# MODELLING THE PARAMETERS OF FIXED STRESS IN SLAVIC WITH DYNAMIC COMPUTATIONAL NETWORKS

Toby Hudson

University of Oxford
toby.hudson@ling-phil.ox.ac.uk

## ABSTRACT

This paper examines the acoustic correlates of word stress in Czech and Polish, Western Slavic languages with fixed accent position. The objective of the study is to model the data using a pair of variables, calculated by an optimisation algorithm. These variables represent lateral excitation or inhibition influencing prominence through the word in each direction.

A repository of recordings of stand-alone trisyllabic words in both languages is built using publicly-available internet resources. The duration, average intensity and average fundamental frequency for each syllable are measured. The optimisation algorithm is applied to the normalised measurements using a computational network that operates recursively until equilibrium and a pair of weights is obtained.

The network performs well where data are not diffuse, e.g. fitting the intensity data with a mean squared error of 0.04. The trendline for this parameter differs minimally between the two languages due to peak delay in the Czech data.

**Keywords**: Computational networks, modelling, word stress, Czech, Polish

## 1. INTRODUCTION

The possibility of modelling word stress patterns by using dynamic computational networks was introduced by Larson [1]. Here, the relative prominence of syllables is considered to be the result of lateral inhibition and excitation in the neighbouring environment, i.e. each syllable exerts a force on the others, akin to edge phenomena in the visual domain. This approach was heralded for its power and descriptive adequacy since it can accurately model phonological structures such as syllabic sonority and word stress with fewer variables than the number of units under scrutiny.

Larson [1] successfully uses this approach to model all attested general prominence patterns of word stress. Words with lexical/morphological stress or syllabic weight-to-stress are given additional weightings (M, H/L). The core settings are a weighting for initial syllable (I), optionally a weighting for the final syllable (F), a leftwards propagating 'wave' ($\alpha$), and a rightwards propagating 'wave' ($\beta$). Word-prosodic shapes are generated by altering the relative values of these variables within the equation

$$(1) \quad \mathrm{d}_{\mathrm{i}}^{\mathrm{t+1}} = \mathrm{u}_{\mathrm{i}} + \alpha^{*}\mathrm{d}_{\mathrm{i+i}}^{\mathrm{t}} + \beta^{*}\mathrm{d}_{\mathrm{i-1}}^{\mathrm{t}}$$

where d is the derived prominence of the syllable in whichever parameter we are investigating, subscript references its position within the word, superscript means time index, $u$ indicates inherent sonority (weighting $I$ above), $\alpha$ is the coefficient of leftward inhibition, and $\beta$ is the coefficient of rightward inhibition [1: 21]. When the network converges we are left with values for $\alpha$ and $\beta$ which characterise the shape of the trajectory.

The question remains whether or not such an iterative network will fit attested phonetic data with sufficiently low error. The present study is a preliminary attempt to apply equation (1) to data from a pair of languages which are understood to have static stress (i.e. no M or H/L): Czech (CZ) and Polish (PL). Since words in these languages are relatively short, the goal is to model the trajectory of potentially salient parameters with fewer variables than the number of syllables: here, this will be to characterise the durational, intensity and fundamental frequency (f0) contours of trisyllabic words in CZ and PL using only $\alpha$ and $\beta$. This is study is therefore both an exploration of the realisation of stress in these languages and a proof of concept for a modelling technique not commonly used for this purpose which may subsequently be used to model the prosody of longer words and more complex stress scenarios.

Since this study begins by obtaining acoustic readings, no assumption is made about whether and how stress is manifested in the languages in question beyond the selection of CZ and PL as likely candidates for manifesting left-edge and penultimate weighting. Duration, intensity and f0 are considered to be the prime acoustic candidates to represent prominence but this investigation could be expanded to include spectral prominence and finer-grained distinctions such as consonant lengthening.

Western Slavic is understood to have lost the earlier dynamic lexical accent – this was one of its common innovations sometime after the second century AD – and regularised a demarcative stress on the first syllable. Polish later regularised the stress on the penultimate syllable, and recent research has shown Polish to exhibit clear acoustic signalling of word stress in duration and f0 [2], and intensity [3]. However, present-day Czech does not systematically manifest clear acoustic patterns associated with stress. Recent studies (e.g. [4]) have demonstrated a pitch peak delayed to the post-stressed syllable, and [5] reports a shorter duration for the canonically stressed syllable than the second syllable, suggesting that differences between the prosody of CZ and PL may be subtle. Since stress is non-contrastive here, the acoustics of word-level prominence have become weak ([5]) though, as [3] demonstrates, this need not be the case *de facto* for non-contrastive stress.

## 2. METHOD

A dataset was built by obtaining publicly available audio files from the internet (primarily from pronunciation sites *forvo.com* and *dict.cc*). For this study only trisyllabic words were searched for; a trial-and-error process consisted of looking up suitable candidates to see which would bring up clear, noise-free audio recordings. Almost all recordings consisted of single words; a few were short phrases.

| **Czech** | **Polish** |
|---|---|
| nadšený | nieufny |
| kajuta | nadymać |
| kdekoli | mojemu |
| nabidnout | kuśtykać |
| jizdenka | księgosusz |
| hniloba | kryjówka |
| smažený | kozodój |
| obézní | koniuszyc |
| dobrodruh | klarowny |
| hoření | kablówka |
| odjinud | jedynie |
| koleno | jajowód |
| kravata | instrument |
| obrovský | gliwicki |
| domácí | geneza |
| německý | esencja |
| vidlička | dopłacić |
| žvýkačka | cesarski |
| pracovník | bajeczny |
| pomeranč | abstrakcja |

**Table 1**: List of words represented in the set of audio recordings

Only those deemed to be male voices were included here. The set of 20 Polish and 20 Czech tokens, covering a variety of phonotactic possibilities, is listed in Table 1.

Acoustic readings were obtained in *Praat* ([6]) with the use of a script to extract the timing of syllables as well as mean f0 and mean intensity values for each syllable. These were tabulated with very few missing values.

Duration measurements were normalised by expressing the value for each syllable of a word as a percentage proportion of the whole word. Intensity measurements were normalised by expressing the value for each syllable as a percentage proportion of the mean intensity for the word in question. F0 measurements were normalised by conversion of the deviation from the mean value of each word to semitones (multiplication by 12*log2). A very few obvious outliers were removed.

A Python 3 script was written which expresses equation (1) as a 3*3 matrix with only $\alpha$ and $\beta$ as variables. Each normalised list of values (duration, intensity, f0) was read into the script as a comma-separated datafile. The script called the *scipy.optimize* package running the *minimize* routine with the *Nelder-Mead* simplex algorithm. This routine looked for a minimum optimal pair of values for $\alpha$ and $\beta$ in a function that fitted each data set as accurately as possible.

The output for each of the six datasets is given as the optimal values for $\alpha$ and $\beta$, and mean squared error (MSE) along with filled contour plots to represent the degree of optimality of each pair of variables to the data at hand (the darkest coloured contour represents the solution space within which the network performs well). The dashed trajectory overlaid on the input trajectories shows in each case how neatly the function matches the input.

There arises the question of the value to which the invariant 'positional activation' $I$ should be set. Larson [1: 94] shows that this must be greater than zero to produce canonical patterns of initial stress. The model was run with $I$ set to values from 0.1 to 1.0 by increments of 0.1; the best fit from these (declared for each output below) was selected for each dataset.

## 3. RESULTS

Figures 1 and 2 represent the optimization for the CZ and PL *duration* measurements respectively. The data for durational measurements were not tightly clustered, such that no obvious outliers could be removed to improve the optimization which suffered from these rather messy sets of data.

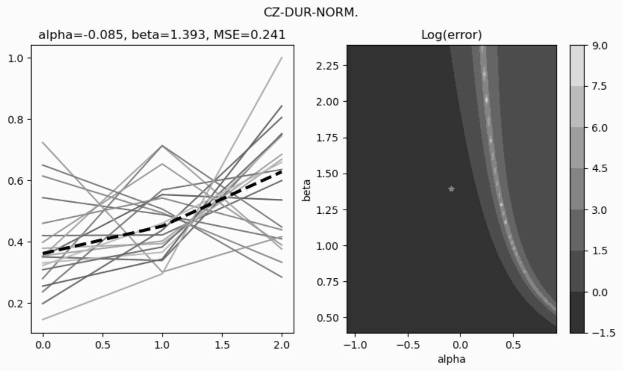

**Figure 1**: Optimization for syllable duration in Czech; *I*=0.4

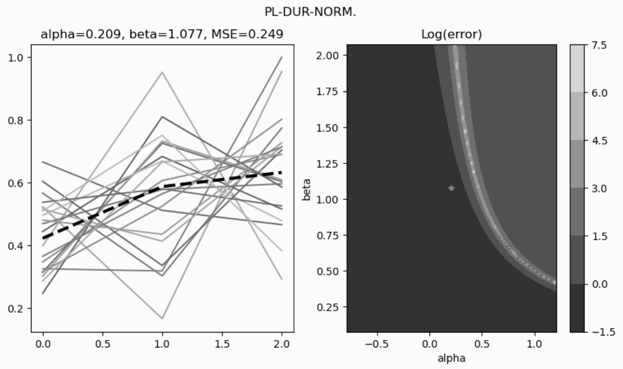

**Figure 2**: Optimization for syllable duration in Polish; *I*=0.3

A second syllable peak is more apparent for the *intensity* readings, and with an MSE of approximately 0.04 for both CZ and PL the model fits the data much better for this parameter (Figures 3 and 4).

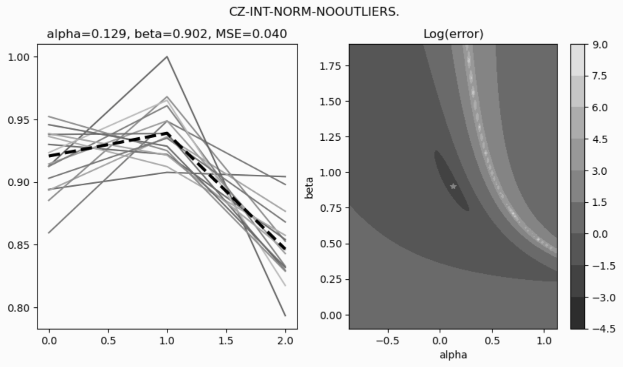

**Figure 3**: Optimization for syllable intensity in Czech; *I*=0.8

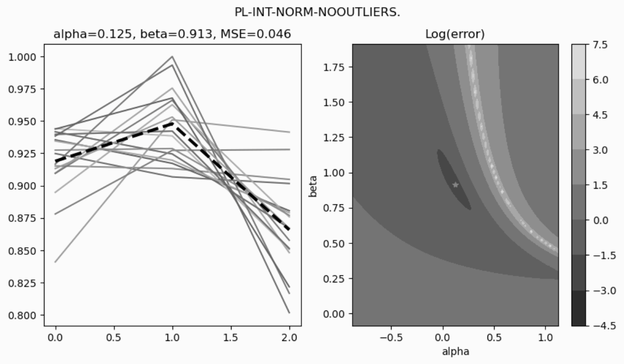

**Figure 4**: Optimization for syllable intensity in Polish; *I*=0.8

Optimization was relatively successful for *f0* with MSE at approximately 0.1 for both languages with a different setting for *I* for CZ and PL (Figures 5 and 6). The reported pitch delay may be in evidence for CZ but there is apparently a split in the data between trajectories with an f0 rise from syllable 1 to syllable 2 and those with a fall in f0 in this position. This same split is evident in the PL data. Optimization would be greatly improved without this split.

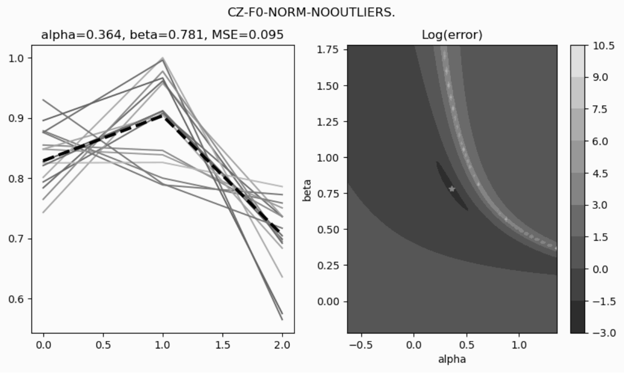

**Figure 5**: Optimization for syllable f0 in Czech; *I*=0.5

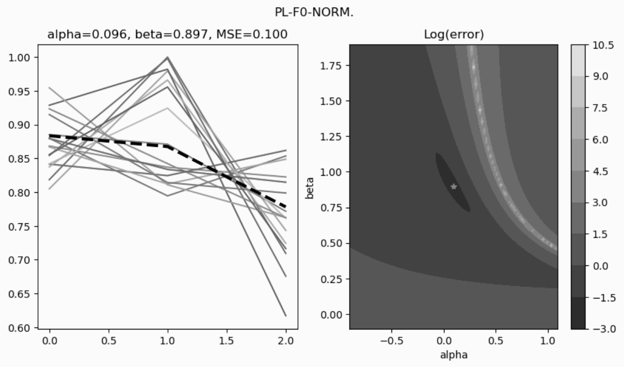

**Figure 6**: Optimization for syllable f0 in Polish; *I*=0.8

Fig. 7 plots the goodness of *I* at the intervals tested. If it were necessary to set this to a single value for all parameters then it would be possible to adopt a value around the centre of the range with the least mean error for all data sets being modelled (from left to right the red arrows indicate mean, median and mode activation weight).

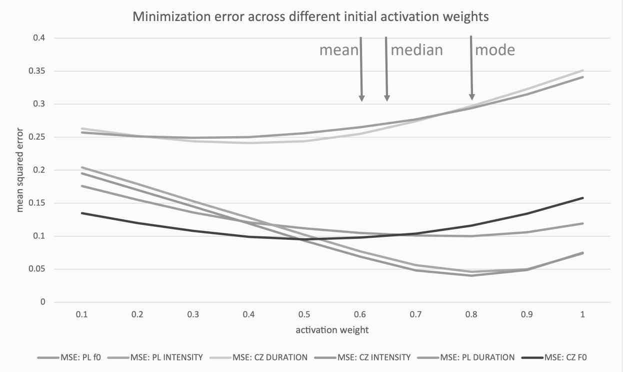

**Figure 7**: Minimization error (mean squared error) as a result of different activation weighting (*I*)

Larson [1:78] shows how different combinations of A and B generate the stress patterns of the world's language. Negative $\alpha$ in combination with negative $\beta$, for example, produces a rhythmic stress pattern, but positive $\alpha$ with positive $\beta$ results in cumulative stress. When $\alpha$ is a higher positive value and $\beta$ is a lower positive value, left edge cumulative stress is generated. With the exception of the more disparate data set for Czech duration (for which $\beta$ had a negative value), all other data sets here do indeed conform to this pattern, as can be seen in Fig. 8 which plots the convergence values for the six optimizations.

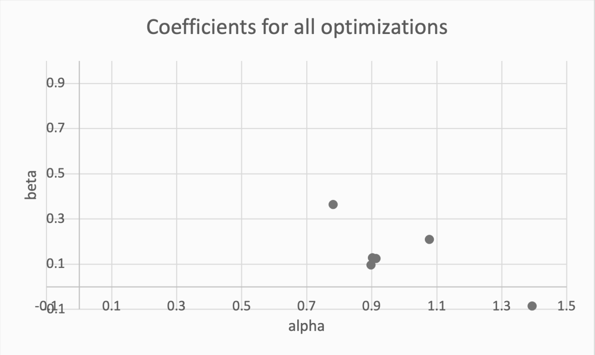

**Figure 8**: Convergence values for all six optimizations

## 4. DISCUSSION AND CONCLUSIONS

On the basis of these findings, average syllable *intensity* is most neatly modelled, and the data themselves are most consistent in this parameter. With optimization at the same initial activation value and a similar goodness of fit for CZ and PL it may be concluded that CZ and PL are modelled equally well for intensity, and that the model is successful in these two data sets. Additionally it would seem that there is little to distinguish the two languages in this parameter, i.e. they both exhibit a second syllable peak – the canonical position for PL, but representing peak delay in CZ.

With an MSE greater than 0.2 for both outputs we cannot deem the model to be powerful for *duration*. It is possible that an improvement would be seen with longer words if the data are split by those examples with a word-final lengthening effect and those without. Visual inspection of the trajectories does not lend strong confirmation to the notion that stress is cued durationally in either language, though the trajectories of several tokens have a durational peak on the second syllable which suggests an inconsistent durational prominence in the expected part of the word for Polish and in delayed position for Czech.

Undeniably the words with long vowels (é, á, í, ý) introduce a durational confound here as might be expected in a language with phonemic vowel length and quantity-insensitive stress. It is also possible that the variability of onset and rime consonant clusters has confounded patterns which would be apparent in a much larger investigation or where only words of matching syllabic complexity are compared.

Similarly to the intensity readings, the *f0* data are split into trajectories which have an f0 peak on the second syllable and others with an f0 trough in the same position. It is possible that intonation has been a confound here, for example if recordings were made with a 'listing' intonation. Alternatively it may be that f0 plays an inconsistent role in cueing stress in both languages.

The relative strength of intensity for CZ and PL here as cue to word stress tallies with the perspective that word stress is allied to *loudness*, a percept which relates to vocal effort in addition to e.g. jaw aperture and vowel timbre ([7]). The similarity in trend for intensity in CZ to that of PL is suggestive of (typologically rare) peninitial stress.

Where the data (once normalised and bereft of outliers) follow the same general trend the network has successfully modelled the trend with a fixed activation value and two variables. The practical implication is that the activation value in combination with $\alpha$ and $\beta$ as cited for each output may be employed in synthesis to generate prosodic trajectories that accord with attested data. A theoretical implication of the above may be that the success of the model in representing intensity contours for Czech and Polish words that fit the data well vindicates its reality, i.e. it may indeed represent oscillatory ('wave') processes of excitation and inhibition in speech production and perception in which successive units in a string are quantifiably related to one another ([8]). However, the implications of these promising preliminary findings would be strengthened by extending the application of the model to longer words, which would also serve to test the applicability of the method to modelling secondary stresses.

## 5. REFERENCES

[1] Larson, G. 1992. Dynamic computational networks and the representation of phonological information. Doctoral thesis.

[2] Oliver, D., Grice, M. 2003. Phonetics and Phonology of lexical stress in Polish verbs. *Proc. 15th ICPhS* Barcelona, 2027–2030.

[3] Malisz, Z., Żygis, M. 2018. Lexical stress in Polish: evidence from focus and phrase-position differentiated production data. *9th International Conference on Speech Prosody* Poznań, Poland, 1008–1012.

[4] Palková, Z., Volín, J. 2003. The role of F0 contours in determining foot boundaries in Czech. *Proc. 15th ICPhS* Barcelona, 1783–1786.

[5] Skarnitzl, R, Eriksson, A. 2017. The acoustics of word stress in Czech as a function of speaking style. *Proc. Interspeech*, 3221–5.

[6] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.2.21, retrieved 1 October 2022 from http://www.praat.org/.

[7] Kochanski, G., Grabe, E., Coleman, J., Rosner, B. 2005. Loudness predicts prominence: fundamental frequency lends little. *J. Acoust. Soc. Am*. 118, 1038–1054.

[8] Goldsmith, J. 1994. A Dynamic Computational Theory of Accent Systems. In: Cole, J., Kisseberth, C. (eds), *Perspectives in Phonology*. Stanford: Center for the Study of Language and Information, 1–28.