in G 1980. But what is important is that the new terminology separates accented syllables from unaccented syllables:
- [+rising tone]: accented syllable, peak in the last 25% of the
vowel;
- [-rising tone]: accented syllable, peak₃ in the first 75% of the vowel (so that peaks₃ in the first 25% are now included);
- nonrising contour: unaccented syllable, peak₃ in the first 75% of the vowel: nonrising nonfalling (25%-75%) or falling (0%-25%).

The beginning of a prosodic word is signalled by a nonrising nonfalling contour, the end of a prosodic word is signalled by a falling contour (ibid.).

Leaving aside the fact that at least unaccented syllables are still described incorrectly here, we now need to know whether or not a given syllable is accented in order to know whether a peak₃ in the area from 0% to 75% comes under the heading of [-rising] or of nonrising. In other words, the hypothesis saying that accent is predictable from tone and prosodic word boundaries becomes internally inconsistent (whereas before, it merely did not entertain a relation with reality), since we can no longer determine the prosodic word boundaries in G's set of data without knowing the place of the accent in the same data. Thus, as I intentionally said less explicitly in 3.5 of Studying, G's attempt to eliminate one type of counterexample to her proposal immediately denounces her earlier publications on the subject.

From this point onwards, things can only become worse, unless G gives up her proposal and returns to the point preceding G 1972 where things started to go wrong. Unfortunately, she proceeds to write Remarks.

6. The most conspicuous new elements in G's Remarks are words and phrases like "systematic(ally)", "normal(ly)", "distinct peak", "without a (systematically occurring) distinct peak". The author unfortunately fails to explain the new terminology, so that I can only list some questions which the author may wish to clarify in her next series of versions:
- In all cases where Remarks uses the words "systematic(ally)" or "normal(ly)", how does G account for the "unsystematic" and "abnormal" cases?
- Which peaks₃ occurring in the first 25% of the duration of a vowel are distinct peaks, and which are not distinct peaks?
- Which peaks₃ occurring elsewhere than in the first 25% of the duration of a vowel are distinct peaks, and which are not distinct peaks?

- What is the difference between a non-distinct peak and a peak which is absent?
- Does "nonrising" mean "peak₃ not in the last 50%" or "peak₃ not in the last 25%"?

- What is the relationship between Purcell's (1976) findings concerning peak₂ location in long accented syllables and peak₃ location?
- Does "falling" mean "peak₃ in the first 50%" or "peak₃ in the first 25%"?
- What is the new definition of "relatively high/non-high/low/non-low"?

- What is the prosodic difference between 1. two prosodic words resulting from "contrastive" usage of a combination of a proclitic word and an "autonomous" word; and 2. one prosodic word resulting from "noncontrastive" usage of the same combination?
- How does one know whether a given peak₃ location, say a peak₃ at 40%, signals a prosodic word boundary, and how does one know whether the signal concerns the beginning or the end of a prosodic word?
- What happens to prosodic word boundaries and to prominence perception if peak₃ location in, say, postaccentual syllables, is affected by sentence intonation?

- Where did G find the information that word-initial falling accents with a peak₃ in the first 25% seem possible only in unaccented sentence-final positions? What about the cases mentioned in L&I 1986: 162?

- What is the source of G's knowledge that word-initial preaccentual syllables in dialects where no non-initial falling accents occur, are characterized by a relatively high fundamental frequency? If there is no source, how do these people hear G's prosodic word boundaries?
- Which are the areas where non-initial falling accents cannot occur?
- If accent is predictable from tone (without prosodic word boundaries) in all areas where non-initial falling accents occur, what aim is served by retaining prosodic words?
- Is there a motivation for assuming that accent is predictable from tone other than the fact that an accented short vowel may be followed by an unaccented long vowel?
- If accent is independent of tone if duration is the relevant cue to accent, and if accent is predictable from tone if pitch is the relevant cue to accent, what about the possibility that both duration and pitch contribute to prominence perception?  
- Is there a reason to assume that only either duration or pitch is a cue to accent?

Alternatively, as I said in 3.5 of Studying, G might wish to start considering the possibility that Jakobson had a bad day when he thought up Jakobson 1931, and that G had a bad day when she got it into her head that the "evidence from various languages showing that a change of the fundamental frequency pattern is perceived by the native speakers of those languages as prominent, i.e. accented" pertains to a sequence of peaks, one on every syllable, rather than to one peak on a sequence of syllables.

7. Apart from the fact that since 1972, G has been misusing I&L's work by stating that it provides the factual basis for her speculations, in Remarks she does grave injustice to I&L by suggesting that, although I&L have obtained production data, it is G (1980) who has related production to perception. In reality, I&L have conducted a series of perception tests, and the single test reported in G 1980 has contributed nothing to our understanding of Serbocroatian prosody, because it is based on incorrect quotations from, inter alia, I&L's work. The description "in relative terms, i.e. as discrete, relative and oppositional values" which G gives in section 2 of Remarks, is entirely I&L's in so far as it is correct; compare, for example, the quotation from L&I 1963 given at the beginning of 3.3 of Studying. (G's reference (ibid.) to Purcell 1976 in her description of the difference between the two short accents in some dialects must be a slip of the pen, because Purcell 1976 investigates the perception of peak₂ location in long vowels, in Serbocroatian in general.)

It is true, as G observes in the last sentence of her Remarks, that G's "level of classification has neither been aimed at nor reached by" I&L, but the cause of this fact is evidently not that I&L presumably did not pay sufficient attention to perception. Rather,
I&L understand too much of their subject to aim at the kind of "solid basis" created by G, and if they did not have this understanding, they would still be too honest to reach a level of classification which can be arrived at only by applying G’s "method of abstraction from all the predictable, unsystematic and inconclusive characteristics".

Leiden University

NOTES

1 The references to this paper can be found in the reference sections of Gvozdanović's "Remarks on production and perception of Standard Serbo-Croatian tonal accents" and of my paper "Studying Neostokavian Serbo-Croatian prosody", both published elsewhere in this volume.

2 G’s picture suggests that at the beginning of the postaccentual vowel following a falling accent, pitch is much higher than it is at the end of the accented vowel. Even without consulting I&L’s data one can see that something must be wrong here, because the upward jump between the end of the accented vowel and the beginning of the postaccentual vowel in G’s picture is a typical mode-1 rise, i.e. the type of rise we find at the beginning of an accented syllable with a falling accent (see 3.2 of Studying). In a realistic picture, the highest point of the postaccentual vowel would have to be not only lower than the highest point of the accented vowel (cf. G’s "relatively low"), but also not higher than the end of the accented vowel. In other words, there is no peak in the postaccentual syllable following a falling accent.

In my own proposal (3.6 and 3.7 of Studying), I depart from the idea that the height of postaccentual syllables must be related to that of accented syllables, but here I start from the information which G could have found in I&L’s work; I am not blaming her for the fact that she has made no contribution to the fund of knowledge available at the time.

3 In 3.5 of Studying, I cited G’s 1984-1985 summary of her 1980 result, viz. "In other words, only the syllable with a clear F₀ peak can be perceived as accented". I added the following comment: "which is true if one does not regard the peak in the postaccentual syllable after rising accents as a clear F₀ peak". I now regret having formulated my criticism implicitly here, because in Remarks, G starts from the idea that all peaks in postaccentual syllables must be eliminated (by calling them non-distinct or absent). Her formulation in Remarks amounts to the statement that the highest point of the configuration is not a distinct peak if it occurs in the postaccentual syllable. This is, of course, incorrect. A more explicit version of my comment on G’s summary of her 1980 result runs as follows: "which is true if one does not regard the peak in the postaccentual syllable after rising accents as a clear F₀ peak; however, since the highest point of the configuration of rising accents mostly occurs in the postaccentual syllable, and since it would be nonsensical not to regard the highest point of the configuration as a clear F₀ peak, this highest point being the only clear F₀ peak, it is not true that only the syllable with a clear F₀ peak can be perceived as accented:
with rising accents, it is mostly the syllable before the syllable with the clear F\textsubscript{0} peak which is perceived as accented; it is this fact which makes it necessary to search for cues to accent other than the location of the clear F\textsubscript{0} peak".

I would like to explain the history of the present discussion between G and myself.

As is, I think, normal in scientific communities, before deciding on publication of Studying, I asked several people to read through, and to comment upon, parts or all of the paper. To G I gave a copy of the draft of section 3.5, requesting her to check whether I had reproduced the 1984-1985 version of her hypothesis correctly: since G 1984-1985 does not make explicit that it contains a new version as compared with G 1980 (the same holds true, in fact, for all subsequent versions, as compared with the respective preceding ones), I was not entirely sure whether I had kept track of the development in G's reasoning. On that occasion I invited her to discuss things among ourselves, because I realized that the contents of 3.5 might be painful to her.

Instead of answering my request and/or accepting the invitation, G reacted by writing Remarks and handing it to one of the editors in cameraready form. To the best of my knowledge, she had not seen more of my paper than section 3.5 when she wrote her reaction (plus the end of 3.4 and the beginning of 3.6, which happened to be on the pages where the draft of 3.5 started and ended). Therefore, the subtitle added to her paper is, for all I know, not quite accurate. For the same reason, I have refrained from commenting on points other than those derivable from 3.5 of Studying.

The editors kindly informed me of the existence of G's paper and, when G failed to contact me, provided me with a copy and allowed me to write a reply. I regret it very much that G did not tell me or write to me about her plans to react to my paper publicly. I could have given her the rest of Studying, so that she would have had a fuller understanding of what she was reacting to. More specifically, she might have wished to modify her reaction if she had seen my opinion on the following issues crucial to her position:

- The relation between tone and accent (see Chapter One of Studying, especially 1.4);
- The relative contribution of duration and pitch movement to prominence perception (see 2.1 of Studying);
- The relation between (prosodic) word boundaries and "contrastive"/"noncontrastive" usage of accents (see 2.3 of Studying);
- The relation between L\&L's acoustic data and perception (see 3.1 ff. of Studying).