L2 RHYTHM EFFECTS ON INTELLIGIBILITY

Sarah Waldmann

Freie Universität Berlin, Germany
sarah.waldmann@fu-berlin.de

ABSTRACT

While most scholars agree that prosodic errors can significantly affect the intelligibility of non-native speech (L2), it remains unclear to what extent this applies for speech rhythm. To find out which rhythm features correlate with intelligibility, the present study compares L2 speech of German learners with L1 Spanish and European Portuguese (EP), two languages with different rhythm shapes. Thus, rhythmic differences are expected for L2 speech as well. A first pilot study was carried out, recording L1 Spanish and EP speakers reading 12 true/false statements. Recordings were annotated for rhythm-related errors. Intelligibility indices (reaction time, dictation, scalar ratings) of the recordings were identified through an intelligibility experiment with 12 native German listeners. First results show that strongest correlations can be found between rating scores and syllable-based processes (vowel epenthesis/elisions, ΔC), which are also likely to lead to misunderstood items.

Keywords: Speech rhythm, intelligibility, SLA, Spanish, European Portuguese

1. INTRODUCTION

Prosodic features like rhythm and intonation are likely to be transferred from a source language (L1) into a target language (L2) [23]. Despite the important role of these features for comprehensible speech [2, 16, 17, 27, 34], few studies have systematically examined how rhythm is realized by L2 learners from controlled L1 backgrounds [11]. Even fewer studies to date have focused on German as target language.

Assuming that prosodic errors are caused by language transfer from an L1, and that they reflect properties of the L1 [22, 35], the present research aims to answer the following questions:

1. Which rhythmic properties of the L1 appear in L2 speech and lead to non-targetlike rhythmic realizations?
2. Which of these properties display the most significant influence on intelligibility?

The underlying concept of intelligibility in the present study is that of actual understanding of L2 speech on an acoustic-phonetic level [e.g. 8, 24, 26] and it is measured through a combination of different perception tasks: a true-false decision task (reaction time, accuracy [25]), transcription, and scalar ratings.

To answer the research questions, a small-scale quantitative pilot study on L2 speech of German learners with L1 Spanish and EP was designed. Despite wide structural similarities, these Romance languages differ in several rhythm-related features, and they also both differ from German (see 2.1.).

In order to measure rhythm properties objectively rather than on an impressionistic overall level, the investigation focuses on micro-level metrical, durational and phonological rhythm-related features (2.1.), including several rhythm metrics (2.2.).

2. THE NOTION OF RHYTHM

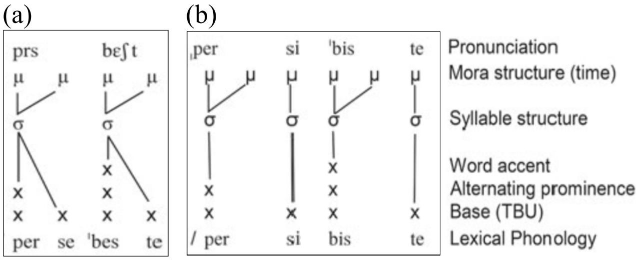
2.1. Rhythm-related features

While the well-known dichotomy of stress- vs. syllable-timed rhythm [1, 28] only takes into account the temporal organization of speech, Metrical Phonology [14, 19] deals solely with the organization of prominence through metrical algorithms, neglecting the dimension of time. Dufter [6] suggests that both time and prominence are crucial for the rhythm contour of a language, depending on functions of (post-)lexical phonology: If time is semantically distinctive (Japanese), spaces between prominences cannot be modified. On the contrary, if prominence is semantically distinctive (English, German), the dimension of time is phonologically less significant. Prominences can be isochronized (often interpreted as stress-timing). If neither time nor prominence are semantically distinctive, an alternating rhythm can be observed (Spanish, Italian).

As rhythm is the result of a performative speech act [31], rhythm shapes also depend on the context of performance. To a certain extent, performance can become a part of the language norm as is the case with Spanish and EP: While the lexical and morphological functions of time and prominence are almost the same in Spanish and EP, both languages differ strikingly when it comes to rhythm [7]. These differences are due to the selection of different metrical grids for the performance of full sonority (syllable, foot, phrase), which leads to different phonological processes [31].

Fig. 1 gives an example of these processes depending on the selection of the metrical grid. While in EP the selection of the second and third (but not the first) level of metrical strength leads to elision or reduction of vowels of low metrical strength and therefore to complex syllable structures, in Spanish all vowels are performed with full sonority. Contrary to EP, phonological processes in Spanish refer to the domain of the syllable level.

Figure 1: Realization of *percebeste* (you perceived) in EP (a) and *percibiste* in Spanish (b) [31:415]



Considering metrical algorithms, temporal organizations, and phonological processes as linked to rhythmic realization, the following rhythm-related features can be deduced: placement of word accent (WA), phonological and durational marking of WA (linked to syllable complexity and vowel restrictions), vowel reduction in unstressed positions, and syllabication processes. Regarding these features, German, Spanish, and EP show clear differences:

Table 1: Rhythm features of German, Spanish, EP

Feature	German	Spanish	EP
WA placement	free, complex rules	free within 3-syllable-window	free within 3-syllable-window
Complex syllables	+	-	(+)
Vowel reduction	+	-	+

In German, a prominence-based language [6], several features underline the auditory salience of prominences (marking both prominence and non-prominence). Spanish lacks vowel reduction and shows primary and secondary word accent [15], while in EP, vowel reduction and elision processes strengthen the salience of primary accents at the cost of secondary prominence and lead to a higher syllable complexity in spoken language [10]. Based on these differences and following [22], the L1 EP learners of German (EP_{GER}) are expected to show more targetlike realizations of reduction syllables and complex consonant clusters than L1 Spanish learners (SP_{GER}). It is assumed that both learner groups will have similar difficulties with accent placement (rightward shift).

2.2. Rhythm metrics

The attempt to quantify rhythm in speech has resulted in numerous rhythm metrics (RMs) during the past twenty years (see [36]). Based on the claim that rhythm is a surface phonetic property only [30] and is therefore reflected in the durations of vocalic and consonantal intervals, several algorithms have been suggested as measures of rhythmic shaping. For reasons of comparability, this study only includes the most common ones: proportion of vocalic material in speech (%V, [30], lower for languages with vowel reduction), consonantal/vocalic standard deviation (ΔC, ΔV, [30], higher for languages with vowel reduction), variability in vocalic/consonantal duration (Varco-V, Varco-C, [5]) and the normalized Pairwise Variability Index (nPVI, [12, 20]). Although several studies questioned the validity of these measures (cf. [3, 13]), numerous works on rhythm consider them as valid (cf. [9, 18, 36]). The present study suggests that they can serve as a supplement to the phonological and durational features mentioned in 2.1. Based on the comparison in Tab. 1, the RMs for EP_{GER} utterances are expected to be closer to L1 German (GER_{GER}) utterances than SP_{GER}.

3. METHODS

3.1. Speech samples

To test the adequacy of the stimuli and the perception task, a small-scale pilot study with speech samples of 4 German learners (2 SP_{GER}, 2 EP_{GER}, level B1, age 23–33) and a control set of 2 native German speakers (age 26–30) was performed. A set of 12 read-aloud true/false statements per speaker was recorded (e.g. (1), (2), total n=72). Each statement has a similar information structure and contains 10 syllables, ≥1 reduction syllable, ≥1 consonant cluster. The small number of speakers is due to the pilot character of the study and will be increased in future research. Nevertheless, results are expected to provide meaningful insights into rhythmic transfer in the I-grammars of the analyzed speakers.

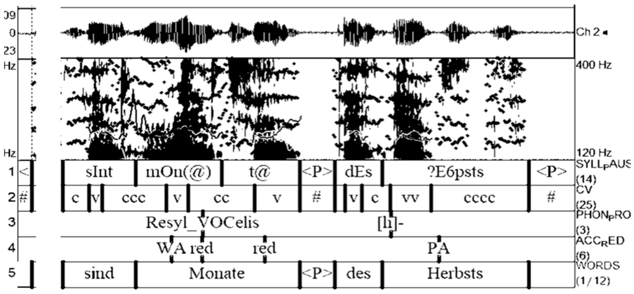
(1) *Benzin ist ein exzellentes Getränk.*
Gasoline is an excellent drink.
(2) *Schnellstraßen verursachen großen Lärm.*
Highways cause a lot of noise.

All 72 samples were analyzed in Praat [4] for rhythm-related features and non-linguistic factors like speech rate and pauses, which affect rhythm [29] and are linked to intelligibility and fluency [33]. Each item was annotated on 5 different tiers: syllables and pauses, consonantal/vocalic intervals, syllabication and segmental processes, accentuation pattern (WA, reduction syllables), and transcription (cf. Fig. 2). Speech rate and syllable durations were extracted in Praat, and RMs were calculated with Correlatore [21].

Table 2: Rhythm-related features used in this study

Feature	Description	Tier
Speech rate	Syllables/second	1
Pauses	Pauses/item (n°, durations)	1
Accent. patterns	Reduction syllables, word/phrase accent placement and shifts	3,4
Syll. struct.	Vowel/consonant elision/epenthesis	3
%V	Proportion of vocalic intervals	2
ΔV, ΔC	Standard deviation of voc./cons. intervals	2
Varco-V, Varco-C	Variability in vocalic/consonantal interval duration	2
nPVI-V, nPVI-C	Normalized durational differences between adjacent vocalic/consonantal intervals	2

Figure 2: Segmental and prosodic annotation in Praat (extract, EP_{GER} speaker)



3.2. Perception task and data analysis

A multimodal perception task with 12 listeners (L1 German) was carried out. For each listener, a randomized set of 12 items (6 true/6 false statements) was created. Each set contained 4 items per L1 background and 2 items per speaker. Correlations were calculated on the basis of 216 observations. Data was gathered via a ShinyApp tool scripted in RStudio [32]. This tool allows collecting multiple perception data and can be customized to specific experimental designs. Participants listened to each statement and decided as fast as possible whether it was true or false. Response accuracy and reaction time (RT) between end of audio file and clicking 'right/wrong' were measured. Subsequently, listeners rated the intelligibility of each item on a 7-point Likert scale (1=not intelligible, 7=perfectly intelligible). Finally, listeners transcribed the items as understood.

For all rhythm-related features and intelligibility indices, univariate correlations were analyzed in R, using scatterplots in combination with Pearson's coefficient. Furthermore, linear regression models were fitted to the data in order to analyze the coefficients' significance according to the corresponding p-values. Listeners' transcriptions were used for a qualitative error analysis.

3.3. Rhythmic properties of stimuli for perception task

Clearest differences between native and non-native speech showed up for speech rate (highest for GER_{GER} speakers) and pauses (overall pauses for GER_{GER}=0). In terms of accentuation, EP_{GER} speakers displayed the same number of rightward WA shifts (n=5), that mostly could be traced back to transfer of L1 accentuation as in [*ʃɪmpənˈze]. Vowel elisions appeared three times more frequently in GER_{GER} speech (n=18) than in utterances of both learner groups (n=6), with SP_{GER} speakers displaying vowel elisions only in reduction syllables and EP_{GER} eliding vowels in reduction and neutral (unstressed, not reduced) syllables (e.g. /Monate/ > *[mɔntə] instead of [mo:natə]). Vowel epenthesis did not occur in GER_{GER} speech, but it did for SP_{GER} (n=3) and EP_{GER} (n=2). Consonant elisions were found much more frequently in SP_{GER} (n=13) than in EP_{GER} (n=3), but GER_{GER} speakers also showed several instances of consonant elisions (n=8), always in word-final positions of unstressed function words.

RMs, however, did not reflect rhythmic differences as predicted: Contrary to expectations, GER_{GER} utterances showed a higher %V and a lower ΔC than those of SP_{GER} and EP_{GER} (Tab. 3). ΔV was found to be slightly higher in utterances of EP_{GER} and almost nativelike for SP_{GER}. A further surprising observation was that SP_{GER} showed highest nPVI-C values (and lowest for GER_{GER}), while the opposite was expected. Neither for Varco-C nor Varco-V significant differences between speaker groups were observed.

Table 3: Rhythm metrics of speaker groups (mean)

Rhythm metric	GER_{GER}	SP_{GER}	EP_{GER}
%V	41.67%	38.98%	38.73%
ΔC	78.98	97.90	103.84
ΔV	57.25	57.92	61.91
nPVI-C	61.94	69.23	67.64

4. RESULTS

4.1. Influence on reaction time and response accuracy

Standard RT was slightly shorter for utterances of GER_{GER}, and almost the same for SP_{GER} and EP_{GER}. While neither %V nor ΔV correlated significantly with RT, non-linguistic features like speech rate and pauses showed the strongest correlations (see Tab. 4). Furthermore, a lower ΔC correlated significantly with a faster standard RT. The same occurs for consonant elisions: The fewer the consonant elisions, the faster the RT. None of the other features (vowel elision/ epenthesis, WA shift) showed a significant correlation with RT. Response accuracy, however, did not correlate with any of the investigated features, which may also be due to the small sample size of listeners.

Table 4: Significant correlations between rhythm-related features and standard RT

Feature	Pearson's	p-value
Speech rate	-0.35	0.00
Pauses	0.19	0.03
C-elision/ epenthesis	0.17	0.04
ΔC	0.25	0.00

4.2. Influence on rating scores

While speech rate and pauses seem to have the strongest effects on rating scores (RSs), vowel processes also correlate with listeners' ratings: more elisions are correlated with higher RSs, whereas vowel epenthesis leads to lower RSs (see Tab. 5). This indicates that RT and perceived intelligibility are influenced differently by syllable-structure-based processes. Surprisingly, a lower ΔC correlated significantly with high RSs, which may be due to the low ΔC values found in GER_{GER} utterances.

Table 5: Significant correlations between rhythm-related features and RSs

Feature	Pearson's	p-value
Speech rate	0.65	0.00
Pauses	-0.43	0.00
V-elision/epenthesis	-0.26	0.01
ΔC	-0.33	0.00

The different intelligibility indices also inter-correlated significantly ($p<0.001$), with the Pearson's for RSs and response accuracy being the highest (0.57), followed by RSs and RT (-0.41), and response accuracy and RT (-0.32).

4.3. Qualitative transcription error analysis

A qualitative analysis of transcription errors (ex. 3–5) confirms the findings of 4.2 that inappropriate vowel elision (3) or epenthesis (4), as well as wrong WA placement (5) can lead to misunderstandings:

(3) *Benzin ist ein *Exzellenz Getränk*
Gasoline is a drink of excellence.
(4) **Schnelle Straßen verursachen großen Lärm*.
Fast streets cause a lot of noise.
(5) *Ein *Schintersee ist eine Vogelart.*
A shinter sea [nonexistent] is a type of bird.

In (3), an inappropriate vowel elision in the reduction syllable [təs] > [ts] in *exzellentes* (EP_{GER}) leads to the perception of a compound noun (*Exzellenzgetränk*) instead of 'adjective + noun'. In (4), a vowel epenthesis in the onset cluster [ʃtʀ] in *Straßen* (realized by a SP_{GER} speaker) leads to the perception of 'adjective + noun' instead of a compound noun. Both inappropriate vowel realizations and wrong WA placement can be traced back to properties of the respective L1: In EP, the last syllable of Portuguese *excelentes* is represented on the first level of metrical strength and therefore syllabified as coda of the preceding syllable: [ʃsəlɛntʃ]. This syllabication process seems to be transferred to the L2. In Spanish, the onset cluster [ʃtʀ] is phonotactically not allowed and normally dissolved by vowel epenthesis (cf. Spanish *estructura* 'structure'), which in (4) is transferred to the L2. In (5) we observe the effect of WA placement transfer on intelligibility: While in German *Schimpanse* is pronounced [ʃɪmˈpanzə], the EP_{GER} speaker realizes it with L1 accentuation patterns [*ʃɪmpənˈze]. The WA shift causes further vowel processes (reduction of stressed syllable /pan/, full sonority of reduction syllable /se/) which lead to misperception (**Schintersee*). Contrary to (3) and (4), the processes in (5) cannot be shown by rhythm metrics, as vocalic intervals stay the same but are placed differently.

5. DISCUSSION

These results show that there are different rhythmic properties of the L1 that lead to non-targetlike rhythmic realization of the L2. However, rhythm metrics only give a blurry idea about how L1 properties influence the rhythmic shape of L2. When it comes to the influence of rhythmic features on intelligibility, they provide even fewer insights. A qualitative analysis of misunderstood items (4.3.) gives first hints on why this may be the case: Syllabication processes and accent placement play a crucial role for word (and word boundary) identification, but they occur 'underneath' the phonetic surface. Therefore, they do not always manifest in rhythm metrics. At the same time, accent placement should be considered on a qualitative level rather than on a purely quantitative one, as consequences of wrong accent placement are not always the same (cf. *Schokolade* 'chocolate' > [ʃokoˈla:də] vs. *[ˈʃokoladə] which does not lead to major changes in syllable structure and/or vowel quality). However, the significant correlation between syllable-structure-based processes and both reaction time and rating scores suggest that future research should put a special focus on these phonological processes.

Concerning the pilot study's research design, several major adjustments are suggested for a large-scale study that is currently being carried out: According to some listeners' comments, several misunderstood items were repaired through semantic frames. Therefore, true/false statements will be replaced by semantically nonsensical sentences.

6. REFERENCES

[1] Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh: Univ. Press.

[2] Anderson-Hsieh, J., Johnson, R., Koehler, K. 1992. The Relationship Between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure. *Lang. Learn.* 42, 529–555.

[3] Arvaniti, A. 2012. The usefulness of metrics in the quantification of speech rhythm. *J. Phon*. 40, 351–373.

[4] Boersma, P., Weenink, D. 2017. Praat. Doing phonetics by computer, http://www.praat.org

[5] Dellwo, V. 2006. Rhythm and Speech Rate: A variation coefficient for deltaC. In: P. Karnowski, I. Szigeti (eds.), *Sprache und Sprachverarbeitung: Akten des 38. Linguistischen Kolloquiums in Piliscsaba 2003*. Frankfurt: Lang, 231–241.

[6] Dufter, A. 2003. *Typen sprachrhythmischer Konturbildung.* Berlin: De Gruyter.

[7] Dufter, A., Reich, U. 2003. Rhythmic Differences within Romance: identifying French, Spanish, European and Brazilian Portuguese. *Proc. 15th ICPhS* Barcelona, 2781–2784.

[8] Field, J. 2005. Intelligibility and the Listener: The Role of Lexical Stress. *TESOL Quarterly* 39, 399–423.

[9] Gabriel, C., Kireva, E. 2014. Prosodic Transfer in Learner and Contact Varieties: Speech Rhythm and Intonation of Buenos Aires Spanish and L2 Castilian Spanish Produced by Italian Native Speakers. *Stud. Second Lang. Acquis.* 36, 257–281.

[10] Frota, S., Vigário, M. 2001. On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Probus* 13, 247–275.

[11] Galaczi, E., Post, B., Li, A., Barker, F., Schmidt, E. 2017. Assessing Second Language Pronunciation: Distinguishing Features of Rhythm in Learner Speech at Different Proficiency Levels. In: Isaacs, T., Trofimovich, P. (eds.), *Interdisciplinary Perspectives. Second Language Pronunciation Assessment*. Bristol: Multilingual Matters, 157–182.

[12] Grabe, E., Low, E. L. 2002. Durational Variability in speech and the Rhythm Class Hypothesis. In: Gussenhoven, C., Warner, N. (eds.), *Phonology and Phonetics: 4-1. LabPhon 7*. Berlin, New York: Mouton de Gruyter, 515–546.

[13] Gut, U. 2012. Rhythm in L2 speech. *Speech and Language Technology* 14/15, 83–94.

[14] Hayes, B. 1981. *A metrical theory of stress rules*. PhD Thesis. Massachusetts Institute of Technology.

[15] Hualde, J. I. 2014. *The sounds of Spanish*. Cambridge: Cambridge University Press.

[16] Isaacs, T., Trofimovich, P. 2012. Deconstructing comprehensibility: Identifying the Linguistic Influences on Listeners' L2 Comprehensibility Ratings. *Stud. Second Lang. Acquis.* 34, 475–505.

[17] Kang, O., Rubin, D., Pickering, L. 2010. Suprasegmental Measures of Accentedness and Judgments of Language Learner Proficiency in Oral English. *Mod. Lang. J.* 94, 554–566.

[18] Kinoshita, N., Sheppard, C. 2011. Validating acoustic measures of speech rhythm for second language acquisition. *Proc. 17th ICPhS* Hongkong, 1086–1089.

[19] Liberman, M., Prince, A. 1977. On Stress and Linguistic Rhythm. *Linguistic Inquiry* 8, 249–336.

[20] Low, E. L., Grabe, E., Nolan, F. 2000. Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Lang. Speech* 43, 377–401.

[21] Mairano, P., Romano, A. 2010. Un confronto tra diverse metriche ritmiche usando Correlatore. In: Schmid, S., Schwarzenbach, M., Studer, D. (eds.), *La dimensione temporale del parlato*. Torriana: EDK Editore, 79–100.

[22] Major, R. C. 2008. Transfer in second language phonology: A review. In: Hansen, J. G., Edwards, M., Zampini, L. (eds.), *Phonology and second language acquisition*. Amsterdam: Benjamins, 63–94.

[23] Mennen, I., de Leeuw, E. 2014. Beyond Segments: Prosody in SLA. *Stud. Second Lang. Acquis.* 36, 183–194.

[24] Munro, M. J., & Derwing, T. M. 1995a. Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Lang. Learn.* 45, 73–97.

[25] Munro, M. J., Derwing, T. M. 1995b. Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech. *Lang. Speech* 38, 289–306.

[26] Munro, M. J., Derwing, T. M. 2015. A prospectus for pronunciation research in the 21st century: A point of view. *JSLP* 1, 11–42.

[27] Pickering, L. 2001. The Role of Tone Choice in Improving ITA Communication in the Classroom. *TESOL Quarterly* 35, 233–255.

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

[29] Prieto, P., Vanrell, M. d. M., Astruc, L., Payne, E., Post, B. 2012. Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Commun.* 54, 681–702.

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

[31] Reich, U., Rohrmeier, M. 2014. Batidas Latinas: On rhythm and meter in Spanish and Portuguese and other forms of music. In: Caro Reina, J. C., Szczepaniak, R. (eds.), *Syllable and word languages*, 391–420.

[32] RStudio Team. 2016. RStudio: Integrated Development for R. RStudio, Inc., Boston http://www.rstudio.com/

[33] Tajima, K., Port, R., Dalby, J. 1997. Effects of temporal correction on intelligibility of foreign-accented English. *J. Phon.* 25, 1–24.

[34] Trofimovich, P., Baker, W. 2006. Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech. *Stud. Second Lang. Acquis.* 28, 1–30.

[35] Waldmann, S. 2014. *Rhythmus und Fremdsprache*. Master thesis, University of Potsdam.

[36] White, L., Mattys, S. L. 2007. Calibrating rhythm: First language and second language studies. *J. Phon.* 35, 501–522.