Sönning, Lukas

(Re-)viewing the Acquisition of Rhythm in the Light of L2 Phonological Theories

Date of secondary publication: 10.04.2025

Accepted Manuscript (Postprint), Bookpart

Persistent identifier: urn:nbn:de:bvb:473-irb-1074563

Primary publication

Sönning, Lukas (2023): (Re-)viewing the Acquisition of Rhythm in the Light of L2 Phonological Theories, in: Robert Fuchs (Ed.), Speech rhythm in learner and second language varieties of English, Singapore: Springer, pp. 123–157, doi: 10.1007/978-981-19-8940-7_6.

(Re-)viewing the Acquisition of Rhythm in the Light of L2 Phonological Theories

Lukas Sönning

Abstract Previous work on non-native speech rhythm has often drawn on L2 phonological theory for the interpretation of findings. The explicit confrontation of theory-derived hypotheses with data remains scarce, however. This paper illustrates how a hypothetico-deductive approach can contribute to our understanding of L2 speech rhythm. We consider cross-sectional data on prominence alternations in German learner speech from the viewpoint of two dynamic frameworks: The Ontogeny Phylogeny Model (OPM) and the Linguistic Theory of L2 Phonological Development (LTD). While both theories deal with L1-independent, universal forces in L2 acquisition, the OPM further considers the role of L1 transfer, similarity, and markedness. The predictions we formulate based on the two models lead us to pursue distinct methodological strategies. While our reading of the OPM prompts us to measure speech rhythm as a single, global category of speech, the LTD suggests a more nuanced, componential approach to L2 rhythm. Our application of the OPM confronts us squarely with the limited utility of rhythm metrics for L2 speech research and points to a number of issues at the theory-data interface. Overall, the LTD generates more informative predictions and provides a richer framework for the empirical study of prominence grading in L2 speech.

Keywords Speech rhythm · German Learner English · Rhythm metrics · Vowel duration · Vowel reduction · L2 phonology · L2 acquisition · Theory · Interlanguage development

I wish to thank the editor, two anonymous reviewers, and Ole Schützler for helpful comments and attention to detail.

L. Sönning (✉)
Institute of English and American Studies, University of Bamberg, An Der Universität 9, 96047 Bamberg, Germany
e-mail: lukas.soenning@uni-bamberg.de

1 Introduction

Current empirical research on speech rhythm in second language (L2) acquisition can be described as predominantly descriptive in the sense that data analysis is not explicitly guided by theories of L2 phonological acquisition. While previous work has frequently turned to theoretical frameworks for a post-hoc interpretation or explanation of findings (e.g. Li & Post, 2014; Ordin & Polyanskaya, 2014), this contrasts with the approach taken in analytical studies, which rely on theory to generate hypotheses and then confront these with data (see, e.g. Colantoni et al., 2015: 31). The aim of this paper is to illustrate how analytical approaches can contribute to our understanding of the L2 acquisition of speech rhythm.

While it has been argued that existing models of L2 phonology are ill-suited to account for prosodic phenomena (Li & Post, 2014: 224), this view may require qualification: Among others, the Ontogeny Phylogeny Model (OPM; Major, 2001) and the Linguistic Theory of L2 Phonological Development (LTD; James 1988) provide rich frameworks for studying prominence alternations in speech. This paper aims to illustrate that the OPM and LTD not only encourage new methodological approaches beyond the well-trodden path of rhythm metrics but also constitute unified frameworks for the (contrastive) study of speech rhythm across different varieties of English.

After a survey of theoretical and empirical approaches to speech rhythm (Sect. 2), Sect. 3 offers a contrastive analysis of prosodic properties of English and German. In Sect. 4, the central tenets of the OPM and LTD are outlined. Section 5 then presents our case study, a set of cross-sectional data on German Learner English (GLE). Following this, both the models are applied to the development of speech rhythm in GLE. To this end, theoretical assumptions of the OPM (Sect. 6) and LTD (Sect. 7) are translated into predictions about timing patterns in learner speech, which are then compared to empirical data. Sect. 8 closes with a general summary and discussion, which recapitulate methodological and theoretical implications for research on L2 speech rhythm.

2 Speech Rhythm Research

2.1 Theoretical Approaches

Over the past 80 years, the notion of speech rhythm has been approached from different perspectives. In the following, three views will be discussed: The isochrony view, the phonological view, and the prosodic view.

One of the earliest conceptualizations is based on the notion of isochrony and focuses on the presumed temporal regularity of prominent units (James & Arthur, 1940; Pike, 1945; Abercrombie, 1967). In this traditional *isochrony view* of rhythm, languages fall into two broad classes, ‘stress-timed’ and ‘syllable-timed’. It is assumed that in 'stress-timed' languages, a perceived regularity applies to the duration of feet (i.e. inter-stress intervals) while in 'syllable-timed' languages it is syllables that tend to be of equal duration. Acoustic studies have not corroborated the existence of such isochronous patterns in spoken language (e.g. Bolinger, 1965; Roach, 1982 for English; Borzone de Manrique & Signorini, 1983 for Spanish; Wenk & Wiolland, 1982 for French). Nevertheless, a number of perceptual studies have reported that listeners are able to discriminate between languages traditionally assigned to different classes (e.g. Nazzi et al., 1998; Ramus et al. 2003; but also see White et al., 2012).

In the absence of empirical evidence for the durational equalization of feet or syllables, the notion of isochrony has given way to alternative accounts. In what is commonly referred to as the *phonological view* of rhythm, Dauer (1983) proposed that rhythmic differences between languages reflect a number of lower-level phonological properties. These include syllable structure, length as a distinctive feature in vowels, and the (non-)existence of vowel reduction. The percept of different rhythm classes, it is argued, results from a combination of phonetic and phonological properties. In general, stress-timed languages show vowel reduction in unstressed syllables and greater phonotactic complexity, thus permitting a larger variety of onset and coda clusters.

Recent endeavors have extended this componential view of rhythm to include structural properties at the prosodic level of representation (e.g. Prieto et al., 2012; White, 2014; White et al., 2012). In this *prosodic view,* a focus has been on the durational marking of prosodic heads and edges, that is, local lengthening effects that are observable in prominent elements and at the boundaries of intonation phrases. In general, accented and phrase-final syllables are lengthened relative to unaccented and non-final syllables, respectively. It has been argued that the degree of prosodic length marking contributes to perceived differences between rhythm classes. Prieto et al. (2012), for instance, noted that the lengthening effect in accented and final syllables is much greater in English than in Spanish or Catalan. Utterance-final lengthening has also been observed to affect perceptual discrimination between languages such as English and Spanish (White et al., 2012).

In summary, our current understanding of speech rhythm suggests that the labels 'stress-timed' and 'syllable-timed'[1] may be considered as cover terms for a range of phonological and prosodic properties, or components, that are shared by languages traditionally assigned to the same end of the continuum. The componential perspective puts forward a set of features for our contrastive analysis of English and German in Sect. 3. First, however, let us turn our attention to instrumental approaches to speech rhythm, that is, different attempts by researchers to quantify the rhythmic properties of speech.

[1] While the terms 'stress-timing' and 'syllable-timing' are not descriptively adequate, they may be considered, at a general level of classification and comparison, a useful shorthand description. In the interest of simplicity, these labels will be used to refer to rhythm prototypes whose phonological and prosodic profiles are characteristic of language varieties that have been traditionally assigned to these rhythm classes. I will use single quotation marks as a signal to distance myself from the original, literal meaning of these terms.

2.2 Empirical Approaches

Recent empirical work on speech rhythm relies quite strongly on the application of rhythm metrics to measure rhythmic characteristics of speech (see Fuchs, 2016, Chap. 3, for a comprehensive overview and discussion). These measures build on insights gained from a componential view of rhythm and condense the degree of prominence variation in an utterance into a single score. This score expresses differences on a continuous scale and thereby offers a more fine-grained description than a binary distinction between 'stress-timed' and 'syllable-timed'. In general, rhythm metrics differ along the following lines:

(i) *Focal acoustic correlate*: While rhythm metrics can be applied to various acoustic correlates of prominence such as intensity and fundamental frequency, most work has so far relied on durational measurements.
(ii) *Unit of analysis*: In order for rhythm metrics to be applied, speech must be segmented into units, which then form the basis for analysis. These units can be vocalic and consonantal intervals or higher-level structures such as syllables or feet.
(iii) *Level of comparison*: To assess variation in prominence, comparisons can be made locally, that is, between adjacent units, or globally, across all units of analysis. Only the former level of comparison takes into account the linear arrangement of units.
(iv) *Quantification*: At the *local level*, units of analysis can be compared in *absolute terms*, where the focus is on absolute differences (e.g. a difference of 50 ms), or in *relative terms*, where relative differences are used (e.g. the duration of two units differs by a factor of 1.2, or by 20 percent). At the *global level*, metrics can be subdivided into *dispersion* measures, which quantify prominence variation in a batch of units (e.g. expressed as a standard deviation) and *proportion* measures, which document the share of certain unit types in the utterance.

Members of the family of rhythm metrics arise from different combinations of these attributes. To illustrate, let us briefly discuss some commonly used scores, which rely on durational measurements of vocalic intervals. Global measurements include the proportion of vocalic intervals (%V) in an utterance and the standard deviation of vocalic interval durations, originally as a raw (ΔV, Ramus et al., 1999) and now usually as a rate-normalized measure (VarcoV, Dellwo & Wagner, 2003). Low %V values and high VarcoV values indicate a high degree of vowel reduction and/or accentual lengthening (i.e. 'stress-timing' properties). The group of local measures, which rely on differences between successive interval durations, includes Low and Grabe's (1995) Pairwise Variability Index (PVI), which is an average of the absolute differences of successive intervals. Thus, a higher degree of temporal variability is reflected in higher PVI values. A normalized version, the nPVI, was proposed by Low et al. (2000) to adjust for differences in speech rate. Table 1 gives a summary of the nPVI-V, %V, and VarcoV. We will encounter these metrics again in the next section, which offers a contrastive analysis of the rhythm profiles of English and German.

Table 1 Comparison of rhythm metrics focusing on the duration of vocalic intervals

Metric	Correlate	Unit of analysis	Level	Quantification	Reference
nPVI-V	Duration	Vocalic intervals	Local	Absolute differences	Low et al., 2000
%V	Duration	Vocalic intervals	Global	Proportion	Ramus et al., 1999
VarcoV	Duration	Vocalic intervals	Global	Dispersion	Dellwo & Wagner 2003

3 Contrastive Analysis: English and German Speech Rhythm

English and German are both considered 'stress-timed' (Giegerich, 1992; Kohler, 1995). From a componential perspective, then, the two languages share a number of phonological and prosodic properties that are characteristic of this rhythm class. This section compares English and German in terms of the phonological and prosodic features discussed above and offers a survey of relevant quantitative work.

Let us first turn to phonetic and phonological components. Both languages have a complex syllable structure (König & Gast, 2009: 38, 42; Maddieson, 2013) and phonetically distinguish stressed and unstressed syllables in terms of quality and quantity. Both have the short central vowel [ə] and show schwa deletion and syllabic consonants as extreme forms of reduction. However, the distribution of schwa vowels in German is restricted (Kaltenbacher, 1998). In simple lexemes, they only occur in stem-final syllables (*Hase* [ˈhaːzə]) and inflectional affixes (*ge-dacht* [ɡəˈdaxt]; *denk-e* [ˈdɛŋkə]). Differences in [ə]-distribution are also found in complex lexemes. In both languages, morphophonological processes apply to derived words such as *photography*/*Fotografie*. In German, vowel reduction as a result of stress shift can be observed as a shortening of long vowels (*Foto* [ˈfo:to]—*Fotograf* [fotoˈɡra;f]—*Fotografie* [fotoɡraˈfi;]) but vowels are never reduced to schwa in these contexts. The morphophonology of English, on the other hand, produces unstressed vowels that are shortened and centralized to(wards) schwa (*photo* [ˈfəʊtəʊ]—*photograph* [ˈfəʊtəɡrɑ;f]—*photography* [fəˈtɒɡrəfi]). In general, therefore, unstressed vowels in polysyllabic lexemes show a higher degree of reduction in English. In connected speech, function words can undergo reduction in both languages (*und* [ʊnt] → [(ə)n(t)]; *and* [ænd] → [(ə)n(d)]). In German, however, these reduction processes are stylistically marked—they only occur in informal speech (Kohler, 1995; Wesener, 1999). In English, the weak form of function words (which involves [ə] in many cases) is the unmarked variant, even in formal speech. Thus, while both languages show reduction in function words, a centralization of vowel quality is much more common in English.

At the prosodic level, accentual and final lengthening have been identified as key correlates of rhythm classes. While accentual lengthening is observed cross-linguistically, its magnitude varies between languages. In connected speech utterances, four levels of syllable prominence are often distinguished: (i) unstressed, (ii) secondary stressed and unaccented, (iii) primary stressed and unaccented, and

Table 2 Summary of the contrastive analysis

Feature	English	German
Phonological components		
Complex syllable structure	++	++
Vowel reduction: Length	++	++
Vowel reduction: Quality	++	+
Phonological vowel length distinctions	(+)	+
Prosodic components		
Accentual lengthening	++	+
Final lengthening	+	+

(iv) accented (Vanderslice & Ladefoged, 1972; Gussenhoven, 2004: 20; Fletcher, 2010: 530). The distinction between prominence grading at the lexical and post-lexical level is commonly captured by the labels ‘stress’ and ‘accent’, respectively. In both languages, accented syllables are longer. Comparing the duration of stressed to unstressed syllables, Delattre (1956: 189) reports a ratio of 1.60 for English and 1.44 for German. Similar values were presented by Li (2014), who compared AmE and German speech and observed ratios of 1.55 and 1.43, respectively. In terms of durational marking of prosodic edges, English and German behave similarly, as shown by the lengthening effects in English (1.53) and German (1.50) observed by Delattre (1965). These results were corroborated by Li (2014), who found ratios of 1.63 and 1.67, respectively. Delattre (1965) further reported on the combined effect of accentual and final lengthening, which was greater in English for both open (2.78 vs. 2.25) and closed syllables (2.63 vs. 2.06).

Table 2 gives a summary of the structural profiles suggested by a componential contrastive analysis. While similarities outweigh differences, ‘stress-timing’ properties that are more pronounced in English include the reduction of vowel quality in unstressed syllables and the degree of accentual lengthening.

As mentioned in the preceding section, rhythm metrics aim to quantify these properties. Given the similarities between English and German, we would expect the two languages to exhibit similar, more ‘stress-timed’ scores relative to languages that have been traditionally classed as ‘syllable-timed’ (such as Spanish, for instance), with prominence grading being perhaps slightly more pronounced in English. Since the focus in this paper is on durational properties of vocalic intervals, our literature survey concentrates on the metrics summarized in Table 1 (nPVI-V, %V, and VarcoV). Figure 1 offers a graphical summary of measurements reported in 15 studies (see Appendix 1 for details). For each metric, the y-axis is arranged to reflect ‘syllable-timing’ values at the bottom and ‘stress-timing’ values at the top. Differences in materials and tasks contribute to the variation among empirical estimates (see Arvaniti, 2012). Data points from the same study and/or condition, however, are directly comparable and therefore connected with lines. Figure 1 demonstrates some gross trends:

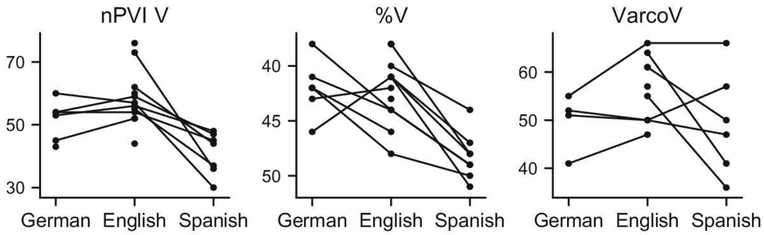


Fig. 1 Graphical summary of rhythm metrics reported in empirical work on German, English, and Spanish (15 studies; see Appendix 1 for details). ©ⓘ[2]

- English/German versus Spanish: Overall, nPVI-V and %V pattern in the expected direction: Spanish scorses tend toward the lower end. VarcoV appears to be less successful at differentiating between representatives of different rhythm classes.
- English versus German: As expected, differences between English and German are minor. On average, English shows slightly higher durational variability of vocalic intervals (nPVI-V).

The next section addresses theoretical work on L2 phonological acquisition. Existing contributions will be examined from the viewpoint of 'rhythmic acquisition', with an eye to whether or not they are capable of accounting for the acquisition of suprasegmental prominence variation.

4 Speech Rhythm and L2 Phonological Theories

4.1 Structural Scope of Theoretical Contributions

Theoretical contributions to the field of L2 phonological research can be grouped along several lines, including their scope, by which we mean the types of structures to which they extend. Table 3 lists several frameworks and indicates whether a particular approach covers segmental and/or prosodic units (see Sönning, 2020: 5–35). Speech rhythm, a special case of the latter level of analysis, is listed separately. The overview suggests that segmental structures receive more extensive coverage. As for speech rhythm, several contributions offer guidance for the study of L2 speech. It should be noted, however, that the influential family of perception-based models (e.g. Best, 1995; Flege, 1995) is concerned exclusively with individual segments. In the following, the tenets of Major's (2001) Ontogeny Phylogeny Model and James'

[2] Images with the symbols ©ⓘ in the figure caption have been published under the Creative Commons Attribution 4.0 licence (CC BY 4.0, http://creativecommons.org/licenses/by/4.0) in the accompanying OSF project (https://osf.io/25kq4/).

Table 3 Structural scope of L2 phonological theories

Contribution	Reference	Segmental	Prosodic	Rhythm
Contrastive Analysis Hypothesis	Lado, 1957	●	●	●
Articulatory Settings	Honikman, 1964	●		
Speech Learning Model	Flege, 1995	●		
Similarity Differential Rate Hypothesis	Major & Kim, 1996	●		
Perceptual Assimilation Model	Best, 1995	●		
Desensitization Hypothesis	Bohn, 1995	●		
Phonological Interference Model	Brown, 1998	●		
Feature Competition Model	Hancin-Bhatt, 1994	●		
Markedness Differential Hypothesis	Eckman, 1977	●	●	●
Structural Conformity Hypothesis	Eckman, 1991	●	●	●
UG Model of Stress Acquisition	Archibald, 1994		●	●
Natural Model of L2 Phonological Acquisition	Dziubalska-Kołaczyk, 1990	●	●	(●)
Naturalness Differential Hypothesis	Schmid, 1997	●	●	(●)
Functional Model of Phonological Acquisition	Boersma, 1998	●	●	
Model of Segmental Acquisition	Colantoni & Steele, 2008	●		
Gradual Diffusion Model	Gatbonton, 1978	●		
Model of Sociolinguistic Variation	Fasold & Preston, 2007	●	●	(●)
Linguistic Theory of L2 Phonological Development	James 1988	●	●	●
Ontogeny Phylogeny Model	Major, 2001	●	●	●
Dynamic Systems Theory	De Bot et al., 2007	●	●	(●)

Note Parentheses indicate that while the framework may be argued to extend to rhythm, this would require the stipulation of non-trivial auxiliary assumptions or premises, for which the literature gives little guidance

(1988) Linguistic Theory of L2 Phonological Development will be discussed and applied to rhythmic properties of German Learner speech.[3]

4.2 The Ontogeny Phylogeny Model (Major, 2001)

Major's (2001) Ontogeny Phylogeny Model (OPM) combines theoretical insights into transfer, similarity, and typological markedness into a formal model outlining the dynamic nature of interlanguage (IL) development. The OPM rests on two basic assumptions: (i) a learner's IL consists of three structural components (L1, L2, U) and (ii) the relationship between these components changes systematically over time. Thus, it is held that every structure found in learner speech is attributable to one of three sources: It may be a transferred L1 structure (L1), a target language structure (L2), or a universal structure that is not part of L1 or L2 (U). The latter component is defined by Major (2001: 83) as 'the universal set of properties of the human language capacity and the resulting universal characteristics of languages, […] [including] anatomical, functional and processing properties of the human mind'.

The OPM stipulates an organized interplay of L1, L2, and U over the course of IL development, which depends on the type of structure that is acquired (and also on speaking style). These assumptions are expressed as four 'corollaries', which are shown graphically in Fig. 2. The basic chronological assumption states that, over the course of five hypothetical developmental stages, L1 influence decreases while L2 structures increase; U first increases and then decreases. As panels (b) and (c) show, this interplay follows a different pattern for similar and marked structures. Compared to 'normal' language structures, i.e. units that do not classify as marked or as similar to an L1 counterpart, marked structures are acquired at a slower rate (7 vs. 5 stages). While equivalent L2 trajectories are posited for similar and marked structures, the relative influence of L1 and U differs: In the acquisition of similar structures, L1 transfer is more persistent; U, on the other hand, exerts no notable influence. For marked structures, transfer is assumed to decrease rapidly, while U rises to exert considerable influence. The OPM thus brings together two well-documented constraints on interlanguage—transfer and universals—and states that their weight depends on developmental stage and properties of the focal structure. Whether L1 influence or language-universal biases (or both) are observable in learner speech therefore depends on characteristics of the learner and the structure.

[3] A reviewer raised the question of why these two models were chosen. Since the model proposed by Major (2001) may be considered a unification of several contributions including Lado (1957) and Eckman (1977, 1991), it covers the explanatory notions proposed in those accounts (i.e. transfer, markedness, and language universals). The only remaining model that is directly applicable to the acquisition of rhythm, then, is Archibald (1994), which is restricted to prominence grading at the lexical level, however.

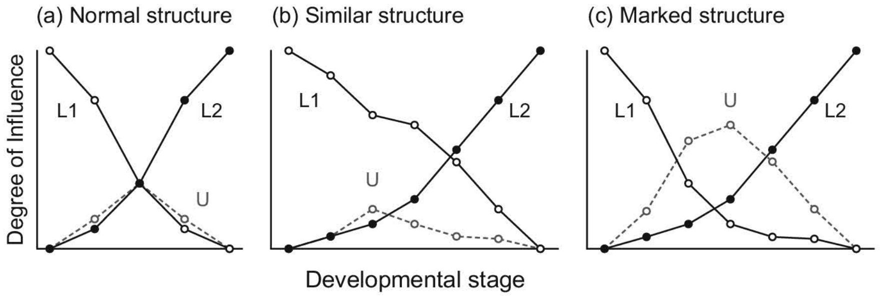


Fig. 2 Corollaries of the OPM: Interplay of transferred structures (L1), target language structures (L2), and language universals (U) in the course of L2 phonological acquisition of normal (left), similar (center) and marked structures (right); from Sönning (2020: 31).

4.3 The Linguistic Theory of L2 Phonological Development (James 1988)

James' (1988) Linguistic Theory of L2 Phonological Development (LTD) focuses on the interplay of different levels of phonological representation in the course of L2 development. Three levels—the lexical, prosodic, and rhythmic—are posited to interact systematically. For the prosodic level of representation, a non-linear framework similar to metrical phonology is employed (see James, 1986 for details). As illustrated in Fig. 3, it comprises seven layers, with binary strength values (*s*-marks) assigned to constituent nodes at each level. These add up to determine the structural weight of a syllable.

At the rhythmic level of representation, the units at each layer are described with the generalized scheme (proclitic) head (enclitic), the head being the obligatory element. These rhythmic features reflect speech rate: Proclitics (P) show increased tempo; heads (H) and enclitics (E) are characterized by decreased tempo. The supra-syllabic rhythmic structure for the example sentence is shown in Fig. 4. Similar to prosodic *s/w*-marks, rate features add up to determine the tempo, or duration, of a syllable. The bottom-up acquisition of rhythmic structure thus yields increased temporal differentiation in the speech stream.

James (1988) posits that L2 phonological acquisition follows a universal progression from the lexical to the prosodic to the rhythmic level. The lower levels provide

Fig. 3 Illustration of supra-syllabic *s*/*w*-marking at five hierarchical levels of prosodic representation (after James, 1986, 1988)

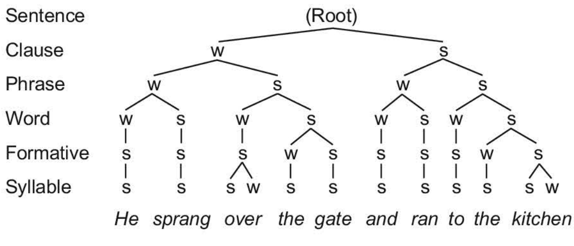
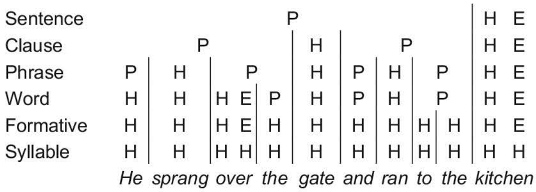


Fig. 4 Illustration of rhythmic structure: proclitics (P, faster tempo), heads (H, slower tempo), and enclitics (E, slower tempo) (after James, 1986, 1988)



substance for the acquisition of higher-level structure: Higher-level *s*-marks rest on lower-level *s*-marks and the same holds for heads. The LTD also predicts acquisition sequences for the prosodic and rhythmic levels. At the prosodic level, *s*-marked units are acquired before *w*-marked units due to 'logical priority' (James, 1988: 155); at the rhythmic level, the properties of peaks are acquired earlier than those of proclitics.

The bottom-up advancement of L2 phonological acquisition generates developmental predictions: With acquisition of the rhythmical level requiring sufficient prosodic structure, and the prosodic representation in turn commencing after the lexical level has been established, rhythmic and prosodic patterning are expected to emerge relatively late. In general, then, strength and rate asymmetries grow during IL development. With structural strength being cumulative, an increasing *s*/*w*-differentiation will yield a gradual increase in strength effects in syllables and segments. The same is true for rate values that are determined based on rhythmical structure. The postulated suprasegmental acquisition sequence 'peaks before proclitics' and '*s*-marked units before *w*-marked units' suggests a bias toward overarticulation at early stages of L2 acquisition; backgrounding or reduction of *w*-marked units and proclitics is expected to emerge at a later stage. Before we discuss the empirical application of the OPM and the LTD in more detail, Sect. 5 introduces the data used in the following analyses.

5 Data

A total of 88 speakers were recorded: 62 German learners and 25 native speakers of English (11 American and 15 British English subjects[4]). The data were collected as part of a more comprehensive investigation of phonological variation in GLE (Sönning, 2020) and therefore not specifically designed for the purposes of the present study. The sampling of instructional-setting German learners aimed at capturing a broad range of proficiency levels. Biographical details of subjects are given in Table 4. A foreign accent rating (FAR) was used to obtain a global accentedness score

[4] All BrE and AmE informants reported that both of their parents' native language was English. Neither group can be considered as representing a well-defined variety of English. Nevertheless, with the reported minimal education level of all subjects being a bachelor's degree, the native speakers recorded in this study may be described as speaking an educated variety of English. Most BrE informants ($M_{Age} = 27$; $SD = 6$) had grown up in the London area and the Midlands. Native speakers of AmE ($M_{Age} = 24$; $SD = 5$) were predominantly from the northeastern part of the US.

Table 4 Descriptive statistics for the sample of 62 German learners

Variable	Distribution
Gender	39 female (63%); 23 male (37%)
Age	*Mean* = 18; *SD* = 4; *Min* = 11; *Max* = 30
Grade	6 (*n* = 4); 7 (12); 8 (1); 9 (3); 10 (6); 11 (1); 12 (9); tertiary education (26)
AOL	Age at onset of learning: *Mean* = 10; *SD* = 1.5; *Min* = 3; *Max* = 13
FAR	Foreign accent rating, 12-point scale (scores from 1 to 12): *Mean* = 6.0; *SD* = 2.5; *Min* = 1.6; *Max* = 10.8

for each learner. Two British English native speakers rated the degree of foreign accent on a 12-point scale from 1 ('strong foreign accent') to 12 ('native speaker level'), based on 4 utterances per speaker. The averages of the raw scores, which range from 1.6 to 10.8, were converted to *z*-scores, which, by definition, have mean 0 and standard deviation 1.

As for the recordings, a reading task was used to elicit 10 sentences embedded in short question–answer sequences (see Appendix 2). This aimed at eliciting consistent accentual patterns. Participants were given time to familiarize themselves with the materials and then asked to read both turns of the dialogue; they were allowed to correct themselves and re-read a sequence. The sentences included 105 vocalic intervals in total, producing 6509 measurements for learners (1 missing) and 2618 measurements for native speakers (7 missing). The acoustic analysis was carried out in Praat (Boersma & Weenink 2014) and the data were segmented manually following the principles outlined in Machač and Skarnitzl (2009). Specifically, the boundaries of vocalic intervals were determined using the onset and offset of the second formant and changes in waveform amplitude. Onset /w j r l/ were assigned to consonantal intervals, coda /l/ was labeled as consonantal (except when syllabic), coda /r/ was treated as consonantal but *r*-coloring as part of the nucleus /ɜ˞ː ə˞/. The interval between stop release and the onset of voicing was treated as part of the consonantal stretch. To facilitate the rescaling of the durational measurements (see below), deleted vowels were coded as having a duration of 2 ms. The complete data are available from TROLLing (Sönning, 2022).

6 L2 Rhythm in German Learner English: An OPM Perspective

6.1 Working Assumptions

In order to derive OPM predictions about prominence grading in GLE, assumptions must be made about (i) the status of English speech rhythm as a normal, similar, or marked category of speech, and (ii) the nature of L1, L2, and U influences on speech production.

A classification of English speech rhythm as ‘similar’ to German speech rhythm may seem warranted in the light of our contrastive analysis. It should be noted, however, that in the field of L2 phonology research the notion of similarity is rooted in perception-based models of L2 phonological acquisition. Perceptual similarity statements rely on the assumption that listeners are able to selectively and contrastively perceive relevant units in the speech stream. It is unclear whether (non-native) listeners can (and do) consciously attend to structures above the level of the segment and whether they are able to make similarity judgements. For the present, we will therefore exclude perceptual similarity as a relevant structural property.

As for markedness, ‘stress-timing’ can be considered more marked than ‘syllable-timing’ on several grounds. Research into the L1 acquisition of ‘stress-timed’ languages indicates that children develop from ‘syllable-timed’ to ‘stress-timed’ speech (Allen & Hawkins, 1980; Cruttenden, 1979), which has received support from acoustic studies using rhythm metrics (Grabe et al. 1999; Bunta & Ingram, 2007; Payne et al., 2011). Certain properties of ‘stress-timed’ languages appear to be more marked than those of ‘syllable-timed’ languages. Thus, the reduction of weak syllables has been observed to emerge relatively late (Allen & Hawkins, 1980), as children selectively attend to stressed syllables (Blasdell & Jensen, 1970; Risley & Reynolds, 1970). It has also been noted that children acquiring a ‘syllable-timed’ language show adult-like timing patterns at a younger age, both at the word (Vihman et al., 2006) and utterance level (Grabe et al., 1999). Research on rhythm development in L1 acquisition has shown parallels between children from typologically different languages (Grabe et al., 1999 for English, German and French; Payne et al., 2011 for English, Spanish, and Catalan). This suggests that ‘syllable-timing’ properties may, in general, be considered the default setting in L1 acquisition. Biases toward ‘syllable-timing’ properties in World Englishes are also consistent with their status as the unmarked type of rhythmic organization. Thus, Nishihara & van de Weijer (2012) consider ‘asymmetric borrowing’, that is, the tendency in varieties of English to adopt ‘syllable-timed’ properties instead of those typical for L1 varieties such as American or British English, as an indication of markedness. In light of these observations, the type of prominence variation found in L1 English will be considered, collectively, as a marked property of speech.

Next, we need to state our assumptions about the nature of L1, L2, and U influence in the acquisition of English rhythm by German learners. Given the findings of our contrastive analysis, L1 and L2 should produce similar surface patterns, with L2 showing a slightly higher degree of prominence grading. The role of U, on the other hand, can be derived from markedness considerations. Assuming that U reflects universal influences that are also operative in L1 acquisition, its effect should surface in a tendency toward prominence-leveling in speech production. This assumption is coherent with previous empirical research on reduction phenomena in interlanguage phonology, which suggests this to be an area of difficulty in learner speech (Aoyama & Guion, 2007; Flege & Bohn, 1989; Gut, 2006). It has also been noted that overarticulation of unstressed syllables is a general feature of non-native speech (Barry, 2007).

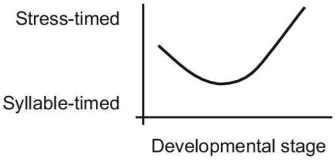


Fig. 5 Predictions based on the OPM: U-shaped developmental pattern.

In summary, the application of the OPM to rhythmic properties of GLE will rely on the following assumptions:

- English speech rhythm is a marked category of speech.
- L1 and L2 will surface in a tendency toward prominence variation, i.e. 'stress-timing' properties.
- U will surface in a tendency toward prominence leveling, i.e. 'syllable-timing' properties.

6.2 Predictions

Based on the tenets of the OPM and the assumptions stated in the preceding section, we can formulate the following expectations: Initial transfer from L1 should surface in (near) target-like prominence variation. The pronounced influence of U, which is expected due to the relatively more marked status of 'stress-timing' patterns, is predicted to surface in a tendency toward prominence-leveling at intermediate stages. The final increase in L2-like patterns and the decreasing influence of U should yield more temporal differentiation between units in the speech stream. In short, we expect prominence grading to follow a U-shaped pattern (see Fig. 5).

6.3 Method and Data

For the application of the OPM to speech rhythm in GLE, we will rely on the rhythm metrics described in Table 1.[5] The next paragraph gives information about statistical procedures and may be skipped without loss of continuity.

Given the controlled method of elicitation (the same 105 syllables were produced by each speaker), the data are highly structured. Measurements are grouped by syllable, and syllables in turn are nested in 10 sentences. Further, measurements are

[5] The choice of these particular metrics was motivated by the following considerations: (i) in the interest of simplicity, the focus in the present study is restricted to the analysis of vocalic intervals, (ii) these metrics are widely used in the literature comparing different languages (see Fig. 1 and Appendix 1), and (iii) they feature prominently in previous work on speech rhythm in German Learner English (e.g. Ordin et al. 2011; Li & Post 2014; Ordin & Polyanskaya 2015).

clustered by speaker. Accordingly, the data were analyzed with a hierarchical (mixed-effects) model. The nPVI-V data include 95 adjacent comparisons per speaker. These were analyzed with a hierarchical linear regression model with random intercepts for speaker (level 2) and adjacent pair (level 2). The adjacent pairs are nested in sentences and random intercepts for sentences were therefore included at level 3. As in all other models reported in this study, proficiency (as measured by the foreign accent rating) is a property of the individual speaker (i.e. a between-speaker variable); it is therefore a level-2 predictor. The %V scores are based on 10 proportions per subject (one for each sentence). These proportions were transformed to logits (i.e. log odds) and analyzed with a hierarchical linear regression model with random intercepts for speaker (level 2) and sentence (level 2). Scores were back-transformed to percentages for presentation and interpretation. VarcoV values require different treatment, as they are measures of dispersion rather than location. To preserve statistical uncertainty in VarcoV estimates, the standard deviation was modeled using a hierarchical linear regression model with random intercepts for subject (level 2) and sentence (level 2). This is to say that the variation of measurements (rather than their central tendency) was the outcome of interest. In line with the rationale behind VarcoV, the durational measurements were rate-normalized, i.e. converted to express duration relative to the average vocalic interval duration for each subject. The statistical analyses were carried out in R (R Core Team 2016), relying on the "brms" package (Bürkner 2016), which in turn builds on the Bayesian inference engine Stan (Stan Development Team 2016). The posterior distributions generated by the models were processed in R and the packages 'lattice' (Sarkar 2008) and 'latticeExtra' (Sarkar & Andrews 2016) were used for data visualization. Technical information about model parameters and priors are deferred to the online appendix (https://osf.io/25kq4/), which also includes the complete R code.

For each rhythm metric, we will compare three candidate models. These encode three possible relationships between prominence variation and proficiency level: (i) no systematic relationship, (ii) a straight-line trend, and (iii) a U-shaped trend. Model (iii) is the one suggested by the OPM and the other two are simpler descriptions, which will serve as a point of reference. Our primary concern is to determine, for each rhythm metric, which of these patterns receives most support from the data. Such a ranking can be established using information criteria.[6] These can be re-expressed as Akaike weights, a heuristic and more intuitive measure of the relative goodness of a model. Such weights range from 0 to 1 and can be interpreted as the probability that a given model is the best one in the set (Burnham & Anderson, 2002: 75; McElreath, 2016: 197–201). The type of information criterion we will rely on is LOOIC (Vehtari et al., 2017), whose scores are then translated into Akaike weights. In addition, the

[6] The purpose of information criteria is to provide an assessment of how well the model—in our case, the pattern (horizontal vs. linear vs. U-shaped trend)—is likely to generalize to new observations (i.e. other speakers from the population of L1 German learners of English). Information criteria help the researcher guard against 'overfitting', that is, reporting and interpreting idiosyncratic features of the sample in hand, which may not replicate in a new sample of observations. They report what is referred to as the out-of-sample predictive accuracy, with lower values signaling higher accuracy, that is, a higher goodness rating.

Table 5 Model comparison results

Model	Pattern	LOOIC	(SE)	Akaike weight	
nPVI-V					
(i)	None	5337.9	(141)	0.19	|||||||
(ii)	Straight-line	5336.6	(141)	0.35	||||||||||||||
(iii)	Curvilinear	5336.1	(141)	0.46	||||||||||||||||||
%V					
(i)	None	− 642.1	(41)	0.24	||||||||||
(ii)	Straight-line	− 643.5	(41)	0.49	|||||||||||||||||||
(iii)	Curvilinear	− 642.4	(41)	0.27	|||||||||||
VarcoV					
(i)	None	11,428.4	(134)	0.02		
(ii)	Straight-line	11,421.5	(134)	0.60	||||||||||||||||||||||||
(iii)	Curvilinear	11,422.5	(134)	0.38	|||||||||||||||

data are shown graphically, with native speaker values added for comparison. The way scores pattern across proficiency levels is captured by a flexible regression line, more specifically a cubic B-spline[7] (Fahrmeir et al. 2013: 426–431). The trend and its uncertainty, then, are purely data-based, which will allow us to appreciate visually the degree to which the data support different candidate models.

6.4 Results

Model comparison results are shown in Table 5, where a semi-graphical representation of Akaike weights is added (a single bar denotes 0.025 units of weight). The metrics differ in the extent to which they allow us to differentiate between the three patterns. While %V fails to identify a best candidate, nPVI-V and VarcoV provide some indication against 'no trend'. However, the data do not discriminate well between models (ii) and (iii).

While these comparisons do not suggest a single best model, a visualization of the patterns in the data is still revealing. Figure 6 shows, for each rhythm metric, estimates for the 62 learners. The raw, data-based trend is superimposed, with error bands denoting 50% and 90% uncertainty intervals. The patterns provide virtually no indication of a U-shaped profile.

[7] Three knots were chosen to keep the flexibility at a reasonable level. See R code for further details.

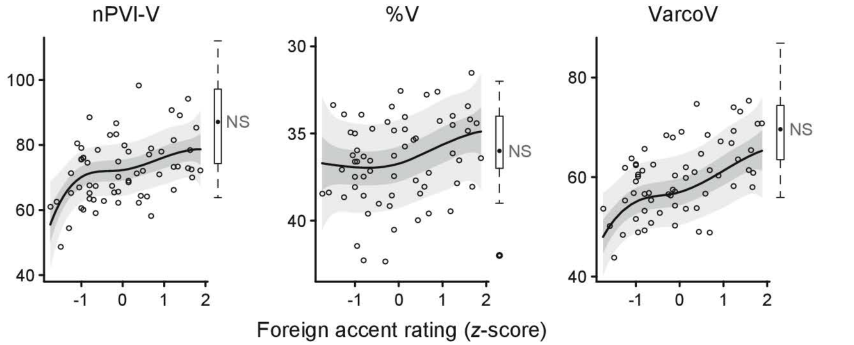


Fig. 6 Rhythm metrics by proficiency level with flexible regression lines. Error bands denote 50% and 90% uncertainty intervals. Boxplots at the right margin show the distribution of scores for the 25 native speakers recorded in this study.

6.5 Discussion

To recapitulate, we considered the acquisition of English speech rhythm by German learners from the viewpoint of Major's (2001) OPM. A contrastive analysis and a survey of markedness properties of rhythm types led us to postulate a U-shaped trend in the overall degree of temporal variability of vocalic intervals. This expectation is rooted in the OPM assumption that initial stages of L2 acquisition should show a disproportionate influence of the L1. Having identified several rhythmic parallels between English and German, transfer from L1 was expected to surface in near-target timing patterns, followed by a U-induced reduction of temporal variability, that is, more 'syllable-timed' speech at intermediate levels. However, no evidence was found for the hypothesized curvilinear trajectory and nPVI-V and VarcoV scores showed that, compared to native speakers, on average, all developmental stages are characterized by a lack of temporal variability among vocalic intervals.

These findings are consistent with previous work on timing patterns in GLE (Li & Post, 2014; Ordin & Polyanskaya, 2015), in which a monotonic increase in temporal variability across proficiency levels was reported. The present study has extended the empirical scope toward lower-proficiency levels by including early-stage instructional-setting learners. Against the backdrop of the OPM, we would have expected this sub-population of learners to not have progressed beyond the stage of L1 influence. In light of the present findings, then, there is growing indication of a linear (i.e. straight-line) increase in durational variability across different (if not all) levels of pronunciation ability in GLE. A U-shaped trajectory, on the other hand, does not seem to provide an adequate description.[8]

[8] It should be noted that existing research, including the present study, offers limited information on genuinely developmental patterns due to its cross-sectional nature. In order to make reliable statements about change across different stages of L2 development, longitudinal data would be

Instead of questioning the OPM, we first need to cast a critical eye on the set of assumptions we had to state and rely on to formulate predictions. The expectation of a curvilinear trend rests on a contrastive analysis, which revealed similar rhythmic profiles in English and German. We therefore expected L1 transfer of the full set of phonetic, phonological, and prosodic components to yield timing patterns that are close to those of native speakers. We must recognize, however, that we are not able to pin down the precise point along the rhythmic continuum that would characterize the hypothetical first stage of the OPM, that is, a full L1 transfer scenario. This greatly compromises our ability to distinguish between different explanations of the observed patterns, specifically, the delineation of L1 and U influence. We cannot fully rule out the possibility that L1 transfer might also yield at least a certain degree of prominence leveling, given that German is somewhat less 'stress-timed' than English. Our survey of the empirical literature (see Fig. 1) suggests that these concerns are valid: The deviation of learner scores from those of native speakers is within the range of variation that has been observed between English and German, that is, about 10 to 20 points on each the nPVI-V and the VarcoV scale. We are thus facing considerable uncertainty when it comes to interpreting the observed patterns in terms of IL components: The steady upward cline could reflect (i) L1 slowly giving way to L2 structures, with U playing no role; (ii) U gradually giving way to L2 structures, with L1 playing no role; and (iii) simultaneous influence of L1 and U gently giving way to L2.

Li and Post (2014: 244) partly steered clear of this interpretive dilemma by also recording utterances of German learners in their L1. They observed that lower-intermediate learners showed a drop in 'stress-timedness' compared to their L1 control values. This amounted to just under 5 points on each of the nPVI-V and VarcoV scale. Assuming perfect comparability between the English and German materials, these data suggest that L1 alone may not be able to fully account for the observed patterns, leaving (ii) and (iii) as possible structural constellations. Given the data in the present study, however, any claim of Major's model not providing an adequate account of the acquisition of prominence variation by German leaners is poorly probed. Only via interpolation with findings in the literature may we arrive at the interpretation that a clear facilitative effect of L1 'stress-timing' properties does not seem to be borne out by the data.

What, then, can be learned from taking an OPM perspective on the acquisition of speech rhythm by German learners? In order to extract informative predictions from L2 phonological frameworks, we needed to state a number of auxiliary assumptions. This is necessary whenever a theoretical model leaves considerable interpretive leeway. An example in the case of the OPM is U as a structural speech rhythm component and markedness as a property of rhythm classes. We stated that U would operate to produce prominence-leveling and that 'stress-timed' properties of speech rhythm are, collectively, a marked category of speech. Both attempts to add rhythmic substance to OPM components arguably arrived at equally vague formulations. We

required. Nevertheless, current knowledge about cross-sectional patterns consistently points to a steady increase in temporal prominence grading across proficiency levels.

thus did not manage to make a very general theory concrete for the phenomenon under study. Rather, we allowed an equal level of fuzziness to enter our predictions by relying on imprecise and weak links between theory and predicted pattern.

To a large extent, this unsatisfactory exchange between theory and data may be due to the fact that we have attempted to apply a model that has been formulated based on segmental phenomena to a much more complex category of speech. It is questionable whether notions such as markedness and similarity can be meaningfully applied to an assemblage of features that collectively produce the percept of speech rhythm. This would suggest that linguistic constraints that have been observed in the acquisition of segmental structures may not be directly extended to speech rhythm. Similar reservations may be voiced about the way we measured speech rhythm. By relying on rhythm metrics, we treated speech rhythm conceptually and empirically as a complex but single structure (or ‘category of speech’), when in fact it is currently understood as an ensemble of lower-level properties (see Sect. 2). In other words, we may have blindly followed the lead of rhythm metrics by coercing a multidimensional construct into a single descriptive category. While this allowed us to (unsuccessfully) operate from an OPM perspective, we may have confused the logical relationship between construct (speech rhythm) and measurement (rhythm metrics), understanding the former in terms of the latter. Thus, it could also be argued that our approach to the object of interest was too coarse and simplistic. We will return to this point in the general discussion.

7 L2 Rhythm in German Learner English: An LTD Perspective

7.1 Working Assumptions

Concerning the acquisition of prominence variation, the key aspect of James’ (1988) model is the bottom-up advancement of learners, who are assumed to build prosodic representation level by level. To apply the LTD to L2 data, the way in which speech is organized above the level of the segment must be specified. The system proposed by James (1986, 1988) will not be used in the present study for several reasons. For one, its application to the materials was in many cases not straightforward, which arguably questions the reliability with which surface structures can be mapped onto this set of hypothesized representations. Further, the strata constituting this model create a considerable level of complexity, with 5 levels of prosodic *s/w*-structure and 6 levels of rhythmic structure. Given these limitations, a simplified and largely theory-neutral template was chosen, which connects to the componential view of prominence variation outlined above. Specifically, prosodic strength asymmetries were coded at four levels:

- *Lexical* (level 1): At the level of the lexical/grammatical word, a simplified two-way distinction is made between unstressed syllables with a reduced vowel (*w*-marked) and stressed syllables with a full vowel (*s*-marked). Primary and secondary stress both receive *s*-marks and monosyllabic words are considered *s*-marked.
- *Syntactic* (level 2): At the post-lexical level, content words (*s*-marked) and grammatical words (*w*-marked) are distinguished. The *s*-mark is assigned to syllables with primary lexical stress in content words.
- *Nucleus* (level 3): At the level of the intonation phrase, the syllable carrying the nuclear accent is *s*-marked.
- *Final* (level 4): Intonation phrase boundaries at the right margin are *s*-marked. Specifically, the final syllable that carries strength at level 1 and subsequent units are *s*-marked.

While this coding scheme may draw legitimate criticism, it also offers benefits. Despite its rudimentary structure, it captures a number of phonological and prosodic components of durational prominence grading. Binary distinctions, which are adopted from James' (1988) notion of *s/w*-marking, allow for a reliable encoding of strength asymmetries. Further, this template offers a parsimonious way of representing relevant supra-syllabic properties. As to the hierarchical organization of these strata, levels 1 to 3 are layered in the sense that higher-level *s*-marking requires structure at the lower level(s)—that is, higher-level *s*-marks always rest on lower-level *s*-marks. The status of boundary effects, on the other hand, is less clear. James (1988) claims that rhythmic organization emerges last. Since final lengthening resembles the role of enclitics in his scheme, we will in the following assume that phrase-final *s*-marking is acquired last.

To illustrate the implications of this 4-level scheme for the acquisition of prominence grading, consider the sentence *He's from the north of Germany* (the coding of the materials is documented in the online appendix (https://osf.io/b34p8). Figure 7 illustrates the bottom-up construction of durational variability, which proceeds from left to right. Prominence leveling, which we will consider as the default setting in L2 acquisition, is illustrated in the leftmost arrangement: At this pre-lexical level, all units are of equal prominence. As learners build up prosodic representation level by level, systematic variability emerges, starting with strength asymmetries at the 'Lexical' level. We will assume that the acquisition of *s/w*-marking at this level yields a backgrounding of *w*-marked units relative to *s*-marked units. Similarly, the emergence of *s/w*-marking at the 'Syntactic' level will be observable as a backgrounding of function words relative to lexical elements. At the 'Nucleus' level, *s/w*-marks will surface in a foregrounding of the syllable carrying the nuclear accent. Likewise, differentiation at the 'Final' level results in a lengthening of units in final position. Prominence grading in the right-most pattern then reflects the aggregation of *s/w*-marks across all four levels.

The patterns in Fig. 7 then reflect different stages in L2 acquisition. Before we go further, we will consider these stages from a different perspective, as this will help us follow the methodological procedure. Let us assume that the last ('Final') stage

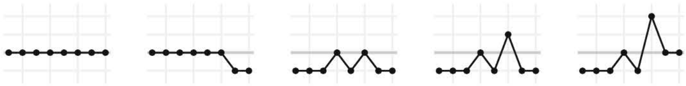


Fig. 7 Acquisition of prosodic structure: Development of *s*/*w*-contrasts

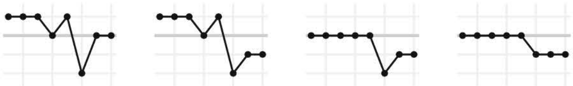


Fig. 8 Theoretical deviations at different developmental stages: Points (i.e. syllables) below the reference line show hypoarticulation (too short), those above hyperarticulation (too long).

approximates the type of prominence variation found in native speech. Relying on the LTD and the assumptions outlined above, we expect learners to deviate systematically from this target pattern. Based on the developmental stage a learner is at, the model makes predictions about whether a syllable will show hyperarticulation (surplus of prominence) or hypoarticulation (lack of prominence). In terms of timing patterns, this enables us to state whether a vowel is expected to be too long or too short. This is illustrated in Fig. 8, which shows four hypothetical stages. For ease of exposition, we will proceed 'backwards':

- Stage 3: Learners have advanced to the 'Nucleus' level. They deviate from native speakers in that they show no pre-boundary lengthening. In the example sentence, pre-boundary lengthening affects the three final syllables (*Ger-ma-ny*). As the learner has not progressed to the 'Final' stage, these three syllables will show hypoarticulation: they are too short.
- Stage 2: The learner utterance also lacks nuclear accentual lengthening and the third-to-last syllable (*Ger-ma-ny*) is therefore further hypoarticulated.
- Stage 1: The failure to background function words at this stage yields an overarticulation of syllables 1 (*He's*), 2 (*from*), 3 (*the*) and 5 (*of*).
- Stage 0: Complete prominence leveling further yields an overarticulation of unstressed syllables in content words (*Ger-ma-ny*).

The patterns in Fig. 8 were arrived at by subtracting from the 'Final'-level pattern the pre-lexical (stage 0), 'Lexical' (stage 1), 'Syntactic' (stage 2), and 'Nucleus' pattern (stage 3) shown in Fig. 7. As explained in more detail below, these deviations will be the key quantity in the following analyses.

7.2 Predictions

The four levels of prosodic representation allow us to distinguish five syllable types. Thus, the addition of *s*-marks at the lexical, syntactic, and nucleus level yields four levels of prominence:

- P1a: Unstressed syllables in lexical words
- P1b: Monosyllabic function words
- P2: Syllables carrying lexical stress
- P3: Syllables carrying lexical stress and the nuclear accent

These prominence levels can be discerned in the coding scheme, where they are reflected in the number of *s*-marks resting on a syllable (see online appendix at https://osf.io/b34p8). Besides these four prominence levels, we can distinguish final from non-final syllables. Table 6 summarizes the properties of these syllable types in terms of *s/w*-marking and position in the intonation phrase.

Our focus will be on deviation patterns in these five syllable types. For each learner, we can determine, based on instrumental measurements, whether a certain type is too long or too short, on average. As illustrated in Fig. 8, the LTD makes predictions about the direction of these deviations and how they change over the course of L2 development. In Fig. 9, expected deviations are shown schematically. P2 serves as a baseline, since our assumption is that prominence levels P1a and b are affected by backgrounding (shortening) whereas prominence level 3 and final units are affected by foregrounding (lengthening). Thus, at stage 0, syllables of prominence 1a and 1b are overarticulated in contrast to those of prominence 3 and type 'Final', which are too short. Deviations disappear gradually and in accordance with the stages shown in Figs. 7 and 8. At stage 3, then, the only discrepancy that remains is that between final and non-final syllables: The former are too short in relative terms. Stage 4, which was not shown above, is characterized by no systematic deviations from target language timing patterns. In the following section, we discuss how, by means of a quantitative analysis, the present study aims to detect and describe the nature of these deviation patterns in learner speech.

Table 6 Description of syllable types

Syllable prominence	'Lexical'	'Syntactic'	'Nucleus'	'Final'
P1a	*w*	*w*	*w*	No
P1b	*s*	*w*	*w*	No
P2	*s*	*s*	*w*	No
P3	*s*	*s*	*s*	No
Final position	-	-	-	Yes

7.3 Method and Data

The aim of the following analyses is to assess whether learners in our sample show a bottom-up progression, with low-proficiency subjects exhibiting the hypothesized deviation patterns for early stages (cf. Figure 9) and high-proficiency levels resembling those posited for later stages. To begin with, however, two methodological concerns must be addressed. First, we need to take into consideration that speech rate will affect vowel duration. Differences in tempo can be canceled out by centering durations, whereby the duration of each vowel is expressed relative to the speaker's average vowel duration (a vowel may be, say, 50 ms longer than the speaker's average vowel duration). Positive deviation scores then indicate relatively long vowels. This leads us to the second concern: If we take the average vowel duration as the within-speaker reference point, it should be representative of the distribution of measurements—that is, it should be roughly located at its center. The distribution of durational measurements, however, is typically not symmetric but skewed, as values are bounded at the lower but not the upper end of the scale. For the ensuing analyses, vowel durations were therefore rescaled using the square root transformation[9] to more closely approximate normal, or at least symmetric, distributions at the speaker level. To cancel out differences in speech rate, these square root durations were then centered at the speaker mean. This rescaling aims for better comparability; we must accept, as a trade-off, the fact that we will be comparing rather abstract scores.

Recall that we are interested in how learners deviate from target language timing patterns. It therefore makes sense to express vowel durations produced by learners relative to those of native speakers. For each vowel, this new measurement reflects whether it was longer or shorter than in native speech. To this end, the utterances of the 25 native speakers in the study were used to establish a 'target' durational profile for the 105 vowels. The construction of this target profile for sentence 1 is illustrated in Fig. 10, where the grey lines show individual durational profiles (i.e. 25 profiles, one for each native speaker). The target profile, which is shown in black, is based on

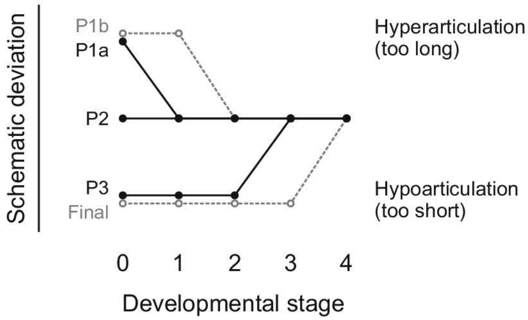


Fig. 9 Schematic illustration of deviation patterns at different developmental stages with prominence level 2 (P2) as the baseline of comparison.

[9] For this data set, the log transformation was less successful at establishing within-speaker symmetry.

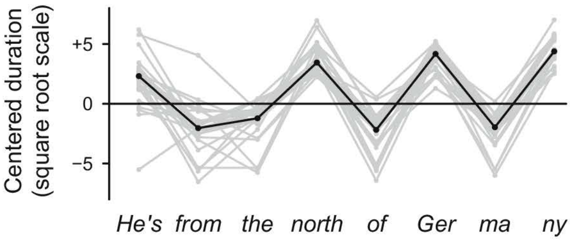


Fig. 10 Construction of the target durational profile: Grey lines show the 26 patterns for the native speakers. The black line shows the target profile, which connects the medians.

the median duration of each vowel. This median profile will serve as the baseline of comparison in the following analyses.

Deviations from temporal prominence patterns in native speech can now be assessed by comparing vowel durations to this target profile. The difference between a learner vowel and the target profile will be referred to as a *deviation score*. Positive deviation scores reflect hyperarticulation (the vowel was too long), negative scores reflect hypoarticulation (the vowel was too short). These deviation scores form the basis of the following analyses.

Our aim, then, is to determine whether the five syllable types show the expected deviations across proficiency levels. Translating the schematic representation in Fig. 9 into deviation scores yields the constellation shown in Fig. 11, where deviation patterns are not expressed relative to a certain prominence level (as in Fig. 9), but relative to the speaker's average vowel duration (i.e. a deviation score of zero). While this may seem counterintuitive, we need to bear in mind that changes in timing patterns systematically affect the within-speaker reference point, that is, his or her average vowel duration. To illustrate, consider a learner progressing to Stage 1. The LTD states that syllables of type P1a will now be properly backgrounded, while no changes occur for the other prominence levels. The speaker's average vowel duration therefore decreases. As a consequence, the remaining prominence levels artificially shift upwards, as they receive a new deviation score relative to the new reference point. Due to our methodological approach, then, deviation scores are re-centered at zero at each stage, which results in a relative shift of the other prominence levels. Note, however, that this adjustment of hypo- and hyperarticulation patterns does not yield qualitatively different predictions. In fact, the expected arrangement of deviation scores by syllable type remains very much the same as that shown in Fig. 9.

The crucial question in the following analyses is whether the empirical deviation patterns in our sample resemble the theoretical values shown in Figs. 9 and 11. The remainder of this paragraph lays out the statistical procedures and can be skipped without losing the main thread of the argument. To extract the empirical deviation patterns from the sample of German learners, a hierarchical linear regression model was fitted, including random intercepts for subjects ($n = 62$) and syllables ($n = 105$). Subjects further received a random slope for each syllable type described in Table 6. These random slopes document learner-specific deviations from the target profile. The model included, as fixed effects, the five syllable types and their interaction with the foreign accent rating as a measure of proficiency. Thus, the random slopes

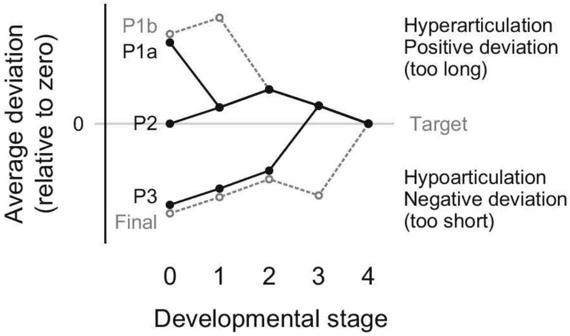


Fig. 11 Expected average deviation patterns by syllable type and developmental stage: Translation of the schematic patterns shown in Fig. 9 to the scale of deviation scores.

were modeled conditional on foreign accent rating to detect systematic changes in deviation patterns across proficiency levels. This allows the model to capture trends in the magnitude and direction of deviation scores for each syllable type. For the analyses presented in the following section, use was made of the resources outlined in Sect. 6.2. Details about the model and the complete R code for the analysis can be found in the online appendix (https://osf.io/25kq4/).

7.4 Results

Figure 12 provides a summary of the empirical deviation patterns, where each line represents a syllable type. The trends in these lines show how the direction and magnitude of deviation scores change across proficiency levels. The lines fan out to the left, which indicates that the timing patterns at lower proficiency levels correspond least to those of native speakers. At higher proficiency levels, there is convergence toward the target, which reflects alignment with the median profiles derived from the native speaker data.

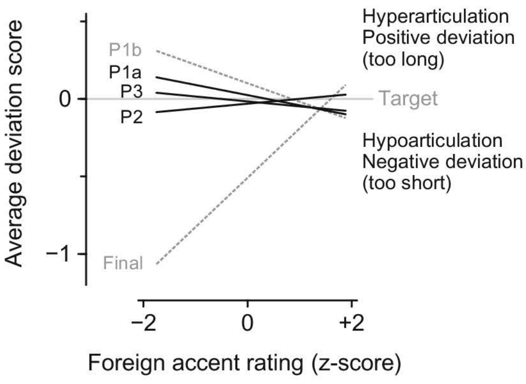


Fig. 12 Empirical deviation patterns by syllable type and proficiency level.

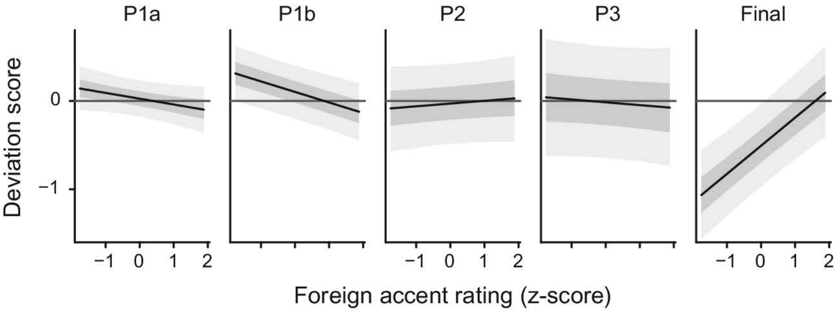


Fig. 13 Empirical deviation patterns by syllable type and proficiency level, including information about the variation among learners and statistical uncertainty. Error bands denote 50% and 90% uncertainty intervals.

A comparison of the prominence levels reveals that unstressed syllables in lexical words (P1a), syllables carrying lexical stress (P2), and syllables carrying lexical stress and the nuclear accent (P3) show near-horizontal trend lines around the target baseline. This suggests that these types of units were, on average, close to the native speaker profile with no pronounced changes of deviation patterns across proficiency levels.

Deviations from TL timing patterns are discernible for prominence levels P1b and in final contexts. Vowels in monosyllabic function words (P1b) show excess duration at the beginner stage, where they reflect a notable level of hyperarticulation. At high proficiency levels, on the other hand, we see target alignment. Final syllables show the greatest deviation from the target profile: Learners with low pronunciation ability exhibit a systematic lack of lengthening, which persists well into the intermediate stages.

While the patterns in Fig. 12 allow for direct comparison between the five syllable types, more detail is provided in Fig. 13, which adds information about statistical uncertainty. For each syllable type, 50% and 90% uncertainty intervals are added to the trend lines. The uncertainty bounds suggest that there is indeed scarce evidence for a sensitivity of P1a, P2, and P3 deviation patterns to proficiency level. The trends for P1b and final syllables, on the other hand, appear to be more robust.

7.5 *Discussion*

To summarize, we applied James' (1988) model to the acquisition of prominence variation by German learners to determine whether rhythmic timing patterns emerge in a bottom-up fashion. To this end, we adopted a simplified scheme of prosodic representation, which encodes strength asymmetries at the lexical, syntactic, and nucleus level, as well as between final and non-final syllables. These levels were assumed to form a hierarchy, which, in accordance with the LTD, allowed us to formulate expectations about how learners establish prosodic *s/w*-marking level by level. The hypothesized universal progression from the lowest to the highest level corresponds to a step-by-step adaptation of native speaker timing patterns. In empirical terms, this is reflected in a predictable decrease in local hypo- and hyperarticulation. Expected deviation trends were compared to the deviation scores recorded for the sample of German learners.

A level-by-level emergence of prominence grading is consistent with some, but not all, patterns in the present data. At a broad level of comparison, the empirical patterns agree with expectations in two regards. For one, the overall constellation is coherent with the fan-shaped predictions in Fig. 11. Second, there is a general match in the direction of deviation for syllable types P1a and b, which tend to be too long, and Final syllables, which are too short, on average. Lack of agreement between theoretical and predicted profiles is found for P3 syllables, which carry lexical stress and the nuclear accent. While the statistical uncertainty depicted in Fig. 13 suggests that caution should be exercised in interpreting this pattern, it appears that this syllable type does not show the type of hypoarticulation predicted by the LTD.

In line with the model, vowels at prominence level 1b were hyperarticulated by low-proficiency learners and converged with the native speaker target at the advanced stages. This suggests that monosyllabic function words are susceptible to overarticulation in GLE. P1a deviation patterns showed weak alignment with predictions. Bearing in mind the statistical uncertainty represented in Fig. 13, the early stages of L2 acquisition do show the expected divergence of unstressed syllables in lexical words: They are slightly overarticulated by lower-proficiency learners. The patterns in the data are also consistent with the LTD prediction of a delayed acquisition of P1b relative to P1a prominence backgrounding, suggesting tentatively that the acquisition of the two lowest levels may conform to the sequential order postulated by James' (1988) model.

Syllables carrying lexical and syntactic stress—that is, vowels at prominence level 2—are coherent with the hypothesized pattern in that they show no notable change across proficiency levels. Syllables carrying the nuclear accent, however, appear to violate LTD predictions. There is no evidence for hypoarticulation, that is, lack of lengthening, at lower proficiency levels. In fact, vowels in P3 contexts were relatively close to the target profile. Concerning the theoretical predictions, a delayed acquisition of P3 vowels therefore does not materialize in the present data set, indicating that P3 may not conform to the hypothesized level-by-level progression. We might be dealing with an instance of L1 transfer since German, like English, shows accentual lengthening.

The durational marking of final syllables yields the greatest discrepancy between learners and native speakers. The direction of deviation is in accordance with LTD predictions. In comparison to the four prominence levels, the magnitude of divergence from native speech is striking. Note that the theoretical predictions shown in Figs. 9 and 11 encode the simplistic assumption that prosodic heads and edges show the same extent of relative lengthening. More generally, these graphs suggest identical temporal effects of prominence grading across all levels. This arbitrary assumption merely served illustrative purposes. Upon reflection, the amount of accentual and final lengthening could have relied on empirical evidence reported in Sect. 2. This is to say that specific values for durational fore- and backgrounding could have been chosen based on findings in earlier work (e.g. Delattre, 1965; Li & Post, 2014). What matters most, however, is the relative position of syllable types.

Returning to the empirical results, the deviation patterns show that there is notable durational hypoarticulation in final syllables. It appears that L1 transfer may not be able to fully account for this finding, as empirical studies have observed comparable levels of final lengthening in English and German. Based on the logic of Major's (2001) model, we would look to universal constraints as a possible explanation. James' account of L2 acquisition in fact offers a U-perspective, as the model makes no allowance for L1 effects in the bottom-up emergence of timing patterns. The present data suggest that prosodic boundary marking may be considered an area of L2 prosody where universal forces toward prominence leveling operate.

8 General Discussion

Our exploration of speech rhythm in GLE from the viewpoint of Major's (2001) OPM and James' (1988) LTD has demonstrated that neither model offers a satisfactory account of rhythmic acquisition in this population of L2 speakers. Nevertheless, much can be learned from the above exercise, both from a methodological and a theoretical viewpoint.

Arguably, the inaccuracy of OPM predictions is rooted in the adopted conceptual approach to speech rhythm. Guided by existing quantitative research, the notion of rhythm was captured by means of rhythm metrics, which dictated a coarse approach to the subject matter. Thus, the line of argumentation proceeded along the broad notions of prominence variation, unequal timing, and durational variability as descriptive cover terms for a complex set of features. Accordingly, 'stress-timing' was treated as a category of speech that can, as a whole, be described in terms of markedness. This abstract level of analysis glossed over concrete, lower-level surface phenomena that can be described more transparently, both theoretically and empirically. This lack of transparency is carried forward to the metrics that were applied for an operationalization of rhythm in learner speech. For the field of L2 speech rhythm, however, metrics offer only a limited amount of information. Thus, if such quantities suggest a mismatch between native and non-native speech, the next step, of course, is to explore the nature of this mismatch. This gives rise to questions of where learners show hyperarticulation and/or hypoarticulation and whether deviations from the target are systematic in nature. We should bear in mind that one of the unique features of our field of inquiry, L2 phonology, is that it can provide answers to these questions. In contrast to typological approaches, from which the use of rhythm metrics originates (e.g. Ramus et al., 1999), L2 research can compare non-native to native speech directly, through controlled elicitation of the same utterances. This allows us to make targeted comparisons between segments, intervals, and syllables, which allow us to uncover the local components of prominence leveling. In short, it seems that the questions that are of direct interest to our field cannot be answered by rhythm metrics; arguably, we must move beyond the abstract and general level of description they offer.[10] Critics may correctly note that controlled data of the type elicited in this study is at odds with the investigation of more natural speaking styles. We thus face a trade-off between more vs. less natural speech on the one hand, and more vs. less informative descriptions on the other. It would seem that the latter dichotomy should receive greater weight as a criterion for methodological decisions.

We have seen that the LTD shifts the researcher's attention to the individual units of analysis and thereby points to analysis strategies that maximize the information that can be extracted from a set of measurements. This offers a more fine-grained account of prominence variation in speech and allows us to address the questions raised by differences in rhythm metrics. Thus, by opting for deviation scores as the quantities of interest, the present study shifted attention to local hyper- and hypoarticulation patterns in learner speech. Of course, this approach is very much in line with large parts of the current literature on speech rhythm, which adopts a componential view of this category of speech and would therefore proceed along similar lines.

From a meta-theoretical viewpoint, a unique feature of James' model is its predictive adequacy, that is, its ability to generate precise and informative predictions that can be falsified by data. This contrasts quite dramatically with our implementation of the OPM, which yielded fuzzy links between theory and predicted patterns. As for the LTD, the hypothesized universal bottom-up progression translates into quantitative predictions about local deviations from native speaker timing patterns. These predictions were highly informative in the sense that their partial falsification adds to our understanding of the L2 acquisition of rhythm. Thus, even though the LTD fails to account fully for the observed data, the weak spots we may have uncovered nevertheless offer valuable insights. They take us one step further by highlighting which parts of the assumed acquisition mechanism we may maintain as viable explanations and which ones may need revision or enrichment by other processes. After all, the LTD takes a bold stance in disregarding the possibility of L1 transfer in this bottom-up progression.

In terms of the constraints underlying the acquisition of English speech rhythm by L1 German learners, our application of the OPM and the LTD has left us with conflicting evidence for L1 transfer in L2 timing patterns. On the one hand, we interpreted the global information provided by rhythm metrics as perhaps suggestive of a limited or non-existent facilitative effect of L1 transfer. In contrast, the LTD perspective led us to refine this view, as deviation patterns at specific levels of prosodic representation appear consistent with L1 transfer. While the role of L1 transfer in the acquisition of speech rhythm remains to be explored more fully in future studies, it seems that a combination of the two models may offer fruitful perspectives. Thus, future work could reconsider the interplay of L1, L2, and U at the level of different syllable types or different levels of prosodic representation.

[10] The same is true, of course, for comparisons of accents or varieties of the same language. These remarks therefore also apply to the investigation of rhythm in World Englishes.

A final point that deserves to be mentioned concerns the added value of the necessarily broad notion of language universals for collaborative, cumulative efforts in the field of speech rhythm research. Both models stipulate the existence of constraints independent from the L1 and L2, that is, the notion of U (OPM) and that of a bottom-up construction of prosodic representation (LTD). Due to their very nature, such cross-lingual tendencies (or language universals) establish common ground for the study of speech rhythm across different languages. Theoretical and empirical work can contribute to a shared knowledge base about universal constraints on the acquisition of prominence grading. Research in the SLA and the World Englishes paradigm, for instance, can mutually inform each other and consequently draw on a larger body of theoretical knowledge and empirical evidence.

Appendix 1

Empirical evidence on vocalic timing patterns in English, German and Spanish: Rhythm metrics based on vocalic intervals and durational measurements

Language	%V	nPVI-V	VarcoV	*n*	Style	Reference
German	46	60		1	Reading passage	Grabe & Low 2002
	43			7	Reading passage	Dellwo & Wagner, 2003
	42	53		13	Free	Russo & Barry, 2008
	41	53	52	8	Sentences[1]	Arvaniti, 2012
	38	54	51	8	Reading passage	Arvaniti, 2012
	42	54	55	8	Free	Arvaniti, 2012
	42	45	41	5	Sentences	Li & Post, 2014
	42	45	41	5	Sentences	Li & Post, 2014
British English	38	73	64	6	Sentences	White & Mattys 2007
	41	55	55	8	Sentences[2]	Prieto et al., 2012
	38			3	Reading, retelling	Gut, 2005
		60		1	Reading, retelling	Gibbon & Gut 2001
	41	62	61	9	Semi-free	Payne et al., 2011
		76	61	10	Semi-free	Ordin & Polyanskaya, 2015
	41	57		1	Reading passage	Grabe & Low 2002
American English	44	56	50	8	Sentences[1]	Arvaniti, 2012
	44	54	50	8	Reading passage	Arvaniti, 2012
	48	59	66	8	Free	Arvaniti, 2012
	46	52	47	5	Sentences	Li & Post, 2014
		52		20	Free	Thomas & Carter, 2006

(continued)

(continued)

Language	%V	nPVI-V	VarcoV	*n*	Style	Reference
English (variety unspecified)	38			10	Sentences	Low et al., 2000
	42			5	Reading passage	Dellwo & Wagner, 2003
	40			4	Sentences	Ramus et al., 1999
	43	44	57	7	Reading passage	Dellwo et al. 2009
Spanish	44			4	Sentences	Ramus et al., 1999
	48	36	41	6	Sentences	White & Mattys 2007
	49	48	57	8	Sentences[1]	Arvaniti, 2012
	49	45	47	8	Reading passage	Arvaniti, 2012
	50	47	66	8	Free	Arvaniti, 2012
	48	44	50	6	Semi-free	Payne et al., 2011
	47	37	36	8	Sentences[2]	Prieto et al., 2012
	51	30		1	Reading passage	Grabe & Low 2002

Notes [1]Uncontrolled condition; [2]mixed condition (see Arvaniti, 2012 for details)

Appendix 2

Materials used in the reading task. Utterances used for the analysis are printed in bold.

1. Do you want to drink something?

   Oh, yeah. Can I get another cup of tea, please?

2. Where is your friend Peter from?

   Peter? **He's from the north of Germany**.

3. Is your brother home?

   No. He said that he's going to be back at eight o'clock.

4. Oh no! We haven't got any sugar. I want to bake a cake.

   No problem. **I can get some sugar from the market**.

5. Is Sally from England?

   She lives in England. But she was born in America.

6. Can I walk to the city centre from here?

   It's too far. **You must take the bus to the city centre**.

7. How was your trip to England?

   Great! **The weather was sunny and we took a lot of pictures.**

8. Where is the car?

   I parked it at the end of the street.

9. What is your sister doing at the moment?

   Becky? **She's writing an article for the school magazine**.

10. How can I help you?

    Can you tell me the way to the cinema, please?

11. Did you like the book?

    I did. **But the second part of the book was better than the first**.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.

Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology 1: Production* (pp. 227–256). Academic Press.

Aoyama, K., & Guion, S. G. (2007). Prosody in second language acquisition: Acoustic analyses of duration and F0 range. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning* (pp. 281–297). John Benjamins.

Archibald, J. (1994). A formal model of learning L2 prosodic phonology. *Second Language Research, 10*, 215–240.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*, 351–373.

Barry, W. J. (2007). Rhythm as an L2 problem: How prosodic is it? In J. Trouvain & U. Gut (Eds.), *Non-native prosody: Phonetic description and teaching practice* (pp. 97–120). De Gruyter.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 171–206). York Press.

Blasdell, R., & Jensen, P. (1970). Stress and word position as determinants of imitation in first-language learners. *Journal of Speech and Hearing Research, 13*, 193–202.

Boersma, P. (1998). *Functional phonology*. Holland Academic Graphics.

Boersma, P., & Weenink, D. (2014). *Praat: Doing phonetics by computer*. Version 5.3.68. http://www.praat.org/.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 273–304). York Press.

Bolinger, D. L. (1965). Pitch accent and sentence rhythm. In D. Bolinger, I. Abe, & T. Kanekiyo (Eds.), *Forms of english: Accent, morpheme, order* (pp. 139–180). Harvard University Press.

Borzone de Manrique, A. M., & Signorini, A. (1983). Segmental duration and rhythm in Spanish. *Journal of Phonetics, 11*, 117–128.

Brown, C. A. (1998). The role of the L1 grammar in the acquisition of segmental structure. *Second Language Research, 14*, 139–193.
Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English-speaking 4- and 5-year-old children. *Journal of Speech, Language, and Hearing Research, 50*(4), 999–1014.
Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer.
Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics, 29*(3), 489–534.
Colantoni, L., Steele, J. & Escudero, P. (2015). *Second language speech: Theory and practice*. Cambridge University Press.
Cruttenden, A. (1979). *Language in infancy and childhood: A linguistic introduction to language acquisition*. Manchester University Press.
Dauer, R. (1983). Stress-timing and syllable-timing reanalysed. *Journal of Phonetics, 11*, 51–62.
De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition 10*(1), 7–21.
Delattre, P. C. (1965). *Comparing the phonetics features of english, German, Spanish and French*. Groos.
Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. In D. Recasens, M. Solé & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (ICPhS 2003), (pp. 471–474). Universitat Autònoma de Barcelona.
Dellwo, V., Gutiérrez Díez, F., & Gavaldà, N. (2009). The development of measurable speech rhythm in Spanish speakers of English. In *Proceedings of XI Simposio Internacional de Comunicacion Social* (pp. 594–597). Santiago de Cuba.
Dziubalska-Kołaczyk, K. (1990). *A theory of second language acquisition within the framework of natural phonology*. AMU Press.
Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*, 315–330.
Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13*, 23–41.
Fasold, R. W., & Preston, D. R. (2007). The psycholinguistic unity of inherent variability: Old Occam whips out his razor. In R. Bayley & C. Lucas (Eds.), *Sociolinguistic variation: The-ories, methods, and applications* (pp. 45–69). Cambridge University Press.
Fahrmeir, L., Kneib, T., Lang, S. & Marx, B. (2013). *Regression: Models, methods and applications*. Springer.
Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.
Flege, J. E., & Bohn, O.-S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition, 11*, 35–62.
Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (pp. 521–602). Wiley-Blackwell.
Fuchs, R. (2016). *Speech rhythm in varieties of English: Evidence from educated Indian english and British english*. Springer.
Gatbonton, E. (1978). Patterned phonetic variability in second-language speech: A gradual diffusion model. *Canadian Modern Language Review, 34*, 335–347.
Giegerich, H. J. (1992). *English Phonology*. Cambridge University Press.
Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. In P. Dalsgaard, B. Lindberg, H. Benner, & Z. Tan (Eds.), *Proceedings of Eurospeech 2001* (pp. 91–94). Aalborg, Denmark.
Grabe, E., Gut, U., Post, B., & Watson, I. (1999). The acquisition of rhythm in english, French and German.In *Current research in language and communication: Proceedings of the Child Language Seminar* (pp. 156–62). London: City University.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In N. Werner & C. Gusshoven (Eds.), *Papers in Laboratory Phonology 7* (pp. 515–546). Mouton de Gruyter.
Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.
Gut, U. (2005). Nigerian english prosody. *English World-Wide, 26*(2), 153–177.
Gut, U. (2006). Unstressed vowels in non-native German. In Rüdiger Hoffmann & Hansjörg Mixdorff (Eds.), In *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden, Germany.
Hancin-Bhatt, B. J. (1994). Segment transfer: A consequence of a dynamic system. *Second Language Research, 10*, 241–269.
Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. Fry, P. MacCarthy, N. Scott & J. Trim (Eds.), *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday, 12 September 1961*, 73–84. London: Longmans.
James, A. R. (1986). *Suprasegmental phonology and segmental form*. Tübingen: Niemeyer.
James, A. R. (1988). *The acquisition of a second language phonology*. Tübingen: Narr.
James, A. L. (1940). *Speech signals in telephony*. Pitman & Sons.
Kaltenbacher, E. (1998). Zum Sprachrhythmus des Deutschen und seinem Erwerb. In H. Wegener (Ed.), *Eine zweite Sprache lernen* (pp. 21–38). Narr.
Kleber, F., & Klipphahn, N. (2006). An acoustic investigation of secondary stress in German. *Arbeitsberichte des instituts für phonetik Und digitale sprachverarbeitung der universität Kiel, AIPUK, 37*, 1–18.
Kohler, J. (1995). *Einführung in die Phonetik des Deutschen*. Erich Schmidt.
König, E., & Gast, V. (2009). *Understanding english-German contrasts*. Erich Schmidt.
Lado, R. (1957). *Linguistics across cultures*. University of Michigan Press.
Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm. Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition, 36*, 223–255.
Low, E. L. & Esther Grabe. (1995). Prosodic patterns in Singapore English. *Proceedings of the 13th ICPhS*, 636–639. Stockholm, Sweden.
Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterization of speech rhythm: Syllable-timing in Singapore english. *Language and Speech, 43*(4), 377–401.
Maack, A. (1959). Der Einfluss der Betonung auf die Lautdauer deutscher Sonanten. *Zeitschrift Für Phonetik, 3*, 341–356.
Machač, P., & Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Epocha Publishing House.
Maddieson, I. (2013). Syllable Structure. In M. S. Matthew & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/12, Accessed on 2016–12–14.)
Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Erlbaum.
Major, R. C., & Kim, E. (1996). The similarity differential rate hypothesis. *Language Learning, 46*, 465–496.
McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC Press.
Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 756–766.
Nishihara, T., & van de Weijer, J. M. (2012). On syllable-timed rhythm and stress-timed rhythm in World Englishes: Revisited. *Bulletin of Miyagi University of Education, 46*, 155–163.
Ordin, M., Polyanskaya, L., & Ulbrich, C. (2011). Acquisition of timing patterns in second language. In P. Cos, R. de Mori, G. di Fabbrizio & R. Pieraccini (Eds.), *Proceedings of Interspeech 2011* (pp. 1129–1132). Florence, Italy.
Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System, 42*, 244–257.

Ordin, M., & Polyanskaya, L. (2015). Perception of speech rhythm in second language: The case of rhythmically similar L1 and L2. *Frontiers in Psychology, 6*, 316.

Payne, E., Post, B., Astruc, L., Prieto, P., del Mar, M., & Vanrell. (2011). Measuring child rhythm. *Language and Speech, 55*(2), 203–229.

Pike, K. L. (1945). *The intonation of American english.* Michigan University Press.

Prieto, P., Vanrell, M. M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm: Evidence from Catalan, English and Spanish. *Speech Communication, 54*(6), 681–702.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*, 265–292.

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. In D. Recasens, M. Solé amp; J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (ICPhS 2003), (pp. 337–342). Universitat Autònoma de Barcelona.

Risley, T. R., & Reynolds, N. J. (1970). Emphasis as a prompt for verbal imitation. *Journal of Applied Behavior Analysis, 3*, 185–190.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In David Crystal (Ed.), *Linguistic controversies. Essays in linguistic theory and practice*, (pp. 73–79). Arnold.

Russo, M., & Barry, W. J. (2008). Isochrony reconsidered: Objectifying relations between rhythm measures and speech tempo. In A. Barbosa, S. Madureira & C. Reis (Eds.), *Proceedings of Speech Prosody 2008*, 419–422.

Schmid, S. (1997). The naturalness differential hypothesis: Cross-linguistic influence and universal preferences in interlanguage phonology and morphology. *Folia Linguistica, 31*(3–4), 331–348.

Sönning, L. (2020). *Phonological variation in German Learner English.* University of Bamberg Dissertation. https://doi.org/10.20378/irb-49135.

Sönning, L. (2022). Speech rhythm in German Lerner English: Dataset for Soenning 2022 "(Re-) viewing the acquisition of rhythm in the light of L2 phonological theories", https://doi.org/10.18710/GTI2BR, Dataverse NO, DRAFT VERSION. V1.

Thomas, E. R., & Carter, P. M. (2006). Prosodic rhythm in African American English. *English World-Wide, 27*(3), 331–355.

Vanderslice, R., & Ladefoged, P. (1972). Binary suprasegmental features and transformational word accentual rules. *Language, 48*, 819–839.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432

Vihman, M. M., Nakai, S., & DePaolis, R. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, D. Whalen, & C. Best (Eds.), *Laboratory Phonology 8: Phonology and phonetics* (pp. 343–368). De Gruyter.

Wenk, B., & Wiolland, F. (1982). Is French really syllable-timed? *Journal of Phonetics, 10*, 193–216.

Wesener, T. (1999). The phonetics of function words in German spontaneous speech. In K. Kohler (Ed.), *Phrase-level phonetics and phonology* (pp. 327–377). Universität Kiel.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*(4), 501–522.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language, 66*, 665–679.

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication, 63–64*, 38–54.