



Acquisition of English Speech Rhythm by Monolingual Children

Mikhail Ordin¹, Leona Polyanskaya²

¹ Universität Bielefeld, Fakultät der Sprach- und Literaturwissenschaft, Germany

mikhail.ordin@gmail.com, leona.polyanskaya@uni-bielefeld.de

Abstract

We investigated how English rhythmic patterns develop in the course of first language acquisition by children between four and twelve years. We have empirically confirmed that rhythm becomes increasingly more stress-timed as acquisition progresses, which is revealed by higher durational variability of syllables, vocalic sequences and consonantal clusters in speech delivered by older children compared to younger children. The development of speech rhythm was studied using the same speech material produced by children at various ages, while controlling for the differences in phonotactics and syllable structure. The tendency to deliver speech with higher durational variability at later stages of language acquisition emerges even when the phonological differences in speech material – inevitable in uncontrolled speech delivered by children of different ages – are controlled for. This finding indicates that the language-specific phonetic timing patterns are established as a function of age, probably as a result of the motor control development.

Index Terms: speech rhythm, first language acquisition, durational variability, timing patterns, phonological development, motor control

1. Introduction

Languages that are auditorily perceived as rhythmically contrastive (English, German, Russian on the one hand and Spanish, French, Italian on the other hand) also feature syntactic and phonological distinctions, and differ in speech processes like vowel reduction, epenthesis and consonantal assimilations [1]. These differences lead to the emergence of language-specific timing patterns in languages that are considered to have different linguistic rhythms.

The language-specific timing patterns also emerge due to phonotactic constraints. In some languages, these constraints are strict, leading to fewer consonantal clusters and overwhelming number of simple CV (consonant-vowel) syllables (French, Japanese, etc). In other languages (English, German, etc), these constraints are loose, and allow rather complex syllables with multisegmental consonantal clusters in syllable codas and onsets (e.g., CCCVCCCC). Presence and absence of phonological oppositions between long and short vowels or between geminates and non-geminates (double and single consonants respectively) also result in phonetic timing differences. Existence of complex syllables and phonemic length contrasts lead to higher variation in duration of vowel sequences and consonantal clusters. The durational variability can further be enhanced by vowel reduction in unstressed syllables, tempo fluctuations, degree of phrase-final lengthening, durational increase in stressed vowels that have to be more marked in languages with free or flexible lexical stress (Russian, German, etc) than in languages with fixed or

more predictable location of lexical stress (Hungarian, Finnish, etc). A detailed discussion of linguistic factors that enhance or inhibit variation in duration of vowels and consonantal clusters can be found in [1] and [2].

However, differences in durational variability between rhythmically different languages remain even when the phonological and phonotactic differences between the analysed utterances in these languages are controlled for [3]. Therefore, language-specific patterns in durational variability emerge not only due to phonological and phonotactic factors, but also due to language-specific phonetic, surface timing patterns.

The degree of durational variability creates the auditory impression of rhythmic differences. The languages with higher durational variability are referred to as stress-timed, and the languages with low durational variability are referred to as syllable-timed. There is no evidence that the differences in the degree of durational variability between languages are categorical. However, there is evidence that durational variability can vary substantially between utterances in stress-timed languages. Although most utterances in English, for example, are produced with a rather high durational variability, Arvaniti [4] showed that some utterances may exhibit a low degree of variation in duration of speech intervals. Besides, some languages do not fit well into stress- and syllable-timing dichotomy. Therefore, the distinction between linguistic rhythms is probably not categorical, but continuous, and better expressed in relational terms, i.e., language X is more stress-timed compared to language Y, or utterance (or speaker) X exhibits more syllable-timed rhythm compared to utterance (or speaker) Y.

The degree of stress- or syllable-timing is represented by the rhythm metrics that measure durational variability of vocalic sequences (V) and consonantal clusters (C). These rhythm metrics have also been successfully applied to syllables (S) and feet [5] and [6]. These rhythm measures include the standard deviation (Δ) in C and V durations [7] and [8], and the pairwise variability index (rPVI), which is the averaged difference in duration of consecutive intervals pairwise [9]. However, these measures of variability are highly influenced by the speech tempo – when speech is delivered at a faster rate, the mean duration of speech intervals, especially of V and S, decreases, therefore the standard deviation of C and V durations in faster speech also decreases. Different methods of normalization have been suggested to neutralize the effect of tempo fluctuation on the variability measures. The most common normalization method is to divide the metric score by the mean duration of the interval, to which the metric has been applied [10]. Thus, the standard deviation is turned into the coefficient of variation (Varco), and rPVI is turned into nPVI (prefix *n* stands for *normalized*). Ramus et al. in [7] and [8] suggested another commonly used rhythm

metric: The percentage of the vocalic material in an utterance (%V), which has been reported to be robust to tempo fluctuations [11]. A more comprehensive list of the rhythm measures can be found in [12]. Higher %V and lower scores of other metrics (i.e., lower durational variability) are indicative of lower degree of stress-timing.

1.1. Rhythm changes in the course of the first language acquisition

Many studies have been devoted to capturing the rhythmic differences between adult and child speech [13], [14] and [15], and between rhythmic patterns in first (L1) and second (L2) language [16], [17] and [18]. However, very few studies have been focused on the rhythmic changes that occur in the course of acquisition, especially in the L1.

Allen and Hawkins [19] suggested that children, irrespective of their target language, exhibit more syllable-timed rhythm because they often reduce syllables to CV structures, simplify consonantal clusters, and do not reduce vowels to the degree adult speakers do. Besides, children's vocabulary is more limited and usually consists of shorter words. Their utterances are also shorter, therefore child speech offers fewer opportunities to enhance durational contrasts between more and less prominent syllables.

Grabe, Post and Watson [20] applied rhythm metrics to study the development of timing patterns in French and English learning children at 4 years. They found that both French and English children delivered speech with lower durational variability than adults, yet even at this age French children were more syllable-timed than English children. Following Allen and Hawkins [19], the authors concluded that syllable-timing is a default setting, and stress-timing is more marked and developed later in the course of language acquisition.

Bunta and Ingram [14] applied PVI metrics to speech delivered by younger (45-53 months) and older (54-62 months) English-Spanish bilinguals and found that vocalic pairwise durational variability is higher in speech of older children in both languages. This indicates the increase in stress-timing in child speech. No difference in consonantal PVI between younger and older children has been found. Unfortunately, significance tests for monolingual Spanish and monolingual English children of different ages were not provided.

Whitworth [21] compared speech rhythm in German and in English produced by bilingual children between 5 and 13 years. His finding shows that younger children deliver speech with more variable V and C intervals in both languages than their older peers. Payne et al. [13] also found that durational variability of vocalic intervals increases with age between 2 and 6 years in Spanish, Catalan and English. However, their data showed that consonantal variability diminishes with age. The increase of vocalic variability, found in both studies [13] and [21], indicates that child speech becomes more stress-timed as acquisition progresses. Payne et al. [13] explained the opposite tendency regarding consonantal variability by incomplete control over production of consonants, which leads to a greater timing variability in speech production of younger children. However, there is an alternative interpretation that can – at least in part – contribute to interpretation of the finding. Payne et al. [13] used normalized metrics to capture vocalic durational variability, and non-normalized metrics to

measure consonantal durational variability. Younger children deliver speech at a slower rate than older children [15], and faster speech rate can potentially result in the decrease of non-normalized metrics.

Ordin and Polyanskaya [15] used longitudinal data (i.e., speech of the same kids recorded at different ages: 3, 4, and 5 years) to investigate the development of timing patterns in English. They did not find significant differences in consonantal durational variability between utterances produced by the children at different ages, neither in terms of normalized, nor in terms of raw metrics. However, they also detected a significant increase in vocalic durational variability, captured by the normalized rhythm metrics.

1.2. Goals and hypothesis of the current study

The hypothesis that speech rhythm becomes increasingly more stress-timed in the course of L1 acquisition needs further support for the following reasons:

1) With the exception of Payne et al. [13], who made an effort to elicit similar utterances from children and adult participants (within the limitations of the data collection methodology and peculiarities of speech production by young children), earlier studies in rhythm development in the L1 did not directly control for the material. Older children produce longer sentences and words, have larger active lexicon, and produce more complex sentences than younger peers. This results in phonotactic, segmental and structural differences between utterances produced by children of different age, when material is not controlled for. Therefore, it is difficult to disentangle the influence of speech material from the influence of age-related developmental differences on metric scores. We designed our study to control for the differences in speech material and to confirm that the speech rhythm per se becomes more stress-timed with age, irrespective of the phonotactic, lexical, syntactic and phonological characteristics of the delivered utterances.

2) Inter-speaker idiosyncrasies in speech production present another source of substantial variation in metric scores [11] and [12], therefore care should be taken when generalizing the results obtained on a small number of speakers over a particular group of speakers. Although Ordin and Polyanskaya [15] adopted a longitudinal approach and thus removed variation in metric scores due to inter-speaker differences between age groups, other studies are not void of this influence: Payne et al. [13] consider utterances of three children as representative of a particular age group in each language, Bunta and Ingram [14] make generalizations on the basis of five speakers per group, Whitworth [21] is a collection of case studies (one speaker per “age*language group”). Therefore, a larger-scale study is necessary to confirm the hypothesis.

3) Previous reports were based on comparing groups, while it makes more sense to investigate the development of metric scores as a function of age on a continuous scale, when age (e.g., in months) is introduced into analysis as a continuous variable, rather than arbitrarily lumping children of a certain age range into one age category.

4) Finally, we need to increase the age range to investigate the development of L1 rhythmic patterns over a longer time.

Although previous studies present a number of insightful conclusions into the nature of L1 rhythm acquisition and provide collaborative evidence to support the hypothesis that syllable-timing is indeed a default setting irrespective of the

target language, and the degree of stress-timing increases in the course of L1 acquisition [19], we hope to provide corroborative evidence to confirm the hypothesis by taking into account the sketched problems of earlier studies.

2. Method

2.1. Participants

We recorded 52 monolingual children in the age from 4:7 to 11:7. median is 7:6 (years:months). Children were British English speakers from monolingual British English speaking parents. Older children (from 10:3 to 11:7) had regular contact with American English variety, younger children had no exposure to American English in daily life (exposure via mass media cannot be excluded, of course). The recordings were done in schools at British military bases in Germany. The military bases are closed communities with supermarkets, schools, restaurants, entertainment centres, and inhabitants usually do not even leave the base premises except for occasional family outing during weekend, therefore inhabitants have very little exposure to the German language (not even via television). Besides, we selected those children whose families had moved to the base not more than three months prior to the recording, and whose parents reported not to be able to be fluent in foreign languages. German is not spoken at the bases, neither is it taught at schools, and English is the only medium of education.

2.2. Procedure

We prompted and recorded 33 comparable sentences from each speaker using pictures with accompanying descriptive sentences. Pictures were shown to those children who were fluent readers, and we asked them to read the sentences. Each picture was shown several times. After the “memorization phase”, the same pictures without the sentences were displayed one by one, and we asked children to say the sentence they had read on each picture. These productions were recorded and analysed. With younger children, we did not use “memorization”. We showed the pictures and asked children to say what was on the picture. Verbal prompts were suggested to elicit the sentences that are comparable regarding lexical, segmental and structural content. After the child produced the adequate sentence, we asked him/her to say the sentence again into the microphone and recorded the final repetition. A school teacher known to the children, native British English speaker, supervised the procedure and helped to elicit the sentences from non-reading children.

We used acoustic shields set up in a quiet sound-treated room located at school premises, and obtained the recordings of high quality at 48kHz sampling frequency, 16bit, in mono PCI format. We used a studio condenser Samson C01U Pro microphone that gave flat response in the frequency range between 20Hz and 20kHz (the response above 20kHz was not optimal, but this frequency range is not important for speech analysis). The recordings were directly digitized into Toshiba A11 laptop using Audacity software. The boundaries of V, C and S intervals were marked in Praat using the criteria described in details by Ordin and Polyanskaya [6].

2.2.1. Rhythm metrics

As speech tempo grows with age [15], [13], the valued of raw rhythm metrics are expected to be affected by two

conflicting forces: the increase in speech tempo will drive the scores of the raw metrics down, while the growing control of speech timing will enhance durational variability, which will raise the values of the rhythm metrics. Therefore, it will be difficult to interpret any developmental trend for the raw rhythm metrics, and we decided not to include them into our analysis. Besides, Ordin and Polyanskaya in [15] and [6] demonstrated that the developmental changes in speech rhythm both in L1 and L2 are captured only by normalized rhythm metrics, while raw rhythm metrics either yield inconsistent results (due to conflicting factors that affect their values), or demonstrate no clear developmental tendency at all. Thus, only tempo-normalized rhythm metrics were calculated for V, C and S intervals. Although %V was not shown to differ between age groups (between 3 and 5 years) in L1 English or proficiency levels in L2 English by Ordin and Polyanskaya in [15] and [6], a contrary was shown by Payne et al. in [13] for child speech delivered by 2 and 4 year olds. Besides, %V was shown to differentiate adult and child speech, and we might expect significant drop in %V over a longer time between 4:7 and 11:7. Moreover, Wiget et al. [11] showed that %V is robust to tempo variation. Therefore, we decided to include %V into our analysis. To test whether speech rate changes in the course of L1 acquisition, mean durations of S (meanSyl), V (meanV) and C (meanC) have also been calculated. We averaged the scores across all sentences within each speaker. This yielded one mean metric score per speaker. This was done for the by-speaker analysis and building the regression models. A more commonly used by-utterance design does not allow exploring the development of rhythm as a function of age on a continuous scale.

3. Results

To study how rhythmic patterns change as a function of age in the course of L1 acquisition, we built regression models with *Age* as a predictor and *Metrics* (including means) as outcomes. In order to control for multiple error rates and to receive the simultaneous test for all regressions, we ran multivariate analysis prior to estimating the contribution of the predictor on separate outcomes. The analysis showed that *metrics* and *age* were significantly related, $V = .62$, $F(10, 41) = 6.695$, $p < .0005$, $\eta^2 = .62$, which justified building separate regression models for each metric (with corrected α level for multiple tests). The model parameters are presented in table 1. Power analysis showed that with the sample size of $N = 52$ and the desired statistical power level $\beta \geq .8$ (following the recommendation by Cohen in [23]), we could aim at detecting the effect size of $R^2 \geq .15$ or higher.

Table 1. *Parameters of the regression models.*

Metric	Parameters estimates			Model fit	
	β	$t(50)$	R2	$F(1,50)$	p
<i>meanSyl</i>	-.306	-2.272	.094	5.161	.027
<i>VarcoSyl</i>	.241	1.759	.058	3.094	.085
<i>nPVI-syl</i>	.3	2.221	.09	4.931	.031
<i>meanV</i>	-.274	-2.014	.075	4.054	.049
<i>VarcoV</i>	.509	4.179	.259	17.466	.0005
<i>nPVI-V</i>	.543	4.572	.295	20.908	.0005
<i>meanC</i>	-.431	-3.375	.186	11.391	.001
<i>VarcoC</i>	.293	2.168	.086	4.702	.035
<i>nPVI-C</i>	.239	1.74	.057	3.026	.088
<i>%V</i>	.061	.435	.004	.189	.666

The data show that *age* reveals the strongest relations with *VarcoV*, *nPVI-V* and *meanC* (figures 1-2). The R^2 for these

models reach the values needed for the pre-defined desired statistical power. %V, nPVI-C and VarcoSyl reveal absolutely no significant relations between *age* and *Metrics*. Increase in speech tempo in child speech is a cumulative effect of a substantial drop in mean durations of consonantal clusters and slight decrease in V durations. Despite the slight – albeit significant – decrease in V durations as a function of *age*, variation in duration of V intervals reveals a substantial growth, which indicates that some V intervals become shorter, and others become lengthened. This trend is evident both globally, i.e., at the scale of the whole utterance, and locally, i.e., pairwise. This indicates a substantial growth in stress-timing in the course of L1 acquisition by monolingual English children, and confirms our hypothesis, at least in regard to vocalic variability. Variation in C durations at the scale of the whole utterance and syllabic variability pairwise also significantly increase, but the slope is less steep than that for vocalic variability. This, again, confirms our hypothesis.

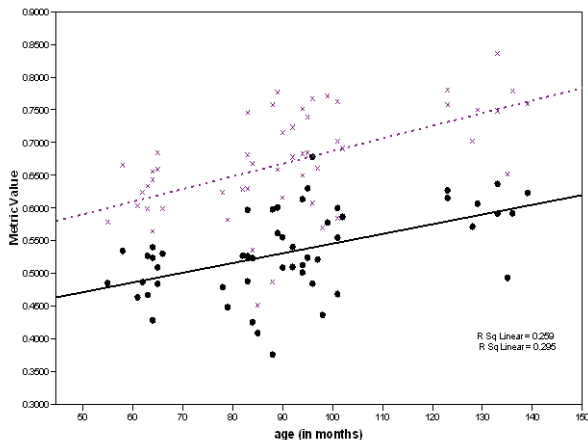


Figure 1: Changes in VarcoV (solid line and filled dots) and nPVI-V (dashed line and crosses) as a function of age.

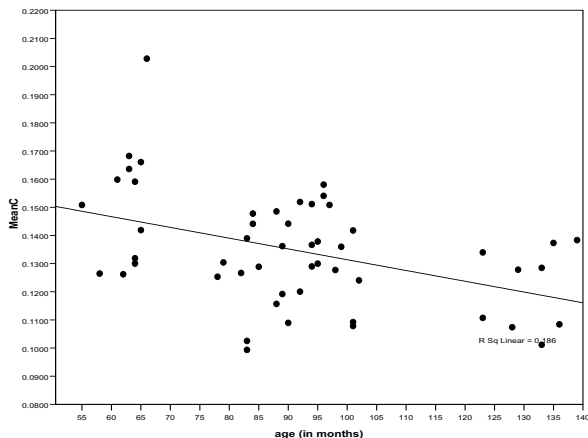


Figure 2: Changes in MeanC as a function of age.

4. Discussion

Rhythmic patterns in child L1 English indeed develop from more syllable-timed towards more stress-timed as a function of age as L1 acquisition progresses. The same developmental trend has been detected in adult L2 English spoken by learners with rhythmically contrastive languages (French and German). Ordin and Polyanskaya [22] used the procedure and prompts described in this study to elicit sentences from learners of

English at different proficiency levels whose native language was either rhythmically similar to (German), or different from (French) the target language (English). The results revealed that speech tempo and variation in V, C and S durations increase irrespective of the native language of the learner, but the French varied durations in L2 English less than Germans. Thus, although the developmental trend towards higher degree of stress-timing in L2 has been confirmed for native speakers of both languages, the actual rhythm in L2 at a certain proficiency level is also dependent on the native language of the learner. These results showed that rhythm in L2 is partly dependent on the native language of the learner, but the general development of rhythmic patterns is a universal tendency.

The universal nature of this tendency has also been confirmed by detecting the same trend in L1 acquisition, and also by corroborative evidence that diachronic changes of phonological parameters in Germanic (and many other Indo-European) languages drive the rhythmic changes towards higher degree of stress-timing [24].

Based on these data, we could tentatively agree with [19] and [20] that syllable-timing is a default setting, and stress-timed rhythm is marked and acquired later and with more difficulty. Even the speakers of other stress-timed languages cannot merely transfer the patterns of durational variability from native into target language and have to develop stress-timing in the course of L2 acquisition, similar to what children do in L1 acquisition. If the marked feature of L2 is also present in the native language of the learner, acquisition is facilitated, therefore German learners of English develop stress-timing to a greater degree and more rapidly than French learners of English [22].

This developmental tendency is very stable and detectable when researchers use different experimental paradigms and analysis methods. The increase in durational variability of speech intervals has been detected using a longitudinal approach and a within-subject analysis [15] as well as using a cross-sectional paradigm and a between-subject analysis [other studies]. Within the latter paradigm, it is possible to use a between-subject (this study and [13]) as well as between-utterance analysis [6]. The developmental tendency for higher degree of stress-timing in the course of L1 and L2 acquisition reveals itself when the speech material is strictly controlled for, as in this study and in [14], [6], [22], somewhat controlled for, as in [13], or not controlled for, as in [15]. It has been detected on read as well as on semi-spontaneous speech. Of course, depending on the method, the exact values of the rhythm metrics may differ significantly, and we are not saying that the method is irrelevant if the researcher needs to compare rhythmic patterns between speakers/proficiency levels/languages/ages/different stages of L1 and L2 acquisition. However, when the researcher is interested in developmental tendencies in acquisition of speech rhythm, the methodological issues become less important. The tendencies are strong, stable, and detectable using a variety of approaches to data collection and analysis.

5. Acknowledgements

The research leading to these results has been supported by the Alexander von Humboldt Foundation. We would also like to thank children, parents and teachers of William Wordsworth School, Paderborn, and St. David's Primary School, Ramstein, who participated or assisted the data collection.

6. References

- [1] R. Dauer, "Stress-timing and syllable-timing reanalyzed", *Journal of Phonetics*, vol. 11, 51-62, 1983.
- [2] R. Schiering, "The phonological basis of linguistic rhythm. Cross-linguistic data and diachronic interpretation", *Sprachtypologie und Universalienforschung*, vol. 60, 337-359, 2007.
- [3] P. Prieto, M. del Mar Vanrell, L. Astruc, E. Payne, and B. Post, "Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish", *Speech Communication*, vol. 54, 681-702, 2012.
- [4] A. Arvaniti, "The usefulness of metrics in the quantification of speech rhythm", *Journal of Phonetics*, vol. 40, 351-373, 2012.
- [5] F. Nolan and E.L. Asu, "The Pairwise Variability Index and coexisting rhythms in language", *Phonetica*, vol. 66, 64-77, 2009.
- [6] M. Ordin and L. Polyanskaya, "Perception of speech rhythm in second language: The case of rhythmically similar L1 and L2", *Frontiers in Psychology*, vol. 6, no. 613, 1-17, 2015.
- [7] F. Ramus and J. Mehler, "Language identification with suprasegmental cues: A study based on speech resynthesis", *Journal of the Acoustical Society of America*, vol. 105, 512-521, 1999.
- [8] F. Ramus, M. Nespors, and J. Mehler, "Correlates of linguistic rhythm in the speech signal", *Cognition*, vol. 73, 265-292, 1999.
- [9] E. Grabe and L. Low, "Durational Variability in Speech and the Rhythm Class Hypothesis" in C. Gussenhoven and N. Warner (Eds.) *Papers in Laboratory Phonology 7* New York: Mouton de Gruyter, 515-546, 2002.
- [10] V. Dellwo, "Rhythm and Speech Rate: A Variation Coefficient for deltaC", in P. Karnowski, I. Szigeti (Eds.). *Language and language-processing*, Frankfurt am Main: Peter Lang, 231-241, 2006.
- [11] L. Wiget, L. White, B. Schuppler, I. Grenon, O. Rauch, and S. Mattys, "How stable are acoustic metrics of contrastive speech rhythm?", *Journal of Acoustical Society of America*, vol. 127, 1559-1569, 2010.
- [12] A. Loukina, B. Rosner, G. Kochanski, E. Keane, and C. Shih, "What determines duration-based rhythm measures: text or speaker?", *Laboratory Phonology*, vol. 4, 259-553, 2013.
- [13] E. Payne, B. Post, L. Astruc, P. Prieto, and M. del Mar Varnell, "Measuring child speech", *Language and Speech*, vol. 55, 203-229, 2012.
- [14] F. Bunta and D. Ingram, "The acquisition of speech rhythm by bilingual Spanish- and English-speaking four- and five-year-old children", *Journal of Speech, Language, and Hearing Research*, vol. 50, 999-1014, 2007.
- [15] M. Ordin and L. Polyanskaya, "Development of timing patterns in first and second languages", *System*, vol. 42, 244-257, 2014.
- [16] L. White and S. Mattys, "Calibrating rhythm: First language and second language studies", *Journal of Phonetics*, vol. 35, 501-522, 2007.
- [17] A. Li and B. Post, "L2 acquisition of prosodic properties of speech rhythm", *Studies in Second Language Acquisition*, vol. 36, 223-255, 2014.
- [18] V. Stockmal, D. Markus, and D. Bond, "Measures of native and non-native rhythm in a quantity Language", *Language and Speech*, vol. 48, 55-63, 2005.
- [19] G. D. Allen and S. Hawkins "The development of phonological rhythm", in A. Bell and J. Hooper (Eds.). *Syllables and Segment*, Amsterdam: North-Holland, 173-185, 1978.
- [20] E. Grabe, B. Post, and I. Watson, "The acquisition of rhythmic patterns in English and French", in *Proceedings of the International Congress of Phonetic Sciences, San Francisco, USA, Proceedings, 1999*, pp. 1201-1204.
- [21] N. Whitworth, "Speech rhythm production in three German-English bilingual families", *Leeds Working Papers in Linguistics & Phonetics*, vol. 9, 175-205, 2002.
- [22] M. Ordin and L. Polyanskaya, "Acquisition of speech rhythm in a second language by learners with rhythmically different native languages", *Journal of the Acoustical Society of America*, under review and still waiting for evaluation reports, submitted on 22.02.2014.
- [23] J. Cohen, "Statistical power analysis for the behavioral sciences" (2nd ed.). Hillsdale, NJ: Erlbaum, 1988.
- [24] M. Ordin, L. Polyanskaya, and P. Wagner, "Development of timing patterns in second language: A diachronic perspective", in A. Leemann, M.-J. Kolly, S. Schmid, and V. Dellwo (eds.), *Trends in Phonetics and Phonology. Studies from German speaking Europe*. Frankfurt am Main / Bern: Lang, 331-347, 2015.