9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Acoustic profiles for prosodic headedness and constituency

Uwe D. Reichel[1], *Katalin Mády*[2], *Štefan Beňuš*[3]

[1]Institute of Phonetics and Speech Processing, University of Munich, Germany
[2]Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary
[3]Constantine the Philosopher University, Nitra & II SAS Bratislava, Slovakia

reichelu@phonetik.uni-muenchen.de, mady@nytud.hu, sbenus@ukf.sk

Abstract

We examined American English, French, German, Hungarian and Slovak data with respect to two dimensions of prosodic typology, namely headedness and the existence or absence of accentual phrases. Based on a computational prosodic stylization we identified several acoustic features distinguishing the given languages in those dimensions. The relevant features were integrated to acoustic profiles characterizing the prosody of languages with regard to the selected typology aspects.

Index Terms: prosodic typology, intonation stylization, American English, French, German, Hungarian, Slovak, headedness, accentual phrase, profile

1. Introduction

1.1. Prosodic typology

Languages differ prosodically amongst others in terms of their prominence type [1], the tone inventory and the type of prosodic constituents [2], headedness [3], and rhythm [4]. The classification of languages with respect to these dimensions can roughly be subdivided into expert- and data-driven. Among the expert-driven approaches [2, 1] developed intonation typologies based on numerous single-language studies in the framework of autosegmental-metrical (AM) phonology [5]. [2] reports that a major difficulty of this approach consists in the comparison of languages across different tonal inventories. [3] circumvent this problem of post-hoc annotation unification using their language-independent INTSINT model. In both accounts the classification of a language's intonation is based on a categorical data representation.

Computational approaches in contrast additionally yield continuous parameters allowing for capturing more phonetic detail. Opposed to the expert-driven approaches they do not require time-consuming manual annotations. However, except of numerous studies on measuring rhythm (e.g. [4, 6, 7]), computational typology accounts are still very rare. [8] propose a Wavelet decomposition of fundamental frequency (f0), duration, and energy contours. Their typology is then derived in a bottom-up way by clustering the Wavelet coefficients.

1.2. Goals of this study

We propose a computational data-driven account to automatic prosodic typology that is based on a superpositional intonation stylization [9, 10].[1] We will show, that this account yields phonetically interpretable acoustic features from which acoustic typology profiles can be inferred.

[1]The stylization code (Python 3) is open source and available here: *https://www.github.com/reichelu/copasul*

For the present study we restrict the examinations on the prosodic dimensions *constituency* and *headedness*, for which Table 1 gives an overview for the examined languages.

Constituency. Accent groups consist of an accented and neighboring unaccented syllables. Languages can be subdivided with respect to whether or not those accent groups form a prosodic phrase on its own, namely the *accentual phrase* (AP). This phrase can be defined by a language-dependent stable f0 pattern and by boundary marking [11, 12, 2].

Headedness. Languages furthermore differ in their tendency to place relevant prosodic events as word stress or pitch accents rather at the beginning or the end of prosodic constituents [3]. This behavior is called *left-* and *right-headed*, respectively.

Table 1: *Prosodic classification of the languages under examination following [3, 2]. English and German are considered as left-headed up to the accent group level. For Slovak headedness is not as clear-cut [13].*

Language	Headedness	AP
English	left	no
French	right	yes
German	left	no
Hungarian	left	yes
Slovak	–	yes

2. Data

For the Hungarian, Slovak, and American English data we randomly selected 150 intonational phrases (IP), respectively, containing about 440 manually segmented accent groups from corpora of collaborative dialogues [14, 15, 16]. The French data was obtained from the Rhapsodie corpus [17, 18] containing dialog and monologue data that is segmented and annotated phonetically, prosodically, and syntactically. From this corpus we extracted a random sample of the same size as the other language's data from the spontaneous-speech dialog part. In the French annotation a different terminology is used for the prosodic units in question. We thus defined according to the specifications given in [18] the constituent type "*rhythmic group*" as accentual phrases and the next-higher level type "*intonational packages*" as intonational phrases. This equivalent treatment of rhythmic groups and accentual phrases is compliant to the AP terminology of [19] (*"rhythmic unit"*, and [20] (*groupe rhythmique*). For the German data a random prosodically segmented sample of comparable size from the Verbmobil 1 corpus [21] was taken.

10.21437/SpeechProsody.2018-142

3. Prosodic parameterization and profile generation

The parameterization of the f0 contour was carried out in the contour-based superpositional CoPaSul stylization framework [9, 10]. In this framework f0 is decomposed into a global component corresponding to the intonation phrase, and local components corresponding to accent groups. From this stylization feature sets were extracted for the intonation phrase and the accent group level as well as for accent group boundaries, which will be described in the following sections.

3.1. Preprocessing

F0 was extracted by autocorrelation (Praat 6.0 [22], sample rate 100 Hz; allowed f0 range from 50 to 400 Hz; default settings). Voiceless utterance parts and f0 outliers were bridged by linear interpolation. Outliers were defined separately for each file as deviating more than twice the standard deviation from the f0 mean. The contour was then smoothed by Savitzky Golay filtering [23] using third order polynomials in 5 sample windows and transformed to semitones relative to a base value b as follows: $\mathrm{F0}_{st} = 12 \cdot log_2(\frac{\mathrm{F0}_{Hz}}{b})$. b was set to the f0 median below the 5th percentile of an utterance and served to normalize f0 with respect to its overall level.

3.2. Intonation phrase features IP

As shown in Figures 1 and 2 a base-, a mid- and a topline are fitted through the f0 contour in an IP by means of linear regression. Time was normalized from 0 to 1. Further details are described in [24, 10]. Following [25], who divide register into *level* and *range*, the base-, mid-, and topline represent register level aspects, and the pointwise distance between base- and topline represents register range aspects. More precisely, in our approach range is parameterized by means of a linear regression through these pointwise distances. A negative slope of the range regression line thus indicates converging base- and topline, whereas a positive slope indicates line divergence. From these four regression lines (base-, mid-, topline and range) the parameters intercept and slope were considered for further analyses (features *bl|ml|tl|rng_c0|c1*, cf. Table 2).

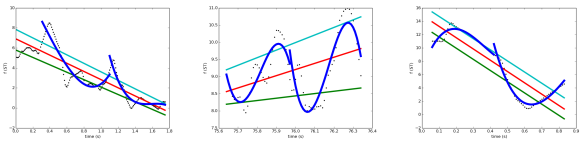

Figure 1: *Superpositional intonation stylization of Hungarian (left), French (mid), and German (right), each for one IP containing two AGs. IP register is captured by a base-, mid-, and topline. AG shape is represented by third-order polynomials. The same IPs are displayed in Figure 2.*

3.3. Accent group features AG

The f0 contour within an accent group (AG) is represented by it's local register, it's shape, as well as by its deviation from the underlying IP. For capturing its shape we fitted 3rd order polynomials to the time-normalized contour (details in [10]), of which we derived the coefficients as shape features for further examination (features *c0–3*, cf. Table 3). The local register in AG segments was derived analogously to the IP level as described in section 3.2. Again the parameters intercept and slope of the four regression lines were considered for further analyses (features *bl|ml|tl|rng_c0|c1*). The deviation of the AG from the IP contour was measured by the root mean squared deviation of the AG-level regression line with the corresponding IP-level regression line stretch (features *bl|ml|tl|rng_rms*). Local AG initial and final register deviations from the IP are captured by calculating the difference for the AG line onset and for the line offset to the corresponding points on the respective IP line (features *bl|ml|tl|rng_d_init|fin*).

Table 2: *Intonation phrase features (set IP) relevant for the typology dimensions headedness (head) and/or the presence or absence of accentual phrases (AP; $p < 0.1$).*

Feature	Description	Relevance
bl_c0	baseline intercept	head, AP
bl_c1	baseline slope	head, AP
ml_c0	midline intercept	head, AP
ml_c1	midline slope	head
rng_c0	range intercept	–
rng_c1	range slope	–
tl_c0	topline intercept	head
tl_c1	topline slope	head

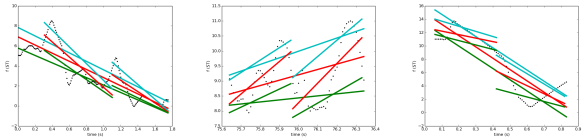

Figure 2: *Superpositional intonation stylization of Hungarian (left), French (mid), and German (right), each for one IP containing two AGs. IP and AG registers are captured by base-, mid- and toplines, respectively. The deviation between AG- and IP-related lines quantifies how much and in which direction an AG sticks out from the underlying IP. The reset between offset and onset of subsequent AG register lines represents the amount of discontinuity at AG boundaries. The same IPs are displayed in Figure 1.*

3.4. Accent group boundary features BND

At AG boundaries we measured the discontinuity of each of the 4 register regression lines this time fitted to the two one second segments adjacent to the boundary as in [24]. As illustrated in Figure 3 discontinuity is expressed as the reset of each regression line, i.e. the difference between the f0 onset of the line in an AG and the f0 offset of this line in the preceding AG (features *bl|ml|tl|rng_r*; cf. Table 4). The deviation of the line pair from a common trend was further expressed by fitting a line of the same type (e.g. a baseline for a AG-related baseline pair) through both adjacent segments and the measuring the RMS between the single segment fits and the joint segment fit (features *bl|ml|tl|rng_rms*). These reset and RMS discontinuity features turned out to be correlated with perceived prosodic boundary strength in [24].

3.5. Directedness

The prosodic dimensions *headedness* and *constituency* differ with respect to the role played by the algebraic sign of the acoustic variables. While for headedness determination it is

Table 3: *Accent group features (set AG) relevant for the typology dimensions headedness (head) and/or the presence or absence of accentual phrases (AP; $p < 0.1$).*

Feature	Description	Relevance
bl_c0	baseline intercept	–
bl_c1	baseline slope	head
bl_d_fin	baseline end diff	head
bl_d_init	baseline init diff	–
c0	0th poly coef	head
c1	1st poly coef	head
c2	2nd poly coef	head, AP
c3	3rd poly coef	–
ml_c0	midline intercept	–
ml_c1	midline slope	head
ml_d_fin	midline end diff	head
ml_d_init	midline init diff	head
rng_c0	range intercept	–
rng_c1	range slope	head
rng_d_fin	range end diff	head
rng_d_init	range init diff	head
tl_c0	topline intercept	–
tl_c1	topline slope	head
tl_d_fin	topline end diff	head
tl_d_init	topline init diff	head

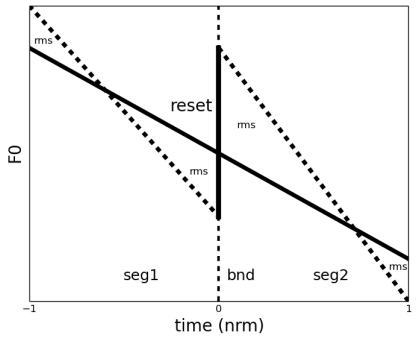

Figure 3: *Boundary stylization in terms of discontinuity of register regression lines fitted in segments seg1 and seg2 (dotted): Line reset (thick solid vertical line) and RMS deviation from a common trend (represented by a regression over both seg1 and seg2; solid diagonal line).*

Table 4: *Accent group boundary features (set BND) relevant for the typology dimensions headedness (head) and/or the presence or absence of accentual phrases (AP; $p < 0.1$).*

Feature	Description	Relevance
bl_r	baseline reset	–
bl_rms	baseline RMS	head
ml_r	midline reset	AP
ml_rms	midline RMS	head, AP
rng_r	range reset	–
rng_rms	range RMS	head
tl_r	topline reset	AP
tl_rms	topline RMS	head

crucial to know whether an f0 trend or discontinuity is positive or negative, for constituency only the absolute values are to be considered. To give an example: By definition APs are hypothesized to be edge-marked by high absolute reset values, while the direction of the reset is determined by headedness only – positive for left-headed, and negative for right-headed languages. Therefore, for constituency we only looked at absolute values for trend and discontinuity variables such as slope and reset. The respective y-axis labels are marked by absolute value bars in the profile plots in Figures 7 and 8 (note that in these Figures values might still be negative due to subsequent z-scoring).

3.6. Profile generation

We applied for all features and each of the two dimensions a linear mixed model with random intercept with the respective feature as the dependent variable, the typology dimension as the fixed effect, and the speaker as the random effect. For headedness the not yet classified Slovak data (cf. Table 1) was exluded from the analyses. Subsequently, for each dimension and each feature set an acoustic profile was generated based on those features that showed a significant difference in the respective dimension. The significance level was set to 0.1 to allow also weakly significant cases to contribute to the profiles. The profiles are shown in Figures 4 to 8. The names of the relevant variables are indicated on the y-axis. Their median values after z-scoring are plotted on the x-axis for each dimension level.

4. Results

26 out of 36 features showed an at least weakly significant difference in at least one of the examined dimensions, which is documented in the final columns of Tables 2, 3, and 4. Overall, the headedness dichotomy was better captured by the examined features than the AP dichotomy. For the former 24 features are indicative, for the latter only 7. Figures 4, 5, and 6 show headedness profiles for the feature subsets IP, AG, and BND respectively, containing those features by which language prosodies significantly differ in the headedness dimension. Figures 7 and 8 show constituency profiles for the feature subsets IP, AG, and BND, containing those features by which language prosodies significantly differ in the AP constituent dimension.

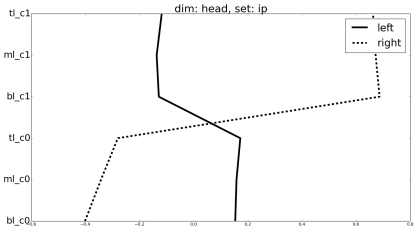

Figure 4: *Headedness profiles for left- (solid) and right- (dashed) headed languages for the IP feature set, i.e. for IP register level and range parameters.*

5. Discussion and conclusions

5.1. Phonetic interpretation

The results indicate that the studied prosodic dimensions are captured by a large amount of the extracted features. Most of these findings are phonetically well interpretable as will be exemplified in the subsequent paragraphs.

Headedness. Prototypical examples for the headedness influence on the shapes of IPs and AGs, on how AGs contrast with IPs, and on AG boundaries are given in the stylization plots in Figures 1 and 2 for left-headed Hungarian (left) and

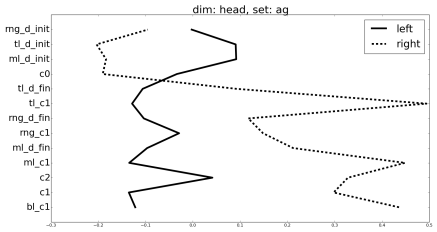

Figure 5: *Headedness profiles for left- (solid) and right- (dashed) headed languages for the AG feature set, i.e. for f0 shape and register parameters in accent groups and for AG deviation from the underlying IP.*

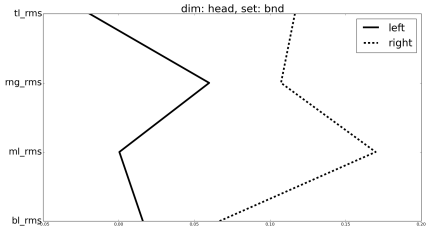

Figure 6: *Headedness profiles for left- (solid) and right- (dashed) headed languages for the BND feature set, i.e. for f0 discontinuity features at AG boundaries.*

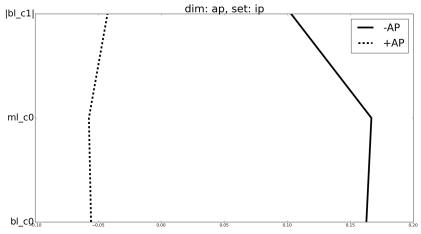

Figure 7: *Prosodic constituency profiles for languages that contain accentual phrases (dashed, +AP) or not (solid, -AP) for the IP feature set, i.e. for IP register level and range parameters. For the BND feature set, i.e. for f0 discontinuity features at AG boundaries.*

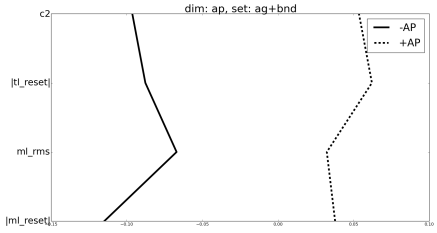

Figure 8: *Prosodic constituency profiles for languages that contain accentual phrases (dashed, +AP) or not (solid, -AP) for the AG and BND feature sets, i.e. for f0 shape and register parameters in accent groups and for AG deviation from the underlying IP as well as for f0 discontinuity features at AG boundaries.*

right-headed French (mid). As reflected in these plots as well as in the IP feature profile in Figure 4, left-headed languages are characterized by a higher register line intercepts (**_c0* features) and negative IP register line slopes (**_c1* features), while for right-headed languages lower intercepts and positive slopes are measured. Left-headed languages thus show IP-initially a high f0 register, whereas right-headed languages show the opposite tendency. The same register level and additionally register range trends are observed on the AG level (cf. Figure 5). Local register (**_c1*) as well as local f0 shape (*c1*) has a falling trend in left-headed, and a rising-trend in right-headed languages. Furthermore, left-headed AGs deviate from the underlying IP more in the beginning (**_init*), and right headed ones at the end (**_fin*). AG boundaries are more strongly marked in right-headed languages (Figure 6, greater mean values for features **_rms*) maybe as a consequence of the opposite trends caused by AG-internal inclination and overall declination, the former raising AG final f0 values, the latter lowering initial f0 values of the subsequent AG. In summary, all mentioned features reflect the tendency of left- and right-headed languages to place relevant prosodic events phrase-initially or finally, respectively. Even if to the current state we examined only one right-headed language, the findings well confirm expert expectations and therefore most likely do not reflect idiosyncratic but topological characteristics.

Constituency. As for headedness Figures 1 and 2 give prototypical examples for the influence of AP constituents on the shapes of IPs and AGs, on how AGs deviate from IPs, and on AG boundaries. In these stylization plots languages with APs (left: Hungarian and mid: French) are compared with the non-AP language German (right). In accordance with these plots the IP profiles in Figure 7 reveal a stronger baseline declination tendency in languages without APs (higher absolute register level slope *bl_c1*). This and the higher intercepts (**_c0*) can be taken as indication that in AP languages the IP is a less salient unit in defining general f0 baseline tendencies compared to non-AP languages. As inferable from the AG feature profile in Figure 8, f0 shapes in AGs in AP languages tend to be convex (falling-rising, which is reflected by more positive *c2* coefficient values). Taking into account studies on initial and final AP boundary signals [26, 12] we conclude that this shape results from AP edge marking. This edge marking is further reflected by the higher absolute reset values (**_reset*; cf. Figure 8) for languages containing APs.

5.2. Comparison of typology accounts

The typology approach proposed in this study combines the advantages of computational and of expert-driven accounts. As with other computational accounts the prosodic representation can be derived automatically and is language-independent, so that it can with little effort be applied to new and understudied languages. Furthermore, the parametric representation allows for a more fine-grained analysis of acoustic typology properties. As with expert-driven accounts, the representation turned out to be phonetically interpretable with respect to pre-defined prosodic dimensions. The present account enabled us to derive interpretable prosodic language profiles, that can provide data-driven evidence for typology research.

6. Acknowledgments

The work of the first author is financed by a grant of the Alexander von Humboldt Foundation.

7. References

[1] S.-A. Jun, "Prosodic typology: By prominence type, word prosody, and macro -rhythm," in *Prosodic Typology II: The phonology of intonation and phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2014, pp. 520–539.

[2] ——, "Prosodic typology," in *Prosodic Typology: The phonology of intonation and phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, pp. 430–458.

[3] A. Di Cristo, "Intonation in French," in *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press, 1999, pp. 195–218.

[4] E. Grabe and E. Low, "Durational variability in speech and the rhythm class hypothesis," in *Papers in Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds. Berlin: Mouton de Gruyter, 2002, pp. 515–546.

[5] J. Pierrehumbert, "The phonology and phonetics of Englisch intonation," Ph.D. dissertation, MIT, Cambridge, MA, 1980.

[6] V. Dellwo, "Rhythm and speech rate: A variation coefficient for deltaC," in *Language and language-processing*, P. Karnowski and I. Szigeti, Eds. Frankfurt/Main: Peter Lang, 2006, pp. 231–241.

[7] L. Wiget, L. White, B. Schuppler, I. Grenon, O. Rauch, and S. Mattys, "How stable are acoustic metrics of contrastive speech rhythm?" *J Acoust Soc Am*, vol. 127, no. 3, pp. 1559–1569, 2010.

[8] J. Šimko, A. Suni, K. Hiovain, and M. Vainio, "Comparing languages using hierarchical prosodic analysis," in *Proc. Interspeech*, Stockholm, Sweden, 2017, pp. 1213–1217.

[9] U. Reichel, "Linking bottom-up intonation stylization to discourse structure," *Computer, Speech, and Language*, vol. 28, pp. 1340–1365, 2014.

[10] ——, *CoPaSul Manual – Contour-based parametric and superpositional intonation stylization*, RIL, MTA, Budapest, Hungary, 2017, https://arxiv.org/abs/1612.04765.

[11] S.-A. Jun and C. Fougeron, "The accentual phrase and the prosodic structure of French," in *Proc. ICPhS*, Stockholm, Sweden, 1995, pp. 722–725.

[12] ——, "Realizations of accentual phrase in French," in *Probus*, 2002, vol. 14, pp. 147–172.

[13] U. Reichel, K. Mády, and v. Beňuš, "Parameterization of prosodic headedness," in *Proc. Interspeech*, Dresden, Germany, 2015, p. paper 929.

[14] Š. Beňuš, U. Reichel, and K. Mády, "Modelling accentual phrase intonation in Slovak and Hungarian," in *Complex Visibles Out There*. Olomouc, Czech Republic: Palacký University, 2014, vol. 4, pp. 677–689.

[15] K. Mády, U. Reichel, and Š. Beňuš, "Accentual phrases in Slovak and Hungarian," in *Proc. Speech Prosody*, Dublin, Ireland, 2014, pp. 752–756.

[16] G. Agustín and J. Hirschberg, "Turn-taking cues in task-oriented dialogue," *Comp. Speech and Language*, vol. 25, no. 3, pp. 601–634, 2011.

[17] "ANR Rhapsodie 07 Corp-030-01, Corpus prosodique de référence du franais parlé," http:www.projet-rhapsodie.fr, June 24th 2014, version 1.0.

[18] A. Lacheret, S. Kahane, J. Beliao, A. Dister, K. Gerdes, J.-P. Goldman, N. Obin, P. Pietrandrea, and A. Tchobanov, "Rhapsodie: A prosodic-syntactic treebank for spoken French," in *Proc. LREC*, Reykjavik, Iceland, 2014, pp. paper hal–00 968 959.

[19] A. Di Cristo and D. Hirst, "Rythme syllabique, rythme mélodique et représentation hierarchique de la prosodie du français," in *Travaux de l'Institut de Phonétique d'Aix*, 1993, p. 924.

[20] E. Delais-Roussarie, "Pour une approche parallèle de la structure prosodique," Ph.D. dissertation, Universit Toulouse le Mirail, 1995.

[21] S. Burger, K. Weilhammer, F. Schiel, and H. Tillmann, "Verbmobil Data Collection and Annotation," in *Verbmobil: Foundations of Speech-to-Speech Translation*, W. Wahlster, Ed. Berlin/Heidelberg: Springer, 2000, pp. 537–549.

[22] P. Boersma and D. Weenink, "PRAAT, a system for doing phonetics by computer," Institute of Phonetic Sciences of the University of Amsterdam, Tech. Rep., 1999, 132–182.

[23] A. Savitzky and M. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[24] U. Reichel and K. Mády, "Comparing parameterizations of pitch register and its discontinuities at prosodic boundaries for Hungarian," in *Proc. Interspeech 2014*, Singapore, 2014, pp. 111–115.

[25] T. Rietveld and P. Vermillion, "Cues for Perceived Pitch Register," *Phonetica*, vol. 60, pp. 261–272, 2003.

[26] J. Vaissière, "Rhythm, accentuation and final lengthening in French," in *Music, Language, Speech and Brain*, J. Sundberg, L. Nord, and R. Carlson, Eds. London: Macmillan Press, 1991, pp. 108–120.