SPEAKING RATE VARIATION AND THE INTERACTION BETWEEN HIERARCHICAL RHYTHMIC LEVELS

Agnieszka Wagner, Jolanta Bachan

Institute of Linguistics, Adam Mickiewicz University in Poznan, Poland
wagner@amu.edu.pl

Abstract: This paper presents linear models built for the purpose of investigating tempo-dependent interaction between two components of Polish speech rhythm, namely *Rhythmic Prominence Intervals* (RPI) and syllables, within the framework of *coupled oscillator model* ([1], [2], [3]). The relative coupling strength parameter (*r*) calculated from regression coefficients shows increasing strength of the lower-level syllabic oscillator with increasing tempo and, conversely, greater dominance of the RPI oscillator at slower speaking rates. These findings indicate the *coexistence* of different timing strategies or patterns within one language and add to the evidence against the traditional, categorical division of languages into syllable- and stress-timed (e.g., [4]). Altogether, the results of our models indicate a *strong linear relationship* between syllable count in RPI and its duration across different speaking rates (very slow – very fast) and reveal some interesting properties of the rhythmic/timing structure of Polish, e.g. greater flexibility of *tempo deceleration* comparing to other languages. The models account for between 85% and 94% of the variability in RPI duration (based on adj. R^2) when *speaker* is added as a random factor, or when the *laboratory measured speaking rate* (in syll./sec.) is used instead of *intended speaking rate* (tempo class).

1 Introduction

This study is part of a large project on the rhythmic structure of utterances in Polish[1] and investigates selected aspects of Polish speech rhythm, namely the relationship between two components of the hierarchical rhythmic structure, i.e. the lower-level *syllabic* component and the higher-level *Rhythmic Prominence Interval* component, across a range of speaking rates. The analysis is carried out using the *coupled oscillator model* of speech rhythm ([1], [2], [3]). The model assumes that complex rhythms such as speech rhythm, but also various biological rhythms, can be modelled as a result of an *interaction* between *component oscillators* (or "subrhythms") which control timing at different levels of a complex hierarchical system (e.g., syllable, foot or interstress interval level, etc.). In isolation, these components may display simple periodic behavior (they represent simple oscillations), but in interaction the oscillators influence each other: The direction, i.e. which oscillator dominates and which becomes entrained, and the strength of this influence is expressed in terms of *relative coupling strength* denoted with *r*. In the model by O'Dell and Nieminen ([1], [2]) *r* is measured as the ratio of linear regression coefficients (*r* = intercept/slope), based on the results presented by Eriksson [5], who found strong linear relationship between interstress interval (ISI) duration and the number of syllables (n) it contained (ISI = a + bn). It was shown in [5] that what distinguishes the purported rhythm classes of syllable- and stress-timing is the value of the intercept which is higher in stress- than in syllable-timed languages. ([1], [2]) proposed to interpret a tendency towards stress-timing as the effect of the dominance of the foot or ISI oscillator over the syllabic one ($r > 1$), resulting in a greater overall timing variability, and, conversely, greater strength of the syllabic oscillator over the foot/ISI oscillator in case of syllable-timing ($r \leq 1$). Coupled oscillator models provide a very good solution to analysis of different timing strategies within a

[1] This research has been carried out in the scope of the project "Rhythmic structure of utterances in the Polish language: A corpus analysis" supported by National Science Centre (NCN) grant no. 2013/11/D/HS2/04486.

single language and can also be useful for tackling typological issues concerning language rhythm.

2 Methodology

2.1 Speech material

The current study is based on a selected material from the *Polish Rhythmic Database* [6], i.e. readings of the story "The north wind and the sun" at five *intended speaking rates* (very slow, slow, normal, fast and very fast) by twenty Polish native speakers (10 females and 10 males). The *Polish Rhythmic Database* consists of 3 subcorpora: *L1 Polish, L2 Polish* (recordings of 12 non-native speakers of Polish with L1 German, Spanish and Korean), and *other L1s* representing languages traditionally associated with different rhythm types: stress-, syllable- and mora-timing (German, Spanish and Korean respectively). Apart from the recordings of the story, the *L1 Polish subcorpus* consists of: a) read sentences of a varying phonotactic structure, b) mini-dialogues, c) excerpts of poems representing different poetic meters, and d) spontaneous monologues. The whole database includes **5 h 30 min.** of speech and was collected for the purpose of studying a wide range of topics related to speech timing, rhythm and prosody in general. It is accompanied with tools and methods developed specifically for the needs of such analyses, e.g., plugins to measure speaking rate and to calculate *rhythm metrics* ([7], [8], [9]), or *Python* scripts to extract prosodic information such as presence of a pitch accent, prominence status of the syllable, or stress group position in the intonational phrase. Most of the tools can be accessed via *Annotation Pro* software [10] which is integrated with the *Polish Rhythmic Database* and additionally enables, among others, automatic phonemic transcription and segmentation of recordings.

2.2 Labeling

The whole speech material used in the current study was automatically segmented at phoneme, syllable and word level [11], and prosodically hand-labeled in *Praat* by four trained annotators and an expert phonetician. In order to facilitate the labeling, to ensure its consistency at multiple tiers and efficient processing of its results, a collection of *Python* scripts was created. They enable, among others, automatic insertion of points (at a *Praat* point tier) corresponding to word boundaries. For each syllable we determined its stress, pitch accent and prominence status and for each word we provided an index (similar, but not identical to ToBI break indices) signaling its position in the prosodic hierarchy according to the phonological model of prosodic structure in [12]. The prosodic labeling was based on objective (phonological/linguistic, acoustic) and subjective (perceptual) criteria. For example, since primary words stress was assigned automatically (it was provided in the output of automatic phonetic alignment), the labelers' task consisted in verifying, if necessary, its presence and position in line with the phonological rules on the one hand ([13], [14]), and perceptual assessment supported by objective acoustic cues on the other ([15], [16], [17]). Syllables which should be stressed, but were not perceived as such, were marked as *destressed*. The same procedure was applied to secondary stress labeling (there were no instances of tertiary stresses). As a result, the annotations reflect *dynamic stress*, i.e. concrete acoustic prominence or salience, rather than an abstract phonological property of a syllable within a word (see discussion in [18], pp. 48-62). *Pitch accent* was defined following [18] as "a local feature of a pitch contour – usually, but not invariably a *pitch change*, and often involving a local maximum or minimum – which signals that the syllable with which it is associated is *prominent* in the utterance" (p. 48). The labelers were instructed to use the information on stress and pitch accent status of a syllable when determining the prominence pattern of an utterance, however in their judgements they were not limited to acoustic cues (cf. [19]). Consequently, it can be assumed that the final results of the labeling reflect the complex nature of

prosodic prominence which does not *necessarily* have to involve acoustic salience, but whose perception is also based, sometimes solely, on (top-down) phonological/linguistic and/or metrical expectations ([18], [19], [20]). Intonational phrases were defined as semantically and prosodically coherent, fluent stretches of speech with the initial boundary marked by high pitch and a pitch accent, and a clear pre-boundary lengthening and boundary tone at the end ([21], [22]). In order to ensure an acceptable level of consistency, the prosodic annotations were cross-checked by at least two transcribers.

2.3 Measurements and methods

2.3.1 Rhythmic prominence intervals

The results of the prosodic labeling were processed automatically by means of *Python* scripts which provided 26 features for every syllable in the analyzed dataset, e.g. stress, pitch accent and prominence status, break index, position in the prosodic structure, duration (of the syllable, its onset, nucleus and coda), identity of the word and the nucleus, etc. The current data contains 1855 instances of intonational phrases and 5428 instances of prominent syllables (almost 35% of all syllables). Altogether 3573 *Rhythmic Prominence Intervals* (RPIs) were extracted from the prosodic annotations by measuring time spans between the *vowels* of the adjacent prominent beats/syllables. One of the reasons for the selection of *vowel onsets* rather than syllable onsets was that the former seem to better correspond to *perceptual (p-) centers* of rhythmic beats than the latter (e.g. [23], [24], [25]). Stretches of syllables preceding the first prominent beat (i.e., anacruses) and following the last prominent beat in the intonational phrase were excluded from the analysis in order to minimize the effect of initial and pre-boundary lengthening at the intonational phrase level (see [26] for a general overview, and [21][15], [21], [22] for Polish). As expected, the overall number of RPIs decreases with speaking rate [very slow: 822, slow: 780, normal: 703, fast: 652, very fast: 616] indicating that some of the prominent beats are *optional* and disappear at faster tempi resulting in a flattening of the rhythmic structure [13]. Bi- and tri-syllabic RPIs were the most frequent (respectively, 36.7% and 30.6% of all RPIs), followed by 4 syllables long RPIs (13%), mono-syllabic RPIs (7.1%), and intervals consisting of 5 (6.1%) and more syllables (6: 3.2%; 7: 1.8%). RPIs longer than 7 syllables were disregarded in the analyses due to very low frequency (below 1%).

2.3.2 Speaking rate

Speaking rate was measured in *syllables per second*, i.e. mean syllable length in each intonational phrase per second was calculated and *averaged* over speakers (20), intended speaking rates (5) and RPI sizes (7), which resulted in 521 data points (cases with the missing values were removed). The averaging procedure was adopted to meet the *independence assumption* of a linear model (the original dataset included multiple responses from the same subject, so these responses could not be regarded as independent from each other; see [27]). A comparison between intended (ISR) and laboratory measured speaking rate (LSR, in syll./sec.) showed that the speakers succeeded in varying the tempo of their reading: the number of syllables per second increased incrementally from very slow to very fast ISR (Figure 1, left), but at the same time it was observed that they differed in their average speaking rate (Figure 1, right). Moreover, the results of ANOVA with ISR as a fixed factor and *speaker* as a random factor, revealed significant differences in the implementation of the five intended tempi among the speakers [for ISR: $F(1855, 4) = 1510$, $p < .001$, $\eta_p^2 = 0.77$; for *speaker:* $F(1855, 19) = 37.9$, $p < .001$, $p < .001$, $\eta_p^2 = 0.28$]. This model accounts for almost 80% of variability in the LSR (based on adjusted R^2).

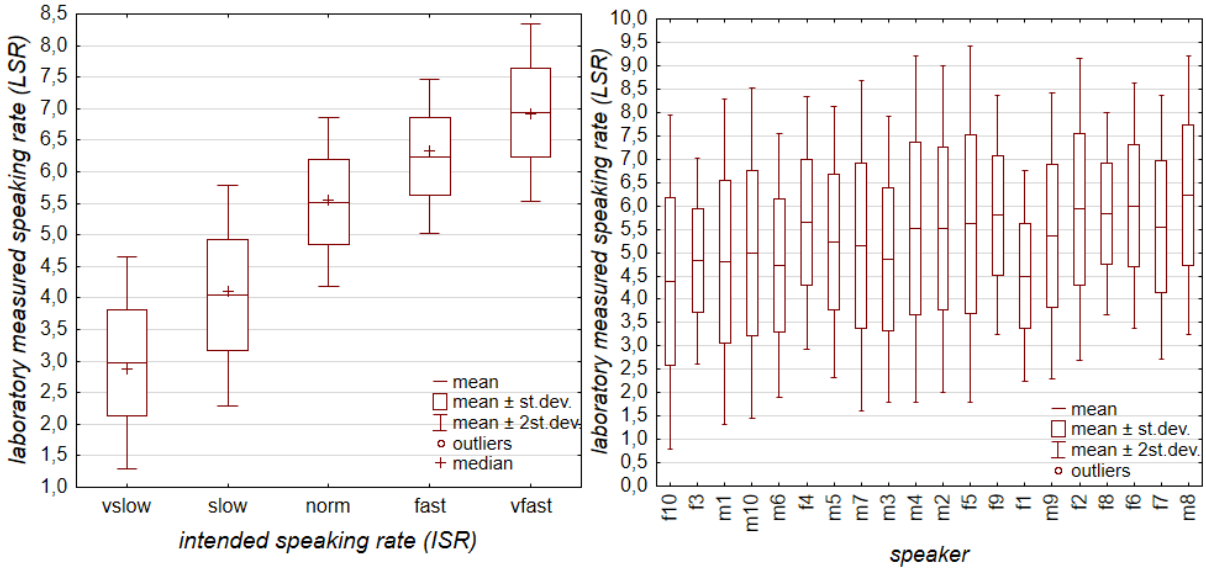

Figure 1 - Comparison between intended and laboratory measured speaking rate (left) and distribution of speaking rate in syll./sec. averaged over the five intended tempi for each speaker (right)

One issue that needs to be considered in order to built a reliable regression model to examine the coupling strength between oscillators operating at various rhythmic levels *across different speaking rates* concerns a possible relation between tempo estimated as syllables per second and syllable count in RPI, so that faster and slower rates are associated, respectively, with relatively greater proportion of longer and shorter RPIs. If such tendency occurs, it may have serious consequences for the regression model, i.e. it will be very difficult to assess the effect of tempo and RPI size separately. Our study requires that *tempo distribution is independent* from syllable count. One solution to such a problem, as proposed by O'Dell & Nieminen and reported in [28] (p. 173), consists in partitioning the sorted data points for each syllable count into groups based on various ranges of quantiles (identical for all syllable counts) and performing a regression on each group. Contrary to [28], where tempo classes were estimated from data as the number of syllables per second, in the current study speaking rate is a *fixed factor*, i.e. tempo was controlled during the recordings by instructing the speakers to accelerate and decelerate their speech delivery, so that their speech samples could be grouped into five distinct tempo classes.

2.3.3 Statistical analyses

In line with the above considerations we examined the speech data using a correlation analysis and a chi-square test (sec. 3.1) to ensure that no confounding factors are introduced into the regression models.

To determine the coupling strength between RPIs and syllables linear regression coefficients were estimated from the data separately for each intended speaking rate and the relative strength parameters *r* were calculated as a ratio of the intercept to the slope ([1], [2], [3], [5]; sec. 3.2). Additionally, we performed linear mixed effects analyses with *speaker* as a random effect. Finally, a multiple linear regression model for tempo-dependent RPI duration estimation was built (sec. 3.3). The statistical analyses were performed using *Statistica 12* software package.

3 Results

3.1 Speaking rate

Firstly, the analysis using Pearson's correlation coefficient indicated *less than a moderate* linear relationship between LSR and syllable count in RPI [$r(521) = 0.36$, $p < .05$]. Secondly, an association between ISR and RPI size was found [$\chi 2(24, 521) = 39.9$, $p = 0.02$]. Examination of the cell frequencies showed this might be due to unbalanced distribution of the "extreme" RPI sizes across the tempo classes: the 6- and 7-syllabic RPIs showed a preponderance of very fast rate (about 40% of all RPIs), while at the same time the proportion of the slowest rate was very low (about 4% and 6% respectively). As concerns the smallest RPIs they were underrepresented at the fastest ISR (only 7 cases, i.e. 6% of all mono-syllabic RPIs). In order to ensure more consistency in the data RPIs shorter than two and longer than six syllables were eliminated and another chi-square test was performed on the reduced dataset. This time we found *no association* between the *RPI size* and intended speaking rate [$\chi 2(12, 358) = 7.26$, $p = 0.84$]. Consequently, we concluded that the procedure suggested by O'Dell & Nieminen is unnecessary and the data is now adequate for the analysis.

3.2 Interaction between RPI and syllabic oscillators and different tempi

Five linear regression models were created to investigate the relation between RPI duration measured in milliseconds (representing the higher-level RPI oscillator) and RPI size counted in syllables (expressing the lower-level syllabic oscillator) at five speaking rates. The relative coupling strength parameter *r* was calculated as a ratio of the intercept to the slope (Table 1).

Table 1 - Simple linear models of RPI duration at five intended speaking rates

tempo	regression equation	coupling strength (*r*)	adj. R^2
v. fast	98 + 116n	0.84	0.87
fast	97 + 128n	0.76	0.83
normal	139 + 130n	1.06	0.78
slow	220 + 161n	1.36	0.52
v. slow	330 + 212n	1.55	0.41

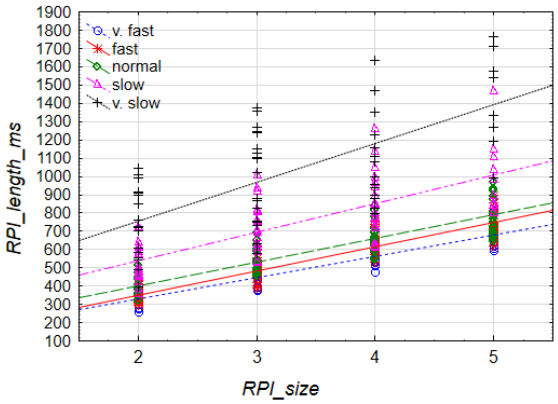

Figure 2 - Regression results for the five tempo classes

All the models and coefficients were statistically significant ($p < .001$). It can be seen in Table 1 that the intercepts and slopes increase with decreasing speaking rate and at the *normal* tempo they take intermediate values. Interestingly, the change in the coefficient values from very fast to normal rate appears to be much smaller comparing to that from normal to very slow rate. As stated in a cross-linguistic study on tempo-dependent timing variability [29] average speaking rate (LSR) and speakers' ability to increase the rate are *language-dependent*, because they are "highly influenced by the language individual phonetic, phonologic and phonotactic syllable structures" (p. 473). It was shown that Spanish, which is traditionally referred to as "syllable-timed", is characterized by faster average speaking rate comparing to languages considered as "stress-timed", i.e. English and German, and that speakers of French are capable of accelerating their speech delivery much more than speakers of English and German. At the same time the authors observed that the LSR changed proportionally in all the languages from normal to very slow rate. The regression results of our study indicate that what may be characteristic of the

Polish language is greater ability of the speakers to *decrease the tempo* than to increase it (Table 1 and Figure 2).

The intercept value for the normal tempo is in between the values reported in [5] for five languages (Spanish, Italian, English, Thai and Greek). With increasing rate our intercepts get closer to the values found for syllable-timed Italian (110) and Spanish (76), whereas with decreasing rate they take values that are similar to (or exceed) those found for stress-timed English (201) and Thai (220). The slopes observed in our study are steeper than those reported in [5].

As suggested by O'Dell & Nieminen (see [28], p. 173) in order to estimate the relationship between tempo and relative coupling strength (*r*) we performed a further regression with r_i as dependent variable and $1/b_i$ (where b_i = individual slopes) as continuous predictor. The results (Table 2) indicate that this relationship is statistically significant and the negative coefficient for $1/b_i$ means that the relative coupling strength decreases significantly with increasing speaking rate. On this basis we can conclude that the lower-level syllabic oscillator gains strength over the higher-level RPI oscillator with increasing rate, which can also be interpreted as a tendency towards syllable-timed rhythm ([1], [2]).

Table 2 - Results of simple linear regression with *r* as dependent and $1/b_i$ as predictor variable

coefficients	**Estimate**	**Std. Error**	**t**	**p**	**-95%**	**+95%**	**Beta (ß)**	**Std. Error ß**	**-95%**	**+95%**
intercept	2.55	0.31	8.29	0.004	1.57	3.53				
$1/b_i$	-205	43	-4.76	0.018	-342	-67.8	-0.94	0.20	-1.57	-0.31
Adjusted R^2: 0.84										

In order to assess how the regression models fit the data we examined adjusted R^2 (Table 1). It can be seen that for the two fast and the normal rate the models containing *RPI size* explain a lot of variability in *RPI duration* in ms (between 78% and 87%), but the models for the slow and very slow ISR perform much worse. These results are in line with the tendencies reported in a previous study based on Polish spontaneous data [28] which, however, did not introduce *intended speaking rate* and obtained adjusted R^2 in between 0.55 and 0.96. Next, we performed *linear mixed effects analyses* with *RPI size* as fixed effect and *speaker* as random effect, and in this way we took by-subject variability in speaking rate (LSR) into account (see sec. 2.3.2). The results indicated that the models with *random intercepts for speakers* had much better fit than the models without the random effect [*adj.* R^2 for: norm = **0.94**, fast = **0.91**, vfast = **0.92**, slow = **0.85**, vslow = **0.87**].

3.3 Multiple regression model of Rhythmic Prominence Interval duration

We built multiple linear regression model for tempo-dependent RPI duration estimation using RPI size (in syllables), laboratory measured speaking rate (LSR in syll./sec.) and the interaction between these two predictor variables. Since the model relies on LSR and not ISR, and we know that there is a relatively weak relation between LSR and RPI size (see sec. 3.1) we use the full dataset, i.e. altogether 521 instances of RPIs varying in size from 1 to 7 syllables. The resulting model and the coefficients are statistically significant ($p < .001$). The adjusted R^2 is very high, indicating a good fit to the data, and the model accounts for **88%** of the variability in the RPI duration. The regression equation for predicting RPI duration is:

$$RPI\_length\_ms = 507 + 286(\text{RPI\_size}) - 76(\text{LSR}) - 22(\text{LSR*RPI\_size})$$

4 Conclusions and future work

In this study, we built tempo-specific simple regression models and used their coefficients to calculate relative coupling strength (r) between two components of Polish speech rhythm that operate at distinct timescales, i.e. the *syllables* and the *Rhythmic Prominence Intervals*. The general findings of our study are consistent with the assumptions of the coupled oscillator model, i.e. dominance of the syllabic oscillator at faster rates (indicated by $r < 1$) and of the RPI oscillator at slower rates ($r > 1$), which can be interpreted, respectively, as a greater degree of syllable- and stress-timing. The long term timing variability that we observe here is just one aspect of the rhythm phenomenon (e.g., we do not know how particular syllables are affected by increasing/decreasing the RPI size and tempo), but it clearly shows that rhythmic patterning within a language can span the continuum between two "extremes", i.e. syllable- and stress-timing, which is consistent with the concept of *coexisting rhythms in language* [30]. Additionally, at normal tempo the coupling strength parameter is close to 1 which is regarded as a *transition "point"* between syllable- and stress-timing [2]. This result illustrates well the fact that Polish is often considered as *rhythmically mixed* or *intermediate*.

The results of *linear mixed effects analyses* indicated that at all speaking rates the models with *random intercepts for speakers* performed much better (in terms of the percentage of variance accounted for as indicated by *adj.* R^2) than the simple regression models. This suggests that average speaking rate and the ability to accelerate/decelerate speech delivery is not only language-specific, but also *speaker-specific*. Therefore grouping of speech data into tempo classes based on intended speaking rate (ISR) may be inconsistent in terms of the actual laboratory speaking rate (LSR), unless ISR is strictly regulated, e.g. by a metronome. Consequently, whenever possible, it seems reasonable to rely on LSR rather than ISR.

In the future we plan to extent the scope of the current study by analyzing additional speech material available in the *Polish Rhythmic Database*, namely read dialogues, poems representing different meters and spontaneous monologues. Moreover, we intend to apply the framework of coupled oscillator model to evaluate the relevance of different components of the hierarchical rhythmic structure (syllables, feet, phonological phrases, etc.) and to investigate the interaction between them (see [31]).

References

[1] O'DELL, M. and T. NIEMINEN: *Speech rhythms as cyclical activity*. In S. OJALA and J. TUOMAINEN (Eds.), Proc. of 21st Meeting of Finnish Phoneticians, pp. 159-168. Turku, Estonia, 2001.

[2] O'DELL, M. and T. NIEMINEN: *Coupled oscillator model of speech rhythm*. Proc. of The 14th International Congress of Phonetic Sciences, Vol. 2, pp. 1075-1078. San Francisco, United States, 1999.

[3] O'DELL, M. and T. NIEMINEN: *Coupled oscillator model for speech timing: Overview and examples*. Proc. of Nordic Prosody, pp. 179-190. Helsinki, Finland, 2009.

[4] ARVANITI, A.: *Rhythm, timing and the timing of rhythm*. Phonetica, 66(1-2), pp. 46-63. 2009.

[5] ERIKSSON, A.: *Aspects of Swedish speech rhythm*. PhD thesis, University of Göteborg, Sweden. 1991.

[6] WAGNER, A., KLESSA, K. and BACHAN, J.: *Polish rhythmic database – new resources for speech timing and rhythm analysis.* In: N. Calzolari et al. (eds.): Proc. of the International Conference on Language Resources and Evaluation (LREC 2016), pp. 4678-4683. Portorož, Slovenia. 2016.

[7] RAMUS, F., NESPOR, M., and MEHLER, J.: *Correlates of linguistic rhythm in the speech signal.* Cognition, vol. 73, pp. 265-292. 1999.

[8] GRABE, E., and LOW, E. L.: *Durational variability in speech and the rhythm class hypothesis.* Laboratory Phonology, vol. 7, pp. 515-546. 2002.
[9] DELLWO, V.: *Influences of speech rate on the acoustic correlates of speech rhythm*. Unpublished PhD thesis, Universität Bonn, Germany, 2010.
[10] KLESSA, K.: Annotation Pro [Software tool]. Version 2.3.1.5. Retrieved from: http://annotationpro.org/ on 2016-02-17.
[11] SZYMAŃSKI M. and GROCHOLEWSKI S.: *Transcription-based automatic segmentation of speech.* Proc. of 2nd Language & Technology Conference, pp. 11-15. Poznan, Poland, 2005.
[12] NESPOR, M. and VOGEL, I.: *Prosodic phonology: with a new foreword,* Studies in Generative Grammar, Vol. 28. De Gruyter Mouton; 2nd ed., 2007.
[13] RUBACH, J. and BOOIJ, G.: *A grid theory of stress in Polish*. Lingua 66, pp. 281–319. 1985.
[14] KRASKA-SZLENK, I.: *The Phonology of Stress in Polish.* LINCOM publishers, 2003.
[15] KLESSA, K.: *Analysis of Segmental Duration for Polish Speech Synthesis Purposes.* Unpublished PhD dissertation, Adam Mickiewicz University in Poznan, Poland, 2006.
[16] WAGNER, A. : *A comprehensive model of intonation for application in speech synthesis*. Unpublished PhD dissertation, Adam Mickiewicz University in Poznan, Poland, 2008.
[17] MALISZ, Z. and WAGNER, P. : *Acoustic-phonetic realisation of Polish syllable prominence: A corpus study*. Speech and Language Technology, vol. 14/15, 105–114. 2012.
[18] LADD, D. R.: *Intonational phonology.* Cambridge University Press, 2nd ed., 2008.
[19] COLE, J., MO, Y., and HASEGAWA-JOHNSON, M.: *Signal-based and expectation-based factors in the perception of prosodic prominence.* Laboratory Phonology, 1(2), 425-452. 2010.
[20] WAGNER, P.: *Great expectations-introspective vs. perceptual prominence ratings and their acoustic correlates.* Proc. of Interspeech, Lisbon, Portugal, 2005.
[21] DEMENKO, G.: *Analysis of suprasegmental features in Polish for speech technology purposes.* Wydawnictwo Naukowe UAM. 1999.
[22] WAGNER, A.: *Acoustic cues for automatic determination of phrasing.* Proc. of Speech Prosody. Chicago, USA, 2010.
[23] ALLEN. G. D.: *The location of rhythmic stress beats in English: An experimental study I.* Language and Speech, 15(1), pp. 72-100. 1972.
[24] MORTON, J., MARCUS S. and FRANKISH, C.: *Perceptual centers (p-centers).* Psychological Review, 83(5), pp. 405–408. 1976.
[25] BARBOSA, P., ARANTES, P., MEIRELES A. R., and VIEIRA, J.: *Abstractness in speech-metronome synchronisation: P-centres as cyclic attractors.* Proc. of Interspeech, pp. 1441–1444. Lisbon, Portugal, 2005.
[26] FLETCHER, J.: *The Prosody of Speech: Timing and Rhythm.* In HARDCASTLE, W. J., LAVER, J., & GIBBON, F. E. (eds.). The handbook of phonetic sciences. John Wiley & Sons. 2nd ed., 2010.
[27] WINTER, B.: *Linear models and linear mixed effects models in R with linguistic applications.* arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf], 2013.
[28] MALISZ. Z.: *Speech rhythm variability in Polish and English: A study of interaction between rhythmic levels.* Unpublished PhD thesis. Adam Mickiewicz University in Poznan, Poland. 2013
[29] DELLWO, V., & WAGNER, P.: *Relations between language rhythm and speech rate.* Proc. of The 15th International Congress of Phonetic Sciences, pp. 471-474. 2003.
[30] NOLAN, F., & ASU, E. L.: *The pairwise variability index and coexisting rhythms in language.* Phonetica, 66(1-2), pp. 64-77. 2009.
[31] O'DELL, M., LENNES, M. and NIEMINEN, T.: *Hierarchical levels of rhythm in conversational speech*. Proc. of Speech Prosody, pp. 355-358. Campinas, Brazil , 2008.