The role of prominences in typologically different rhythms

Gonzalo Eduardo Espinosa[1]

[1]Universidad Nacional del Comahue-CONICET, Argentina

gonzalo.espinosa@fadel.uncoma.edu.ar

Abstract

Speech rhythm is understood as a byproduct of phonetic and phonological characteristics like syllabic structure, vowel reduction and the role of accent. In order to capture these properties, several rhythm metrics have been proposed in the literature so as to give an account of different aspects of rhythm. The aim of this study is to focus on the role of accent, as manifested in the prominences of an utterance, in the timing patterns of two typologically different rhythms represented by Spanish and English. Two rhythm metrics, %V and VarcoV, have been chosen in order to measure the production of a specially designed reading task by speakers from Patagonia, south of Argentina, as well as the southeast of England and the northeast of the United States. Results indicate that (1) there is a significant effect of prominences in the two English varieties, but not in Spanish, as shown by %V and VarcoV, and (2) there is a categorical difference between rhythm types, as shown by VarcoV.

Index Terms: speech rhythm, metrics, accent, Spanish, English

1. Introduction

Speech rhythm has been interpreted as an inherent property of languages [1], [2] or as a byproduct of phonological characteristics [3]–[8]. Two main distinct classes of rhythm are represented by English and Spanish, the former being stress-timed and the latter syllable-timed. From a phonological point of view, i.e., considering rhythm as an epiphenomenon, stress-timed languages show a complex syllable structure, have vowel reduction processes and their speech is greatly influenced by accent; whereas syllable-timed languages exhibit simple syllable structures, do not allow systematic vowel reduction and are not significantly influenced by the role of accent [3], [5], [6].

Acoustic measurements are implemented in order to give an empirical account of speech rhythm. As a result, rhythm metrics have been introduced in prosodic studies as a tool for reflecting syllabic structure and vowel reduction [9]. These metrics compute the duration of vocalic and consonantal intervals in utterances and are able to classify typologically different speech rhythms: %V indicates the proportion of vocalic intervals divided by the overall duration of an utterance, ΔV shows the standard deviation of vocalic intervals, and ΔC reveals the standard deviation of consonantal intervals. As it has been found that speech rate has an influence on the metric results [10], the original duration metrics have been modified in order to control for speech rate, obtaining thus the metrics VarcoV (vocalic intervals) and VarcoC (consonantal intervals) that divide the standard deviations by duration means [10]–[12]. Despite some criticism to the rhythm metrics [13], [14], more proposals and adjustments have been implemented in order to better account for speech rhythm (see [15], [16], for overviews).

With the aim to analyze the role that syllabic structure plays in speech rhythm, there have been attempts to control for the type of syllables and account for typologically different speech rhythms [13], [14], [17]–[20]. For instance, Renwick [20] shows that there is a significant correlation between the phonotactic properties of a language and durational metrics, such as %V. Prieto and colleagues [19] provide evidence that languages like English and Spanish show categorically different speech rhythms even when the syllabic structure is controlled for. Arvaniti [13], [14] considers the strong correlation between phonotactics and metrics as a weak point, whereas authors like Fuchs [15] interpret it as the strong predicting power of these measures.

Vowel reduction, on the other hand, has been considered as another property that influences speech rhythm [3], [5], [6] in the sense that a greater degree of vowel reduction results in a stronger contrast between stressed and unstressed syllables [21]. Reduction in vowel duration has been claimed to be captured by some rhythm metrics like ΔV in combination with other phonological factors such as contrastive vowel length [9]. Alternatively, spectral vowel reduction can be measured by analyzing formant information, particularly F1 and F2, as carried out experimentally by [22], [23] for English. Significant reduction processes are not attested in Spanish, i.e., the same vowel qualities are kept in strong and weak syllables [24].

Having outlined two of the properties that make up the structure of speech rhythm (i.e., syllable structure and vowel reduction [3], [6]), a third main aspect is considered in the present study: the role of accent. Strictly speaking, Dauer [5], [6] refers to *accent* as the highest level of prominence in a word (i.e., lexical stress). In this paper, the term *accent* will make reference to the primary lexical stress that is manifested as higher levels of prominence in an utterance. Taking into account that in stress-timed languages accents play a more significant role in speech rhythm than in syllable-timed languages ([3], [5], [6]), the main goal of this study is to provide empirical evidence for the influence that the number of prominences has on the rhythmic structure of Spanish and English as accounted for by durational metrics. In addition, given the diverse interpretations and results in relation to the lengthening effects of the final interval (FI) that is obtained in the segmentation procedure (e.g. [12], [25]–[27]), the inclusion and exclusion of this interval will be compared so as to see how productive the FIs are when computing rhythm metrics.

2. Methodology

2.1. Participants

A total of 21 participants (11 females, 10 males) were recorded in a reading task. For English, two groups of speakers were selected. One group corresponds to southeast England (BrE) with speakers (*n*=7; 4 females) who were born and raised in this area and their age ranged from 21 to 32 years (*M*=25, *SD*=3.8). The second group corresponds to speakers (*n*=7; 4 females) who were born and raised in the northeast area of the United States (AmE) and their age ranged from 18 to 34 (*M*=25, *SD*=5.1). The speakers (*n*=7; 3 females) in the Spanish group (Sp), on the other hand, were born and raised in Patagonia, the southern region of Argentina. Their age ranged from 21 to 39 (*M*=29, *SD*=6.7).

All speakers were monolinguals and they all shared a similar educational background: they were completing or had already finished their studies at university level. Their parents were born and raised in the same regions or had lived there for more than 30 years. None of the speakers was an expert in linguistics or had in-depth metalinguistic knowledge of the pronunciation of their mother tongue. None of them reported having hearing or speaking impairments.

2.2. Stimuli

For each language, 12 sentences were designed. Each sentence had 10 syllables and they were grouped in 3 conditions: sentences with 4, 3 and 2 prominences (4Prom, 3Prom and 2Prom). For each condition there were equal numbers of syllable types (open and closed) in prominent and non-prominent syllables. During the design of these sentences, special attention was given to their meaning so as to avoid odd phrases. They were piloted with native speakers of Spanish and English who were not included in the experiment. Adjustments were made until the desired number of prominences per utterance was achieved. Examples of the material are presented in Tables 1 and 2, for English and Spanish, respectively. In these tables, the number of prominences (Prom) and one sentence for each condition are shown. Capitals represent prominences.

Table 1: *Examples of sentences in English.*

Prom	Sentences
4	the MAN can SEE the PATH to HAppiness
3	the BROthers will DREAM of the LEgacy
2	the SUBsidies can be the soLUtion

Table 2: *Examples of sentences in Spanish.*

Prom	Sentences
4	el magNAte CANta DOS coPLItas
3	los muCHAchos coMIENzan eTApas
2	yo desinFECto las montaÑItas

2.3. Procedure

Participants were interviewed in special rooms in three different universities: University College London, UK, and Hamilton College, USA, for the BrE and AmE groups, and Universidad Nacional del Comahue, Argentina, for the Sp group. They performed the reading task for the present paper and a series of other tasks that were used for other studies. The sentences were randomized and added to a PowerPoint presentation and participants could click through the slides. Before reading aloud the sentences to be recorded, participants read them in silence so as to make sense of them. With the purpose of obtaining a homogenous prosodic environment, during the recording, the researcher paid particular attention to the segmentation and intonation so that there was only one utterance per sentence with a falling intonation and with the tonic syllable on the last prominence. Participants were asked to read the sentences again whenever they produced more than one utterance per sentence (in approximately 20 cases), placed the tonic syllable somewhere else (in 8 cases) or used a rising intonation (in 7 cases).

A total of 252 utterances (4 sentences × 3 conditions × 21 speakers) and 2520 syllables (252 utterances × 10 syllables) were obtained in the recordings. These were made with Røde NT1-A and CAD GXL2200 condenser microphones. For each speaker a WAV file was obtained at 44.100 Hz, 16 bits mono. Each file was then edited with Audacity [28] so as to get one file for each condition per speaker. The files were segmented manually into vocalic and consonantal intervals by means of Praat [29], according to the criteria followed by [12], [15], [30], [31]. Two rhythm metrics were chosen in the present study: %V and VarcoV, as suggested by [12], [15], [32], because they are able to account for typologically different speech rhythm, especially when they are plotted together. The metric scores were obtained by means of method B in Correlatore [33], which computes the duration of the segmented intervals and considers the values for each inter-pausal string in Praat textgrids. Finally, ANOVAs and Tukey post-hoc tests were carried out with Infostat [34].

3. Results

In relation to the descriptive statistical analysis, Tables 3 and 4 show the mean values and the standard deviations (*SD*) for %V and VarcoV, respectively. For each metric, the scores obtained with the inclusion and exclusion of FIs are shown.

Table 3: *Mean values and (SD) for %V.*

Group	4Prom	3Prom	2Prom
	with FIs		
Sp	40.32 (2.27)	39.99 (2.58)	39.72 (3.17)
BrE	43.56 (2.04)	41.69 (2.38)	43.50 (3.58)
AmE	43.95 (1.90)	40.52 (2.34)	44.03 (2.24)
	without FIs		
Sp	42.63 (2.01)	41.36 (2.42)	41.33 (3.05)
BrE	40.83 (1.75)	37.75 (1.84)	41.97 (3.67)
AmE	41.62 (1.58)	36.86 (1.85)	42.51 (1.87)

Table 4: *Mean values and (SD) for VarcoV.*

Group	4Prom	3Prom	2Prom
	with FIs		
Sp	30.52 (5.02)	28.19 (5.30)	27.53 (4.04)
BrE	63.37 (6.11)	67.05 (10.07)	56.49 (7.05)
AmE	61.49 (3.81)	58.54 (5.81)	54.07 (5.66)
	without FIs		
Sp	29.68 (4.62)	27.17 (3.35)	26.49 (3.18)
BrE	64.90 (5.77)	56.82 (8.24)	53.62 (8.42)
AmE	63.08 (5.24)	48.26 (6.77)	54.03 (7.30)

For the inferential statistical data, all F values have the following degrees of freedom: 2 and 18. Normality and homogeneity of variance have been verified with the Shapiro-Wilks and Levene tests, respectively, which have revealed no serious violations of these assumptions.

In order to analyze the effects of prominences within each language, Language Group (Sp, BrE and AmE) was set as a factor and Number of Prominences (4Prom, 3Prom and 2Prom) as the dependent variable. For the %V metric, the one-way ANOVA analysis indicates that there is no effect of the number of prominences in the Spanish group (F=0.32, p=.7312, with FIs, and F=0.60, p=.5599, without FIs). For the BrE group, there is no significant effect of prominences when FIs are computed (F=1.06, p=.3686), while there is a significant effect when FIs are excluded (F=5.01, p=.0186). For the AmE group, there is a significant effect of prominences in both treatments of FIs (F=5.98, p=.0102, with FIs, and F=20.70, p<.0001, without FIs). As far as VarcoV is concerned, the statistical analysis of this metric has shown that the Spanish group does not have a significant effect of prominences (F=0.75, p=.4887, with FIs, and F=1.39, p=.2740, without FIs). The BrE group only shows a significant effect of prominences when FIs are excluded (F=4.13, p=.0333), but not when they are included (F=3.20, p=.0647). Lastly, the AmE group shows a significant effect when FIs are included (F=3.64, p=.0469) and excluded (F=9.25, p=.0017).

In relation to the effects of language, Number of Prominences was set as a factor and Language Group as the dependent variable. The results for %V indicate that, in the 4Prom condition with inclusion of FIs, languages are significantly distinguished (F=4.71, p=.0226), but the post-hoc Tukey test reveals that there are shared values between the Sp and BrE groups, on the one hand, and the BrE and AmE groups, on the other. When FIs were excluded, no significant difference is found among languages (F=1.77, p=.1987). In the 3Prom condition, there are no significant differences when FIs are included (F=.89, p=.4289), but when FIs are excluded, languages do show a significant difference (F=9.43, p=.0016) and the Tukey test differentiates Sp from BrE and AmE. Lastly, the 2Prom condition shows a significant difference among languages with FIs included (F=4.16, p=.0326) and the Tukey test separates the Spanish group from the two English groups. There is no significant difference, however, when FIs are excluded (F=.28, p=.7586). In relation to VarcoV, results indicate that this metric is able to show a highly significant difference among languages (p<.0001) in the three conditions: 4Prom (F=92.74, with FIs, and F=100.67, without FIs), 3Prom (F=53.70, with FIs, and F=39.17, without FIs) and 2Prom (F=55.25, with FIs, and F=38.93, without FIs). The post-hoc Tukey test distinguishes the Sp group from the BrE and AmE groups.

For ease of illustration, the two metrics implemented in this study are plotted in Figures 1 and 2, with FIs and without FIs, respectively. These figures contain the mean values and the standard errors. The number of prominences is included between brackets following the group of speakers. As these figures clearly show, language groups occupy distinct positions and the number of prominences reflects different grouping effects in each language.

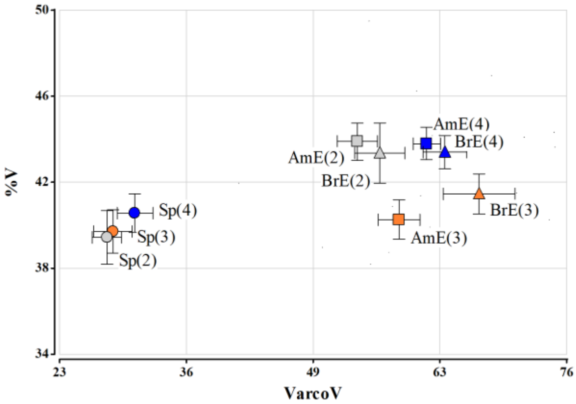

Figure 1: *Values for %V and VarcoV with FIs.*

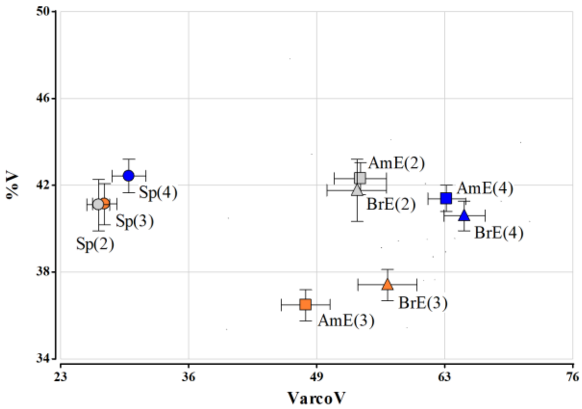

Figure 2: *Values for %V and VarcoV without FIs.*

4. Discussion

The rhythm metrics implemented in the present study have been able to account for the effect of prominences and the effect of language type in different speech rhythms. The results agree with the idea that these measurements are a useful tool for distinguishing rhythm types (e.g. [9]–[11], [19], [31], [32], among many others). Specifically, VarcoV has shown categorical differences between Spanish, on the one hand, and the two English varieties, on the other. This fact indicates that VarcoV might be more productive to represent rhythm types than %V. However, it would be interesting to see the way in which other rhythm metrics proposed in the literature (see e.g. [15], [16], [32]) distinguish language groups with the same elicited production analyzed in this study.

In addition, the results point out that %V and VarcoV have captured the effects of the role of accents (cf. [3], [5], [6]), as manifested by the number of prominences in an utterance. If speech rhythm is understood as the sum of phonetic and phonological characteristics, the main contribution of the present investigation is related to the fact that the role of accents in rhythm has been accounted for in two commonly studied languages. With comparable stimuli in terms of number and type of syllable, the metrics %V and VarcoV have indicated that accents do play a role in stress-timed languages like English, but not in syllable-timed languages like Spanish. It would be of great relevance to replicate the present experiment with other traditionally classified stress-timed languages like German or Dutch and syllable-timed languages like French or Italian.

In relation to FIs, there have been extensive discussions about the inclusion and exclusion of these intervals. Some studies suggest that the exclusion of FIs results in a better discrimination of language types (e.g. [25]). Some others include FIs, because they are considered as part of the intrinsic characteristics of each rhythm type (e.g. [12], [26], [27]). The results of the present study indicate that, when accounting for language types, %V is somehow more sensitive with inclusion of FIs, whereas VarcoV is highly sensitive in both treatments of FIs. When accounting for the effect of number of prominences, both %V and VarcoV seem to be more sensitive without computing FIs. More research needs to be done in order to certainly state how useful it is whether to include FIs or not.

As far as the differences between the two standard varieties of English are concerned, the effects of number of prominences have been different when FIs are computed. Both metrics, %V and VarcoV, have shown that BrE does not show a significant effect of number of prominences with inclusion of FIs (p=.3686 for %V, and p=.0647 for VarcoV), while AmE does (p=.0102 for %V, and p=.0469 for VarcoV). Perhaps there are some interesting dialectal differences with respect to FIs and their final lengthening effects, but more speakers and speech production are needed.

Finally, another promising research area is that of second language acquisition, as studied by e.g. [12], [18], [35]–[37], with a focus on the role of prominences in the structuring of speech rhythm. For instance, it would be of great significance to see how learners whose mother tongue does not show effects of prominence (e.g. Spanish) acquire a second or foreign language that exhibits a great influence of the number of prominences in an utterance (e.g. English).

5. Conclusions

The effects of the number of prominences in an utterance have been reflected in %V and VarcoV, the two rhythm metrics employed in this study. These effects have been observed in the English groups of England and the United States, which indicates a typical characteristic of stress-timed rhythm. The Spanish group has not shown a significant effect of prominences in any of the metrics, which confirms the idea that in syllable-timed languages accents do not play a significant role in the rhythmic structure.

In relation to accounting for different rhythmic typologies, only VarcoV has shown categorical differences between the Spanish group on the one hand and the two English groups on the other. These differences have been observed in the three conditions of the current experiment: 4, 3 and 2 prominences in 10-syllabled utterances.

6. Acknowledgements

The author would like to acknowledge Prof. Faith Liu and Prof. María Gabriela Portal for their invaluable help with the recordings at University College London and Hamilton College, respectively. Furthermore, he would like to thank Paolo Mairano for his assistance with Correlatore.

7. References

[1] D. Abercrombie, *Elements of General Phonetics*. Edinburgh University Press, 1967.

[2] K. L. Pike, *The intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

[3] P. M. Bertinetto, 'Reflections on the Dichotomy "Stress" vs. "Syllable-timing"', *Rev. Phon. Appliquée*, vol. 91–93, pp. 99–130, 1989.

[4] R. Dasher and D. Bolinger, 'On pre-accentual lengthening', *J. Int. Phon. Assoc.*, vol. 12, pp. 58–69, 1982.

[5] R. M. Dauer, 'Stress-timing and syllable-timing reanalyzed', *J. Phon.*, vol. 11, no. 1, pp. 51–62, 1983.

[6] R. M. Dauer, 'Phonetic and phonological components of language rhythm', in *11th International Congress of Phonetic Sciences*, Estonia, 1987, pp. 447–450.

[7] P. M. Bertinetto, '«Syllabic Blood» ovvero l'italiano come lingua ad isocronismo sillabico', *Studi Gramm. Ital.*, vol. 6, pp. 69–96, 1977.

[8] P. M. Bertinetto, *Strutture prosodiche dell'italiano: accento, quantità, sillaba, giuntura, fondamenti metrici*. Firenze: Presso l'Accad. della Crusca, 1981.

[9] F. Ramus, M. Nespor, and J. Mehler, 'Correlates of linguistic rhythm in the speech signal', *Cognition*, vol. 73, pp. 265–292, 1999.

[10] V. Dellwo and P. Wagner, 'Relations between language rhythm and speech rate', in *Proceedings of International Congress of Phonetic Science*, Barcelona, 2003, pp. 471–474.

[11] V. Dellwo, 'Rhythm and Speech Rate: A Variation Coefficient for deltaC', in *Language and language-processing*, P. Karnowski and I. Szigeti, Eds. Frankfurt am Main: Peter Lang., 2006, pp. 231 – 241.

[12] L. White and S. L. Mattys, 'Calibrating rhythm: First language and second language studies', *J. Phon.*, vol. 35, no. 4, pp. 501–522, 2007.

[13] A. Arvaniti, 'Rhythm, Timing and the Timing of Rhythm', *Phonetica*, vol. 66, no. 1–2, pp. 46–63, 2009.

[14] A. Arvaniti, 'The usefulness of metrics in the quantification of speech rhythm', *J. Phon.*, vol. 40, no. 3, pp. 351–373, 2012.

[15] R. Fuchs, *Speech rhythm in varieties of English: Evidence from educated Indian English and British English*. Singapore: Springer, 2016.

[16] A. Loukina, G. Kochanski, B. Rosner, E. Keane, and C. Shih, 'Rhythm measures and dimensions of durational variation in speech.', *J. Acoust. Soc. Am.*, vol. 129, no. 5, pp. 3258–3270, 2011.

[17] J. Brown and E. Matene, 'Is speech rhythm an intrinsic property of language?', in *Fifteenth Annual Conference of the International Speech Communication Association*, Singapore, 2014, pp. 1693–1697.

[18] G. E. Espinosa, 'La adquisición del ritmo inglés por hablantes nativos de español. El caso de aprendientes argentinos en contexto de instrucción formal', unpublished doctoral thesis, Universidad de Buenos Aires, Buenos Aires, Argentina, 2018.

[19] P. Prieto, M. D. M. Vanrell, L. Astruc, E. Payne, and B. Post, 'Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish', *Speech Commun.*, vol. 54, no. 6, pp. 681–702, 2012.

[20] M. E. Renwick, 'What does% V actually measure?', *Cornell Work. Pap Phon Phonol*, vol. 3, pp. 1–20, 2012.

[21] P. Roach, 'On the distinction between "Stress-timed" and "Syllable-timed" Languages', in *Linguistic controversies*, D. Crystal, Ed. London: Edward Arnold, 1982, pp. 73–79.

[22] E. Byers and M. Yavas, 'Vowel reduction in word-final position by early and late Spanish-English bilinguals', *PloS One*, vol. 12, no. 4, p. e0175226, 2017.

[23] L. Rallo Fabra, 'Can Nonnative Speakers Reduce English Vowels in a Native-Like Fashion? Evidence from L1-Spanish L2-English Bilinguals', *Phonetica*, vol. 72, no. 2–3, pp. 162–181, 2015.

[24] J. I. Hualde, *Los sonidos del español*. Cambridge: Cambridge University Press, 2014.

[25] A. Loukina, G. Kochanski, C. Shih, E. Keane, and I. Watson, ‘Rhythm measures with language-independent segmentation’, in *Proceedings of Interspeech 2009*, 2009, pp. 1931–1934.

[26] E. Grabe and E. L. Low, ‘Durational Variability in Speech and the Rhythm Class Hypothesis’, in *Papers in Laboratory Phonology 7*, N. Warner and C. Gussenhoven, Eds. Berlin: Mouton de Gruyter, 2002, pp. 515–546.

[27] E. Kireva and C. Gabriel, ‘Rhythm properties of a contact variety: Comparing read and semi-spontaneous speech in Argentinean Porteño Spanish’, in *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations*, E. Delais-Roussarie, M. Avanzi, and S. Herment, Eds. Berlin: Springer, 2015, pp. 149–168.

[28] Equipo de Audacity, *Audacity (R): Free Audio Editor and Recorder*. 2015.

[29] P. Boersma and D. Weenink, ‘Praat: doing phonetics by computer (Version 4.5.)[Computer program]’, *Version 5408 Httpwwwpraatorg*, 2015.

[30] P. Mairano, ‘Rhythm typology: acoustic and perceptive studies’, Doctoral, University of Turin, 2011.

[31] P. Mairano and A. Romano, ‘Rhythm metrics for 21 languages’, in *ICPhS XVII*, Hong Kong, 2011, pp. 1318–1321.

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

[33] P. Mairano and A. Romano, ‘Un confronto tra diverse metriche ritmiche usando Correlatore’, in *La dimensione temporale del parlato (Proc. of the V National AISV Congress, University of Zurich, Collegiengebaude)*, 2010, pp. 79–100.

[34] J. A. Di Rienzo, F. Casanoves, M. G. Balzarini, L. González, M. Tablada, and C. W. Robledo, *InfoStat*. Universidad Nacional de Córdoba, Argentina: Grupo InfoStat, FCA, 2008.

[35] C. Gabriel and E. Kireva, ‘Prosodic Transfer in Learner and Contact Varieties’, *Stud. Second Lang. Acquis.*, vol. 36, no. 02, pp. 257–281, 2014.

[36] M. Ordin and L. Polyanskaya, ‘Development of timing patterns in first and second languages’, *System*, vol. 42, no. 1, pp. 244–257, 2014.

[37] M. Ordin and L. Polyanskaya, ‘Acquisition of speech rhythm in a second language by learners with rhythmically different native languages’, *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 533–544, 2015.