Acquisition of Timing Patterns in Second Language

Mikhail Ordin[1,2], Leona Polyanskaya[1], Christiane Ulbrich[2]

[1]Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld, Deutschland
[2]School of Communication, University of Ulster, UK

mikhail.ordin@gmail.com

Abstract

The paper presents an analysis of speech rhythm development in second language. 51 German learners of English with varying degrees of proficiency were recorded producing 33 identical sentences of quasi-spontaneous speech. Durational characteristics of syllables, consonantal and vocalic intervals were calculated to allow for analysis of timing patterns on different proficiency levels. We found that 1) durational characteristics in second language depend on the proficiency level of the speaker and can be modeled to predict the proficiency level of the language learner; 2) durational characteristics become more consistent throughout the progress of second language mastery; 3) multiple rhythms operate on multiple timescales.

Index Terms: speech rhythm, second language acquisition, timing differences, rhythm metrics, duration.

10.21437/Interspeech.2011-336

1. Introduction

In studies on second language (L2) acquisition we are interested in how languages differ and how L2 learners acquire these differences. The languages might have systemic and realizational differences [1], [2], [3]. Systemic differences are the differences in linguistic systems, e.g. in inventory of phonological structures and primitives. The differences between phonetic realizations of identical phonological primitives from two linguistic inventories (i.e. languages or language varieties) are referred to as realizational differences.

L2 learners need to acquire the L2 language-specific realizations of the phonological structures which sometimes involves the acquisition of new structures and/or structural elements. Thus the L2 learners system, commonly referred to as interlanguage, may differ from the target first language (L1) both systemically and realizationally.

One of the areas in which languages differ is speech rhythm. Rhythm can be understood as the recurrence of certain events in time, thus timing is fundamental to the speech rhythm. The purpose of the presented study is an analysis of how L2 learners acquire such timing characteristics that may be different between their L1 and their L2. As the timing differences contribute to building language-specific rhythm patterns, we will provide new insights into the development of speech rhythm production in the process of L2 mastery.

The present paper focuses on realisational differences in an L2 that is systemically not different from the L1 of the learner, i.e. German L1 and English L2. Both German and English are assumed to be stress-timed languages (as oppose to syllable-timed languages such as French or Italian). We will track the development of the realization of language-specific timing patterns in English L2 produced by native German speakers with varying proficiency levels in English.

Much effort has been devoted to the understanding of how speech rhythm in general and timing patterns in particular are acquired by L2 learners [9, 15, 16, 17]. These studies, however, compare target productions represented by monolingual adult speakers of the target language with L2 learners' productions. Comparing 'deviant' with 'target' productions sometimes fails to enable us to understand the acquisitional process since L2 learners of the same language may still have different targets. The targets can be represented, for example, by non-native teachers with varying degrees of foreign accent or by specific conditions in the environments in which L2 is acquired, e.g. varying regional or social factors. As exposure and environment create complex and mixed categories of input [4], and these categories may differ for L2 learners even with the same L1 background, there is little point in comparing productions of an adult monolingual L1 speaker with productions of L2 learners. Instead, it seems commendable to consider the interlanguage of L2 learners as the system in its own right. Thus the present paper reports findings of an analysis of the *developmental process* through the comparison of acquirers' productions of varying proficiency levels.

A number of rhythm metrics have been proposed to capture language-specific temporal rhythmic properties, or timing differences, e.g. pairwise variability index (PVI) [5], ΔV, ΔC and %V [6], VarcoV and VarcoC [7].

ΔV and ΔC are standard deviations of vocalic and consonantal durations within the utterance (two or more successive consonants are united into one consonantal interval). ΔC is thought to be indicative of the syllabic structure, syllable complexity and consonantal phonotactic constraints whereas ΔV is supposed to account for the degree of vowel reduction.

VarcoV and VarcoC are the standard deviation of consonantal (C) and vocalic (V) interval durations divided by mean consonantal or vocalic duration within the utterance. They reflect the same properties as ΔC and ΔV do, but Varco measures supposedly neutralize the effect of tempo differences, and thereby reduce the effect of idiosyncrasies in the rate of speech production.

%V is the proportion of vocalic intervals duration (in percent) to the duration of the utterance. It indicates the syllabic structure.

rPVI is a raw pairwise variability index, i.e. quantitative measure in local fluctuations in duration. It expresses variability of duration in pairs of successive intervals and is calculated as:

$$rPVI = 100 \times \left[\frac{\sum_{k=2}^{n} |d_k - d_{k-1}|}{n-1} \right] \quad (1),$$

In the formula in (1), n is the number of intervals in the sentence, d is the duration of the k^{th} interval, and k is the serial number of the interval.

As raw PVI is largely influenced by speech rate, normalized PVI was suggested to compensate for speech rate variations [8]:

Copyright © 2011 ISCA

$$nPVI = 100 \times \left[\sum_{k=2}^{n} \left| \frac{d_k - d_{k-1}}{(d_k + d_{k-1})/2} \right| / (n-1) \right] \quad (2).$$

Dividing each pair difference in duration by the mean interval duration in this pair is a normalization factor, which is introduced to compensate the idiosyncratic speech rate differences, thus this index is thought to be more robust and less influenced by tempo differences between speakers and utterances.

Higher values of %V and lower values of other metrics correspond to the auditory impression of "syllable-timing" [5], [6], [7], [8], [9].

These metrics were used to investigate both cross-linguistic differences in rhythm as well as acquisition of the speech rhythm by monolingual and bilingual speakers in various ages [8], [9], [10], [11], [12], [13], [14], [15], [16], [17]. The general finding is that although people can easily evaluate cross-linguistic timing differences in speech perception, they cannot easily imitate them in speech production. Adult L2 learners can hear fine differences in the realization of the timing patterns, but they are hardly capable of producing them.

Timing differences have been widely investigated using these rhythm metrics. Dies et al. [15] showed that durational characteristics of vocalic intervals differ significantly between L2 and L1 productions. The authors also found significant and reliable changes in the values of the rhythm metrics as a result of acquisition progress (i.e. language mastery). Ordin and Setter [16] showed that there is a significant positive correlation between the nPVI values obtained in utterances of native Russian L2 learners of English and the assessment results of the learners' productions by experienced pronunciation teachers. Ordin and Setter also showed that Hong Kong and Russian learners of L2 English exhibited similarities in syllabic timing patterns in their L2, although Russian is a stress-timed and Cantonese is a syllable-timed language [17]. White and Mattys [9] showed that L2 productions were generally more syllable-timed than productions of native speakers of a target languages, when both L1 and L2 belong to the stress-timed rhythm class.

Based on the results of these previous studies, we expect to find significant differences between the values of the rhythm metrics in utterances produced by German learners of L2 English depending on the levels of language mastery. We expect that the metrics will indicate rhythm development from more syllable-timed to more stress-timed in individual stages of the interlanguage progression. As the previous studies were mostly carried out on read speech, our study will be based on spontaneous speech in order to test and confirm the above hypotheses.

2. Method

For the purposes of the research we recorded 51 German learners of L2 standard British English. The subjects were all native German speakers from families with monolingual German-speaking parents. All participants grew up in Nord-Rhein Westfalia in or near the city of Bielefeld. That area is assumed to be void of regional accents and dialects and the spoken language resembles closely what is understood as a Northern Standard Variety of German (Hochdeutsch). Piske, McKay and Flege [20] analyzed a range of factors which affect the strength of perceived foreign accent in L2. They singled out the factors which have influence on the way a person speaks L2, including factors such as age of arrival, length of residence, exposure to L1, gender, formal instruction, motivation, language learning aptitude, etc. These factors were controlled for to base the analysis on a homogeneous group of homogeneous speakers who varied only in the degree of L2 mastery. These possible influences were controlled for by collecting the relevant information in a short language background questionnaire and verifying it in informal interviews prior to the recording sessions.

To analyse timing patterns using traditional rhythm metrics, it is important that the speech samples are recorded in laboratory conditions, and that speech samples from L2 speakers of different proficiency levels are balanced by content. The traditional rhythm metrics are influenced by the nature of the speech materials, e.g. number of vocalic and consonantal intervals, number of syllables in polysyllabic words, syntactic structures, etc. [21]. Care was taken to elicit the comparable utterances from the speakers. The recordings were made in a sound-treated booth of the audio-visual studio at the University of Bielefeld. To make the recordings of quasi-spontaneous yet comparable utterances, we used elicitation tasks as detailed in [14] with the help of 33 picture prompts for sentence elicitation. The recordings were made in WAV PCM format at 44 kHz, 16 bit in mono.

Prior to the sentence elicitation task, the participants took a pronunciation test and were interviewed (interviews were recorded and the participants were aware of being recorded at all times). Each interview was later assessed by two experienced English language tutors for fluency, vocabulary resources used and grammatical accuracy. The assessments and the results of the pronunciation tests were used to allocate the speakers to three groups of proficiency levels: lower-intermediate (12 speakers), upper-intermediate (9 speakers) and advanced (22 speakers). Eight speakers were not attributed to any group because the tutors did not agree with each other in their assessment and/or because of the discrepancy between the results of the pronunciation tests and the tutor's assessment of language skills (i.e. pronunciation competence and competence in other language areas differed). These speakers were excluded from further analysis.

33 elicited quasi-spontaneous yet identical sentences per speaker were annotated in Praat [22]. Each sentence was divided into vocalic and consonantal intervals and into syllables. The segmentation was carried out manually based on the criteria outlined in [23] and [24] for vocalic and consonantal intervals and in [25] and [26] for syllables.

Traditional rhythm metrics were calculated on the extracted durations of consonantal and vocalic intervals for each sentence per speaker, including raw and normalized PVI for vocalic and consonantal intervals and for syllables (rPVI-v, rPVI-c, rPVI-s, nPVI-v, nPVI-c, nPVI-s), ΔV, ΔC and standard deviation of the syllable duration (ΔS), %V, VarcoV, VarcoC, VarcoS, mean duration of syllables (meanS), vocalic (meanV) and consonantal (meanC) intervals in each sentence. The values of the rhythm metrics were statistically analysed to investigate the differences in rhythm metrics between proficiency groups; and to understand how and if the metrics can predict the proficiency level of the L2 learner.

3. Results

Descriptive statistic data is presented in the table 1.

A one-way between-subject MANOVA was performed to investigate the differences in rhythm metrics between proficiency groups of German learners of L2 English. The values of metrics were used as dependent variables and the proficiency level of the learners was introduced as a factor.

Preliminary data screening did not reveal serious violations of assumptions for MANOVA analysis.

	Lower-inter		Upper-inter		Advanced	
	mean	**st. dev**	**mean**	**st. dev**	**mean**	**st. dev**
MeanS	.227	.045	.217	.043	.203	.037
MeanV	.113	.027	.106	.026	.101	.024
MeanC	.118	.028	.115	.025	.110	.024
ΔS	.108	.035	.109	.033	.111	.031
ΔV	.060	.023	.062	.025	.060	.024
ΔC	.058	.025	.057	.019	.059	.018
VarcoS	.477	.116	.504	.114	.546	.107
Varco V	.531	.151	.575	.160	.592	.152
Varco C	.489	.147	.496	.141	.544	.138
%V	45.8	6.75	44.9	7.08	44.8	6.75
rPVI-s	.144	.052	.148	.051	.145	.049
rPVI-v	.082	.033	.085	.036	.083	.035
rPVI-c	.072	.033	.071	.029	.074	.027
nPVI-s	.634	.177	.685	.190	.710	.175
nPVI-v	.683	.200	.746	.210	.777	.212
nPVI-c	.585	.187	.613	.207	.668	.187

Table 1. *Mean values for the analysed rhythm metrics in lower-intermediate (N=396), Upper-intermediate (N=297) and advanced (N=726) levels.*

We found a statistically significant and substantial effect of the proficiency level on the values of the rhythm metrics, Wilk's lambda (Λ) = .754, F(46, 2788) = 9.591, p <. 0005, eta^2 = .137, power (β) = 1.0

However, when the results of the dependent variables were considered separately, univariate ANOVAs revealed non-significant effect of syllabic, consonantal and vocalic rPVI and of ΔV, ΔC and ΔS.

As it is not recommended to use more than 10 variables in MAVOVA [27], we performed another one-way between-subject multivariate analysis of variance with the proficiency level of the L2 learners as factor. Only those rhythm metrics which were found to differ significantly between the proficiency levels were introduced into this analysis as dependent variables. The model was still significant, Λ = .813, F(20,2814) = 15.3, p < .0005, eta^2 = .1, β = 1.0

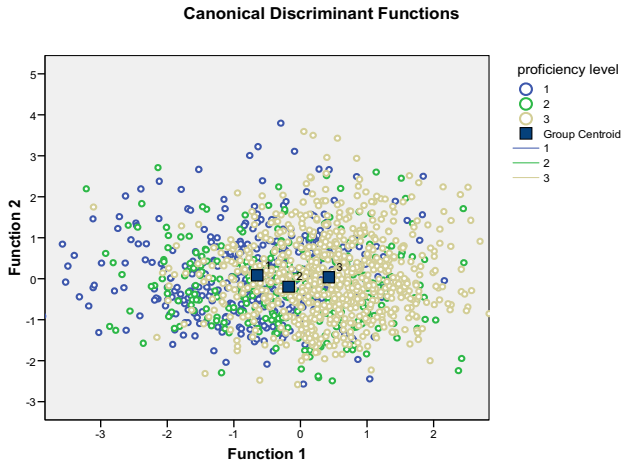

Figure 1. *Discriminant function plot with the centroid values for each proficiency level.*

The latest MANOVA was followed up with discriminant analysis which revealed two discriminant functions. The first function explained 95.5% of variance, canonical R^2 = 0.178, and the second explained only 4.5% of variance, R^2 = 0.01. In combination these functions significantly differentiated the proficiency levels, Λ = .813, $\chi^2(20)$ = 291.4, p < 0.0005. The second function alone did not significantly differentiate the proficiency levels, Λ = .99, $\chi^2(9)$ = 14.434, p = .108. The correlations between the outcomes and discriminant functions revealed that the rhythm metrics calculated on syllabic durations loaded on the first function, and the rhythm metrics calculated on durations of vocalic and consonantal intervals loaded more highly on the second function. As the first function explains substantially more variance that the second function, we conclude that the metrics calculated on syllable durations discriminate between the proficiency levels much better than the metrics calculated on durations of vocalic and consonantal intervals. This can also be seen on the discriminant function plot (figure 1).

4. Discussion

The statistic analysis showed mean durations of vocalic and consonantal intervals and syllables decrease as L2 mastery increases. It can be explained by the fact that speech rate increases with L2 proficiency of the learners.

ΔS, ΔV, ΔC, syllabic, consonantal and vocalic rPVI are dependent on speech rate. Due to the possible idiosyncratic variations in speech rate and due to the speech rate increase with growing mastery in L2, we expected these metrics to be less influenced by the proficiency level than the normalized versions of these metrics, i.e. Varco coefficients and nPVI. This prediction was confirmed with inferential statistic analysis. Varco coefficients and nPVIs significantly increase with proficiency level, while the change of ΔS, ΔV, ΔC, raw PVI metrics was not significant. This indicates that the productions of the advanced speakers are more stress-timed than the productions of identical sentences by lower-intermediate and intermediate speakers, when the effect of speech rate is controlled for.

Inspection of the data in table 1 also reveals that standard deviations of the means of the metrics tend to decrease with the L2 mastery. This might indicate that the timing patterns become more stable and consistent as a result of acquisition progress. Productions of lower-intermediate speakers varied greatly between stress-timed and syllable-timed, but productions of advanced learners were more consistently stress-timed as illustrated in figure 1 (the discriminant function plot shows that the variate scores for the advanced learners are more compact, while the variate scores for the lower-intermediate learners are spread more evenly along the first discriminant function). This indicates, realizations of timing differences become more systematic and consistent as L2 mastery progresses.

Nolan and Asu [28] demonstrated that PVI can be meaningfully applied to various intervals, e.g. syllables, vocalic and consonantal intervals within the utterance, successive feet, etc. It captures the average variability in duration of the analysed intervals. The authors assumed "coexisting rhythms in language". They claime that syllable-timing and stress-timing (or foot-timing) can be independent dimensions of temporal organization in languages rather than opposite ends of a continuum. Syllable-timing operates on the syllable level and foot-timing operates on the a larger timescale, i.e. on a foot level. This idea of coexisting rhythms agrees with the idea of multiple rhythmic systems which are associated with multiple timescales and coordinate prosodic events on multiple timescales [29].

The results of the discriminant function analysis presented here show that the metrics are grouped into two underlying functions. The functions discriminate between the proficiency levels of the L2 learners. The metrics calculated on syllable

durations form the first discriminant function. The metrics related to durational characteristics of vocalic and consonantal intervals form the second function. The first function discriminates between the proficiency levels much better than the second (the second function explained only 4.5% of the variance). In light of the theory of multiple rhythms, we can conclude that with progressing language mastery L2 speech of Germans learners of L2 English becomes less syllable-timed. That is why the first discriminant function (which combines the metrics calculated on syllabic durations) is efficient in discriminating proficiency levels. As stress-timing seems to be less affected by acquisition progresses, the second discriminant function is less powerful in predicting the proficiency level of the speakers. This may be explained by the fact that German and English are close to each other in the dimension of stress-timing, and interlanguage varieties of German learners of L2 English do not differ in stress-timing.

If this is true, than we might expect that the proficiency levels of, for example, French learners of English, will be predicted with the discriminant functions which are made up of syllabic durational characteristics and with the functions that are made up of vocalic/consonantal durational characteristics (English and French should differ in both stress-timing dimension and syllable-timing dimension, and French learners of English will have to acquire interval timing and syllabic timing). We will test this hypothesis in further research on L1 and L2 rhythm acquisition.

Future research will also investigate the coordination between dimensions, i.e. coordination between rhythms operating on different timescales: the timescale of syllables and the timescale of vocalic and consonantal intervals. In addition we plan to investigate rhythms operating on wider timescales, i.e. feet, phrases and maybe even speech turns, and coordination of rhythms on multiple timescales. In the next phase of the project we will carry out similar research with French learners of L2 English. French and English have different rhythm patterns and if there are systemic differences in timing patterns between linguistic systems (not only realizational differences), we expect to find them in Interlanguage of French learners of L2 English on different proficiency levels.

The results of the current and future studies on rhythm development have theoretical and practical implications. Positive correlation between the results of the standardized pronunciation test and rhythm metrics (excluding ΔS, ΔV, ΔC, raw PVI-s and means, meanV, meanC) evidence that the rhythm metrics can imply the level of development of pronunciation skills. Officially recognized language proficiency tests (e.g., IELTS and TOEFL) include speaking modules to assess learners' abilities in areas of pronunciation, depth of vocabulary, fluency and accuracy. It needs to be emphasised that the scores pertaining to the pronunciation of language learners are largely subjective. The examiner lacks precise criteria how to assess pronunciation. The imprecision leads to inter-raters and intra-raters inconsistencies in pronunciation assessment [30, 31]. Expected correlations between rhythm metrics of the French learners of L2 English will conform the expected impact of the research in L2 testing. The expected results could be used to make pronunciation assessment slightly more objective.

The results can also be used to distinguish between proficiency levels of L2 speakers within the Common European Framework of Reference for Languages (CEFR). CEFR includes 6 levels of language proficiency (from A1 to C2), and each level is characterized by certain descriptors of pragmatic and linguistic competence [32]. The descriptors of the phonological control exist only for the levels C1 and B2, and even these are imprecise. Further work on L2 rhythm and correlation between automatically obtained assessment scores and assessment scores by human examiners might feed into the work on CEFR and design of precise descriptors for fluency and phonological control.

5. References

[1] Wells, J., "Accents of English: volume 1", OUP, 1982.

[2] McMahon, A., "An Introduction to English Phonology', Edinburgh University Press, 2001.

[3] Hughes, A., Trudgill, P., Watt, D., "English Accents and Dialects", Hodder Education, 2005.

[4] Fernald, A., "When infants hear two languages: Interpreting research on early speech perception by bilingual children", in P. McCardle & E. Hoff [Eds.], Childhood Bilingualism: Research on Infancy through School-Age, 2006.

[5] Grabe, E., Low, L., "Acoustic correlates of rhythm class", in C. Gussenhoven & N. Warner [Eds.], Laboratory Phonology 7, 515-546, New York: Mouton de Gruyter, 2002.

[6] Ramus, F., Nespor, M., Mehler, J., „Correlates of linguistic rhythm in the speech signal", Cognition, 73(3): 265-292, 1999.

[7] Dellwo, V., Wagner, P., "Relationships between Rhythm and Speech Rate", In Proceedings of the 15th ICPhS, 471-474, 2003.

[8] Low, L., Grabe, E., Nolan, F., "Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English", Language and Speech, 43(4): 377-401, 2000.

[9] White, L., Mattys, S., "Calibrating rhythm: First language and second language studies", Journal of Phonetics, 35(4): 501-522, 2007.

[10] Grabe, E., Post, B., Watson, I., "The acquisition of rhythmic patterns in English and French", In Proceedings of the International Congress of Phonetic Sciences, San Francisco, 1999.

[11] Grabe, E., Gut, U., Post, B., & Watson, I., „The acquisition of rhythm in English, French and German", In Current Research in Language and Communication: Proceedings of the Child, 1999.

[12] Whitworth, N., "Speech rhythm production in three German-English bilingual families", Leeds Working Papers in Linguistics & Phonetics, 9: 175-205, 2002.

[13] Work, R., Andruski, J., Casielles, E., Kim, S., "Paired variability indices in assessing speech rhythm in Spanish/English bilingual language acquisition", JASA, 117(4): 2426-2426, 2005.

[14] Bunta, F., Ingram, D., "The acquisition of speech rhythm by bilingual Spanish- and English-speaking four-and five-year-old children", Journal of Speech, Language, and Hearing Research, 50: 999-1014, 2007.

[15] Diez, F., Dellwo, V., Gavalda, N., "The development of measurable speech rhythm during second language acquisition", JASA, 123 (5):3886-3886, 2008.

[16] Grenon, I., White, L., "Acquiring rhythm: A comparison of L1 and L2 speakers of Canadian English and Japanese", In Proceedings of the 32nd Boston University Conference on Language Development, 155-166, 2008.

[17] Shport, I., "Acquisition of rhythm: Evidence from spontaneous L2 speech", JASA 123(5): 3328-3328, 2008.

[18] Ordin, M., Setter, J., "Objective indicators of rhythmic Russian–English transfer", XX Session of the Russian acoustical society, 649-652, 2008.

[19] Ordin, M., Setter, J., "Comparative research of temporal organization of the syllable structure in Hong Kong English, Russian English and British English", XX Session of the Russian acoustical society, 653-656, 2008.

[20] Piske, T., MacKay, I., Flege, J., „Factors affecting degree of foreign accent in an L2: A review", Journal of Phonetics, 29: 191-215, 2001.

[21] Arvaniti, A., "Rhythm, timing and the timing of rhythm", Phonetica, 66: 46-63, 2009.

[22] Boersma, P., Weenink, D., "Praat: doing phonetics by computer" (Version 5.1.22) Retrieved December 15, 2010, from http://www.praat.org/

[23] Peterson, G., Lehiste, I., "Duration of syllable nuclei m. Enghsh", JASA 32: 693-703, 1960.

[24] Stevens, K., "Toward a model for lexical access based on acoustic landmarks and distinctive features", JASA111(4): 1872-1891, 2002.

[25] Wells, J., "Longman Pronunciation Dictionary", Longman, 2008.

[26] Harris, J., "English sound structure", Blackwell, 1994.

[27] Stevens, J., "Power of the multivariate analysis of variance tests", Psychological Bulletin 88: 728-737, 1980.

[28] Nolan, F., Asu, E. "The pairwise variability index and coexisting rhythms in language", Phonetica, 66(1-2): 64-77, 2009.

[29] Tilsen, S. "Multiscale dynamical interactions between speech rhythm and gesture", Cognitive Science, 33: 839-879, 2009.

[30] Taylor, L., Falvey, P., "IELTS Collected Papers: Research in speaking and writing assessment", Cambridge: CUP, 2007.

[31] Morrow, K., "Insights from the Common European Framework", Oxford: OUP, 2004.

[32] Council of Europe, "Common European Framework of Reference for Languages: Learning, Teaching, Assessment", Cambridge: CUP, 2001.