The influence of syllable structure on lexical tones in Croatian: no effect?

Marija Tabain[1], Mate Kapović[2], Matthew Gordon[3], Adele Gregory[1], Richard Beare[4]

[1] La Trobe University, [2] University of Zagreb, [3] University of California Santa Barbara, [4] Monash University and Murdoch Children's Research Institute
m.tabain@latrobe.edu.au, mkapovic@ffzg.hr, mgordon@linguistics.ucsb.edu, a.gregory@latrobe.edu.au, richard.beare@monash.edu

ABSTRACT

Tonal contrasts in many languages are restricted to syllables with overall greater sonority (i.e. syllables containing long vowels or sonorant codas). However, studies to date have tended to focus on languages with relatively restricted syllable structures. In this study we focus on Croatian (Split variety), where both falling and rising lexical pitch accents are possible on all syllable types (unlike many Western South Slavic varieties that restrict tonal contrasts to long vowels), including those containing short vowels and those with an obstruent coda. We examine the timing of pitch peaks and troughs in recordings from 13 speakers who read a list of words varying in lexical pitch accent (rising or falling), vowel duration (long or short) and syllable type (Open, Closed-by-an-Obstruent or Closed-by-a-Sonorant). We find that although relative timing of pitch events varies according to lexical pitch accent, it is remarkably stable across syllable type for a given accent.

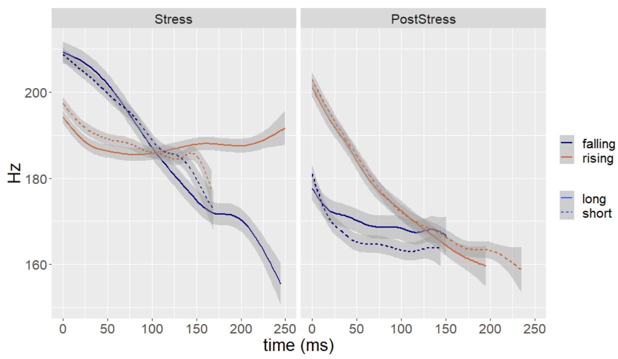
Keywords: lexical pitch accents, timing, syllable structure, Croatian

1. INTRODUCTION

Tonal contrasts in many languages are restricted to syllables with greater overall sonority [1, 2]. Certain languages, e.g. Navajo, thus limit contour tones (falls, rises) to syllables containing a long vowel in the rime, while others, e.g. Kiowa, restrict contours to syllables with either a long vowel or a sonorant coda. These restrictions can be attributed to the relative capacity of different syllables to support f0: syllables with greater sonority provide a better backdrop for tone [2, 3], allowing for crowding of tones that require more time to implement based on number of f0 targets (one, two or more than two) and direction of f0 movement (rising vs. falling tones).

Mitigation of tonal crowding can also manifest itself in gradient fashion through scaling effects, i.e. reducing the f0 distance between high and low targets (undershoot), or temporal shifts, i.e. increasing the distance between tone targets and/or lengthening or adding segmental material on which tones are realized. These strategies are observed not only within syllables but also across longer spans.

This paper explores the role of syllable structure in the realization of tone contrasts in the Split variety of Croatian [4, 5, 6, 7, 8, 9]. As in Standard Croatian, Split Croatian possesses a contrast between a rising and a falling accent on stressed syllables, where the realization of the complete contour often extends into the post-tonic syllable. In the Croatian linguistic orthography, accent type (or "prosodeme") is crossed with vowel length to yield the four categories of short falling, short rising, long falling, and long rising: *pȁra* (SF) 'steam', *pàra* (SR) 'money', *Lȗka* (LF) 'Luke', *lúka* (LR) 'port'). The falling and rising accent are possible on all syllable types, including those containing short vowels and those with an obstruent coda. The Split variety introduces additional tonal crowding in the form of a "double accent", a f0 fall on the post-tonic syllable after the rising tone. The prosodemes of Split Croatian are shown in Figure 1.



This paper examines the timing of f0 peaks and troughs associated with the prosodemes of Split Croatian in order to assess the impact of syllable structure on temporal aspects of their realization. It is hypothesized that there is greater temporal separation between tone targets in syllable types that are relatively unconducive to supporting a tonal contour, e.g. open syllables containing a short vowel and those containing a coda obstruent.

2. METHOD

2.1. Speakers and recordings

13 speakers (1 male) of the Split variety of Croatian were recorded in March 2022 at the Department of Phonetics recording studio at the Faculty of Humanities and Social Sciences, University of Zagreb, under the supervision of a recording technician and the second author. All speakers were born in Split between 1996 and 2002 (with most born between 1999 and 2002), and had arrived in Zagreb for their university studies between 2015 and 2021.
Speakers were recorded using an AKG C414 B-ULS microphone and RME Fireface UFX soundcard. Recordings were saved as mono WAV files using a 44.1 kHz sampling rate and 16 bit bit-depth.

2.2. Stimuli

Speakers read a list of 65 real Croatian words that illustrated the four prosodemes. Words were read in isolation, with no carrier phrase. The words were either stand-alone lexical items, or prosodic words with enclitics. Words contained at least three syllables, and up to six syllables (17 words with three syllables, 24 words with four syllables, 20 words with five syllables and four words with six syllables). The purpose of having longer words was to make sure that there were at least two syllables following the initial stressed syllable, in order to allow for any phrasal tones that may occur in these single-word utterances.
In setting up the wordlist, we endeavoured to find words in which the lexical pitch accent occurred in an <u>Open</u> syllable; in a syllable <u>(Closed by a) Sonorant</u>; and in a syllable <u>(Closed by an) Obstruent</u>. (We also included words that contained a syllabic /r/, and these were treated as a separate category.) We based these decisions regarding syllable structure on the Maximum Onset Principle, and on a word-game called *šatrovački govor*, that allows for the reversal of syllables e.g. *go.vnu* > *vnu.go* 'shit'. Morphological structure was also taken into account if there was a transparent prefix+root structure (as for the prosodic word *ìzvadili su ga* 'they took him out', where /iz/ is a prefix meaning 'out' – this lexical pitch accent was therefore classed as being Closed by an Obstruent, even though the sequence /zv/ is permitted in word-onset position, e.g. *zvono* 'bell').
However, in compiling the wordlist, it became apparent that most words have an Open syllable structure on the stressed syllable, given that Croatian is highly permissive in terms of syllable/word onset clusters. For this reason, syllable type is unbalanced across the wordlist:
11 Long Falling (8 Open, 3 Sonorant)
17 Long Rising (16 Open, 1 Sonorant)
17 Short Falling (12 Open, 3 Obstruent, 2 /r/)
20 Short Rising (11 Open, 4 Obstruent, 3 Sonorant, 2 /r/)
We nevertheless elected to keep all of the words for our study, for purposes of greater ecological validity. The wordlist was randomized and each speaker read four repetitions of the list. (One speaker did not produce the word *klàckali su se* (Short Falling) 'they were see-sawing', due to the addition of this word to the list after the speaker had been recorded.) If we remove the four words with syllabic /r/, which are not considered in the present study, we have a total of <u>3172 possible word tokens for analysis</u> (61 x 4 repetitions x 13 speakers, minus four tokens for the speaker who did not read a particular word).

It should be noted that in the present study, we were not able to control for length of the post-tonic vowel, given the other factors we attempted to control for. For this reason both long and short vowels occur in the syllable following the stressed syllable in our word-list.

2.3. Data analysis

Phonetic transcriptions of the words were imported from a spreadsheet and used for preliminary phonetic segmentation with the Munich AUtomatic Segmentation system (MAUS – [10]) pipeline function G2P->MAUS->PHO2SYL. Manual correction of the phonetic MAUS labelling was conducted using the EMU Speech Database Management System [11, 12], interfaced with the R statistical software package [13]. The Snack signal processor [14] was used for calculating formants. The VoiceSauce software was used to extract voicing measures including Straight f0 [15]. Plots were generated using the *ggplot2* package in R [16].
Dominant peaks and troughs in the f0 traces were identified by progressively smoothing the signal until only two roots (zero crossings) of the first derivative of the smoothed signal were present. This strategy eliminated peaks and troughs caused by noise while selecting the strongest peaks/troughs. Smoothing was implemented using the `smooth.spline()` function in *baseR*, which also allows prediction of the derivative, while the *rootsolve* package [17, 18] was used to identify the roots. In cases where a smoothing iteration produced fewer than two roots, the function returned the (e.g. three or four) roots from the previous smoothing iteration. However, if the first smoothing iteration identified two or fewer roots, it returned the root(s) from this first smooth.
We chose to search for two roots as the target number of roots, since we considered it possible to find both a Maximum and a Minimum in each vowel window, based on our examination of pitch traces in a related

study using the same database. The window chosen for analysis was the vowel (in either the Stress or PostStress syllable), with 10 ms added both before and after the vowel in order to maximise the chances of finding a pitch peak or trough in the f0 trace.

Note that in 17 instances, the function failed to converge even after 20 smoothing iterations (this total of 17 instances includes both the Stress and PostStress vowels).

Examination of data returned by the smoothing function showed that most tokens contained either two or three turning points, and a smaller number of tokens contained only one turning point. Very few tokens contained either zero turning points (i.e. a straight line) or four or more turning points: these tokens were discarded. In addition, of the tokens that contained only one turning point, very few of these contained a pitch Minimum as opposed to a pitch Maximum. For this reason, tokens with only a single Minimum were also discarded. Moreover, examination of the data suggested that where three turning points were found, the last turning point was quite late in the token. Given that we had added 10 ms to the end of the vowel window, we decided to discard the third turning point where three turning points were found, and only keep the first two for that particular token.

Finally, we removed any turning points which had an f0 value of less than 100 Hz or greater than 300 Hz, since these were likely to be a result of tracking errors. All of the above filtering procedures left 3000 tokens where the Stress syllable was analyzed, and 2678 tokens where the PostStress syllable was analyzed.

3. RESULTS

For reasons of space, we only present the results for normalized time, which show the timing of the relevant turning point (Minimum (MN) or Maximum (MX)) in the vowel of the syllable (Stress or PostStress).

Figure 1 shows the timing of the Minima and Maxima across the four prosodemes of Split Croatian. Minima are shown on the top row, and Maxima in the bottom row; Stress is shown in the left column, and PostStress in the right column. It can be seen that the Stress syllable Maxima for the Rising accents (red boxes – bottom left) occur later than the Maxima for the Falling accents (blue boxes – also bottom left). This is what we would expect for a Rising versus Falling pitch accent docked onto the Stress syllable. Conversely, the Maxima in the PostStress syllable (bottom right) occur earlier in the Rising accent than in the Falling accent – this is likewise what we would expect given that the Rising accent has a clear high fall on the PostStress syllable in the Split dialect.

Importantly, the variability in the timing of the Maxima differs between the Rising/Falling accents in relation to the syllable type – the Falling accent has less variability in the timing of the Maximum in the Stress syllable, while the Rising accent has less variability in the Maximum in the PostStress syllable. This reflects the importance of the Stress syllable peak for the Falling accent, but of the PostStress syllable peak for the Rising accent.

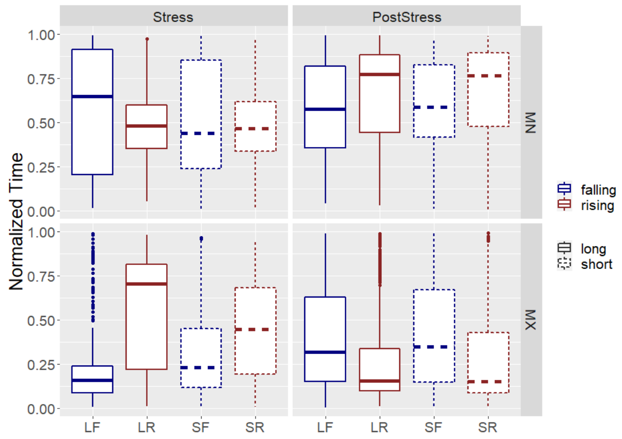


Figure 1: Minima (top row) and Maxima (bottom row) of Stress (left) and PostStress (right) vowels. Rising and Falling pitch accents are shown by colour. Long and short vowels are shown by line type.

Considering now the Minima, we see that the Minimum for the Rising accent tends to occur later in the PostStress syllable, as compared to the Falling accents (top right). This is in line with the high fall that tends to occur on the post-stress syllable of the Rising accent. Turning now to the Minimum in the Stress syllable, we see that although there is less of a clear difference in the means between Rising and Falling accents, there is a clear pattern of less variability in the timing of the Minimum of the Rising accent. This suggests that the timing of a low tone (at around halfway through the Stress vowel duration) is an important identifier of the Rising accent.

To summarize – for the Rising accent, a Minimum is expected at around half-way through the Stress vowel, and a Maximum is expected relatively early in the PostStress vowel. For the Falling accent, a Maximum is expected relatively early in the Stress vowel, while the timing of any Minima or Maxima in the PostStress vowel is less important.

Figure 2 shows the relative timing of Maxima and Minima for the Falling pitch accents according to syllable type. (Note that in all cases, Open~Obstruent~Sonorant refers to the structure of the Stress syllable, not of the PostStress syllable). It can be seen that overall, there are no clear effects of syllable type on the timing of pitch peaks and troughs. It will be recalled that the timing of the Maximum in the Stress syllable shows little variability for the Falling accent, and this is echoed in Figure 2. It may be noted, however, that the timing of the pitch Maximum on the short vowel is more variable than for the long vowel (bottom left, dotted lines for short vowels) – however, there is no effect of syllable type (Open versus Obstruent). Similarly, there is no difference between Open versus Sonorant for the long vowel (solid lines).

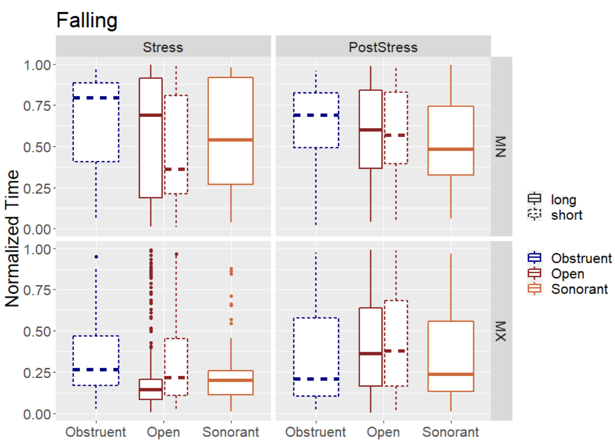


Figure 2: Minima (top row) and Maxima (bottom row) of Stress (left) and PostStress (right) vowels with a Falling pitch accent. Different syllable types are shown by colour. Long and short vowels are shown by line type.

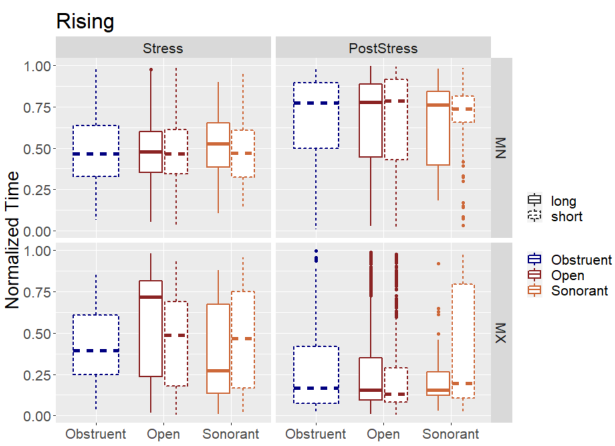


Figure 3: Minima (top row) and Maxima (bottom row) of Stress (left) and PostStress (right) vowels with a Rising pitch accent. Different syllable types are shown by colour. Long and short vowels are shown by line type.

Figure 3 shows the same information as Figure 2, but for Rising pitch accents. It will be recalled that for this pitch accent, the timing of the Minimum in the Stress vowel (top left) and the Maximum in the PostStress vowel (bottom right) are important. However, it is clear that there is neither an effect of vowel duration, nor of syllable type – the Minimum is timed for around halfway through the vowel regardless of any suprasegmental factors, and the Maximum is timed for early in the PostStress vowel, also regardless of any suprasegmental factors.

In summary, the timing of important pitch minima and maxima in this lexical pitch accent language are fixed regardless of syllable type, and perhaps even regardless of vowel duration.

4. DISCUSSION

Analysis of f0 timing and variability indicates a clear division between the rising and falling prosodemes in the importance and location of peaks and troughs. An early f0 peak in the stressed syllable and an early f0 peak in the post-tonic syllable are the most consistent, i.e. least variable, features of the falling and rising prosodemes, respectively. The latter property is a distinguishing characteristic of the Split variety of Croatian and suggests that the contrast between the rising and falling prosodemes may primarily be a contrast between two falling tones, one in the stressed syllable and another in the post-tonic syllable. The timing results thus align with the f0 patterns visually apparent in Figure 1 in that the most salient characteristic of the “rising” prosodeme is the f0 fall in the post-tonic syllable.

The timing of f0 peaks and troughs is not dependent on syllable structure, at least not in a way that suggests sensitivity to the intrinsic ability of different syllable types to support f0 information. The f0 maximum of the falling tone thus does not occur any earlier in a short-voweled syllable that is either open or ends in a coda obstruent. Nor do any peaks or troughs associated with the rising tone reliably shift based on syllable structure.

Overall results suggest that tonal crowding attributed to syllable structure is not alleviated by a temporal shift in f0 targets. Further data would be needed to determine whether tonal crowding is mitigated instead through rescaling of f0 targets. Preliminary results (reflected in Figure 1) indeed suggest that tonal crowding may not be a serious issue if the rising and falling prosodemes both have a single phonological target, a high tone that docks, in the case of the falling prosodeme, on the stressed syllables, and, in the case of the rising tone, on the post-tonic vowel. Investigation of f0 patterns associated with words in different prosodic contexts would be necessary to explore this possibility.

5. ACKNOWLEGEMENTS

We would like to thank our speakers for their time and dedication to language research (Ena Marinković, Ivana Ujević, Jelena Pocedulić, Josipa Papeš, Josipa Teskera, Magdalena Andromak, Mia Mršić, Nuša Vrdoljak, Tea Domić, and four other speakers who did not wish to have their full names published), and Jordan Bićanić for help with the recordings. The ethics protocols for this work were approved by the Office of Research at the University of California, Santa Barbara. Financial support was provided by the School of Humanities and Social Sciences and the Centre for Research on Language Diversity at La Trobe University.

6. REFERENCES

[1] Hyman, L. 1988. Syllable structure constraints on tonal contours. *Linguistique Africaine* 1, 49-60.
[2] Zhang, J. 2002. *The Effects of Duration and Sonority on Contour Tone Distribution: Typological Survey and Formal Analysis.* Routledge.
[3] Gordon, M. 2001. A typology of contour tone restrictions. *Studies in Language* 25, 405-444.
[4] Magner, T. F. "City Dialects in Yugoslavia", *American Contributions to the Eighth International Congress of Slavists* (Zagreb and Ljubljana, September 3-9, 1978). Ohio: Slavica Publishers, Inc., 465–482, 1978.
[5] Kapović, M., *Povijest hrvatske akcentuacije. Fonetika*, Matica hrvatska, Zagreb, 2015.
[6] Godjevac, S., "Transcribing Serbo-Croatian intonation", in S-A. Jun [Ed], *Prosodic Typology*, 146-171, Oxford, 2005.
[7] Browne, E. W. and McCawley, J. D., "Srpskohrvatski akcenat", *Zbornik za filologiju i lingvistiku* VIII:147–151, 1965.
[8] Inkelas, S. and Zec, D. "Serbo-Croatian pitch accent: the interaction of tone, stress, and intonation", *Language*, 64:227-248, 1988.
[9] Rešetar, M., *Die serbokroatische Betonung südwestlicher Mundarten,* Alfred Hölder, K. u K. Hof- und Universitäts-Buchhandler, Wien, 1900.
[10] Kisler, T., Reichel, U. and Schiel, F., "Multilingual processing of speech via web services", *Computer Speech & Language* 45:326–347, 2017.
[11] Winkelmann, R., Harrington, J. and Jänsch K. "EMU-SDMS: Advanced speech database management and analysis in R", *Computer Speech and Language* 45 392-410, 2017.
[12] Winkelmann, Raphael, Jaensch, K., Cassidy, S. and Harrington, J. emuR: Main Package of the EMU Speech Database Management System. R package version 2.0.4, 2019.
[13] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/, 2020.
[14] Sjölander, K. Snack Sound Toolkit, Stockholm: KTH Royal Institute of Technology, Retrieved from http://www.speech.kth.se/snack. 2014.
[15] Vicenik, C., Lin, S., Keating, P. and Shue, Y-L. Online documentation for VoiceSauce. Available at http://www.phonetics.ucla.edu/voicesauce/documentation/index.html, 2020
[16] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York, 2016.
[17] Soetaert K. and P.M.J. Herman (2009). A Practical Guide to Ecological Modelling. Using R as a Simulation Platform. Springer, 372 pp.
[18] Soetaert K. (2009). rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations. R-package version 1.6.