Book chapter to be published: Ratree Wayland, Kevin Tang & Rahul Sengupta. In press. Acquisition of similar versus different speech rhythmic class. In Lars Meyer & Antje Strauss (eds.), *Rhythms of Speech and Language: Culture, Cognition, and the Brain*, Chapter 39. Cambridge University Press, Cambridge"

ACQUISITION OF SIMILAR VERSUS DIFFERENT SPEECH RHYTHMIC CLASS

Ratree Wayland[1,], Kevin Tang[1,2] and Rahul Sengupta[3]*

[1]Department of Linguistics, University of Florida, United States

[2]Department of English Language and Linguistics, Institute of English and American Studies, Faculty of Arts and Humanities, Heinrich Heine University Düsseldorf, Germany

[3]Department of Computer & Information & Engineering, University of Florida, United States

**Corresponding author: ratree@ufl.edu*

1. INTRODUCTION

Rhythm manifests in speech through alternating stronger and weaker elements, occurring at various levels of the prosodic hierarchy. However, the definition and quantification of rhythm in speech remain topics of ongoing debate. Various methodologies have been proposed to measure and categorize rhythmic patterns, reflecting a lack of consensus on the best approach.

According to the original theory of speech rhythm which posits that speech is structured into units of equal duration (i.e., isochrony) (Pike 1945; Abercrombie, 1967), the concept of a dichotomy between *syllable-timed* and *stress-timed* languages was proposed based on whether isochronous units were thought to be the syllables or the intervals between stressed syllables (For further discussion of the concept of rhythm classes, see also Rathcke, Chapter 14 and Fuchs, Chapter 25, this volume). However, empirical evidence did not support the isochrony principle (Roach, 1982, Dauer, 1983), leading to its dismissal. Nevertheless, the term *speech rhythm* and its associated language typology continue to be utilized because the systematic alternations of strong and weak elements at various phonological levels generate a sense of rhythm (Langus et al., 2017), based on several physical dimensions including intensity, duration, pitch, and vowel quality (Terken and Hermes, 2000).

Rather than seeking isochronous patterns, new rhythm measures aim to characterize subtler regularities across multiple dimensions (Bertinetto, 1989; Kohler, 2009; Cumming and Nolan, 2010; Turk and Shattuck-Hufnagel, 2013). For example, Dauer (1983) proposes correlating rhythmic classes with phonological differences among languages, such as syllable structure complexity or the presence of reduced vowels. This approach leads to the development of duration-based speech rhythm metrics defined by variability in the duration of consonantal or vocalic intervals (Ramus et al., 1999; Grabe and Low, 2002) (see also Fuchs, Chapter 25, this volume for a review of various types of rhythm metrics). However, despite their reported success in distinguishing prototypical syllable- and stress-timed languages, they are less robust against variations induced by speech rate and speaking styles (e.g., read story versus spontaneous speech), even when sentence structures are specifically designed to represent maximal *stressed-timed* and *syllable-timed* patterns (Arvaniti, 2009; Wiget et al., 2010; Arvaniti, 2012). These limitations underscore the need for new tools to reassess the acoustic foundations of speech rhythm.

Research into how the rhythm of a non-native language is learned has traditionally received less attention compared to how the sounds of a language, like vowels and consonants, are learned. An outstanding question is whether having a similar rhythmic structure in one's native language (L1) offers

an advantage in learning the rhythm of a second language (L2). The findings thus far seem to support the hypothesis that shared rhythmic properties between L1 and L2 may facilitate the acquisition of the rhythmic pattern of the new language and that the rhythm of L1 sticks, making learning a new, different rhythm, taxing. For example, German learners were found to achieve a level of durational variability resembling that of the target language (British English), while French learners exhibited lower variability compared to native British speakers, even at advanced proficiency levels (Ordin and Polyanskaya, 2015). However, as noted by van Maastricht et al. (2019), the analysis of two distinct L1-L2 pairings (German-English vs. French-English) raises the possibility that disparities in segmental, phonotactic, or prosodic characteristics between German and French could have influenced the observed variations.

To further test the hypothesis that shared rhythmic properties between L1 and L2 may facilitate the acquisition of the rhythmic pattern of the new language, this study examined the rhythm acquisition of different L1-L2 combinations, as well as within the same L1-L2 pair. Specifically, for the distinct L1-L2 combinations, we analyzed English rhythm acquisition by native speakers of German and French and German rhythm acquisition by native speakers of English and French. For the same L1-L2 combination, the acquisition of English and French by native German speakers was examined. Furthermore, unlike prior research where L2 proficiency is based on self-report or length of experience, L2 proficiency in this study is assessed based on the degree of L1-L2 acoustic distance, as described by Bartelds et al. (2020, 2022). Most importantly, unlike earlier studies that primarily focused on the durational variability of segmental intervals, this study measures temporal regularities using amplitude envelope modulation patterns corresponding to syllabic units and larger speech patterns, such as stressed-unstressed rhythms.

2. METHODS

2.1. STIMULI

The stimuli for this study were extracted from The BonnTempo Corpus (Dellwo et al., 2004). The L1 dataset consists of read speech recordings from 15 native speakers of German, 7 of English, and 6 of French. The L2 dataset encompasses recordings of German-accented English (N=8), German-accented French (N=8), French-accented English (N=2), English-accented German (N=3), and French-accented German (N=1). The selection of these languages is based on prior research indicating distinctions in speech rhythm among English, German, and French (Ramus et al., 1999; Loukina et al., 2011). Specifically, English and German are characterized as *stress-timed* languages, where the rhythm tends to be based on regular intervals between stressed syllables. French, on the other hand, is described as a *syllable-timed* language, where each syllable is perceived to have approximately the same duration.

For German speakers, a short German passage from a novel by Bernhard Schlink ('Selbs Betrug') of 76 syllables long served as reading material. This text was then translated into English (77 syllables) and French (93 syllables) for English and French speakers, respectively. The subjects were asked to become familiar with the text before reading in five reading rates: slowest, slow, normal, fast, and fastest. The passage was divided into 7 utterances and saved as separate files. The English version of the utterances are:

1. The next day, I went to Falmouth.
2. It is a voyage to the end of the world.
3. After Lincoln, the hills and woods become monotonous.
4. After Bristol, the town gets boring.

5. And near Saints Bury, the countryside becomes flat and monotonous.
6. If dissidents were banned in our country,
7. They would be banned to the Portishead bay.

All five versions of each utterance were then annotated according to phonological syllable durations and consonantal and vocalic intervals on two separate tiers using Praat software (Boersma, 2001).

2.2. L2 PROFICIENCY MEASURE

To estimate L2 proficiency, we utilized word-based pronunciation differences using self-supervised neural models as explained by Bartelds et al. (2022). In this approach, the Neural Acoustic Distance was computed for pairs of audio files, representing a reference speaker (L1) and a target speaker (L2). The calculation involved averaging the distances between corresponding tokens (words/subwords) that form the given sentence. This procedure was carried out for all combinations of audio from the L1 and L2 groups, considering a specific sentence spoken at a particular rate. It is important to note that the neural model representations are sensitive not only to differences in how individual speech sounds are produced (segmental differences) but also to capturing variations in speech melody (intonation) and timing (duration), as described by Bartelds et al. (2022).

Mean acoustic distances at all five speaking rates between L1 and L2 English, German, and French (for native German speakers only) are presented in Table 1. It is evident that the distance from L1 English is greater for French-accented English than for German-accented English. Similarly, French-accented German is less similar to German than English-accented German.

One-way ANOVAs confirm the statistical significance of these differences, indicating that native German speakers are significantly more proficient in English than French speakers [$F(1, 68) = 35.48$, $p < .001$], and that native English speakers are significantly more proficient in German than French speakers [$F(1, 68) = 64.81$, $p < .001$]. Additionally, the distinction between German-accented English and English, compared to German-accented French and French, was significant [$F(1, 68) = 7.89$, $p = 0.006$], indicating the superior English proficiency of German speakers over their French proficiency.

TABLE 1. ACOUSTIC DISTANCE

Means and standard deviations (SD) of acoustic distance between L1 vs L2 English and German.

Language-pair	*Speaking rate*	*Mean*	*SD*
English vs. German-accented English	slowest	2.95	0.05
	slow	2.83	0.04
	normal	2.77	0.08
	fast	2.72	0.08
	fastest	2.68	0.06
English vs. French-accented English	slowest	3.09	0.08
	slow	3.02	0.07
	normal	2.97	0.10
	fast	2.87	0.08
	fastest	2.83	0.06
German vs. English-accented German	slowest	3.03	0.10
	slow	2.90	0.12
	normal	2.84	0.11
	fast	2.75	0.08
	fastest	2.77	0.08
German vs. French-accented German	slowest	3.42	0.07
	slow	2.75	0.08
	normal	3.16	0.08
	fast	3.28	0.07

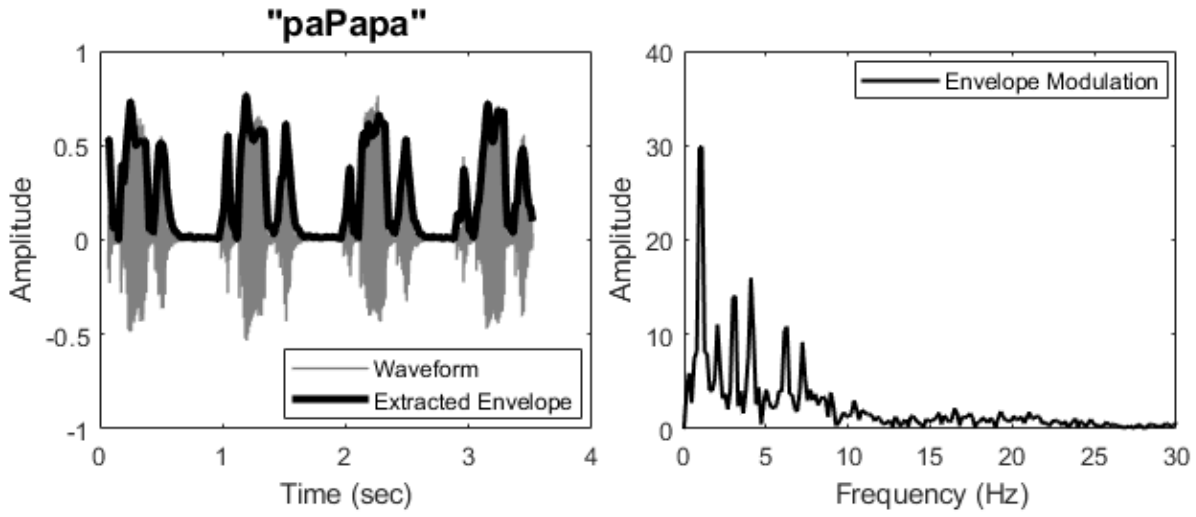
French vs. German-accented French	fastest	2.84	0.09
	slowest	3.04	0.06
	slow	2.94	0.05
	normal	2.89	0.06
	fast	2.76	0.05
	fastest	2.73	0.09

2.3. AMPLITUDE ENVELOPE MODULATION SPECTRUM (AEMS)

AEMS involves the spectral analysis of low-rate amplitude modulations within the envelope of the speech signal. The analysis covers the entire amplitude modulation spectrum as well as specific frequency bands of amplitude modulations. Fourier analysis is applied to the speech envelopes to identify the dominant amplitude modulation rate (Figure 1) (for different metrics, see Braun, Chapter 12, Zhang et al, Chapter 16, and Fuchs, Chapter 25, this volume).

In Figure 1, the left graph illustrates the amplitude envelope of a speech waveform (composed of 4 repetitions of /paPApa/ tri-syllabic non-sense words). This envelope captures temporal fluctuations in amplitude, including those corresponding to syllabic patterns. Regular patterns such as stressed-unstressed rhythms are also discernible. These regularities can be quantified by subjecting the envelope to Fourier analysis, resulting in a depiction of the dominant amplitude modulation rates present in the signal, as shown in the right graph. Note that the highest energy peak (in decibels) occurs at 1 Hz, with additional peaks observable at higher rates.

FIGURE 1.
The left graph shows the amplitude normalized waveform and the amplitude envelope (dark solid line) of a male saying "paPapa" 4 times. The right graph is its corresponding down-sampled envelope modulation spectrum (in dB).



To generate the AEMS spectrum, the signal undergoes several processing steps. First, it is filtered into seven-octave bands using eighth-order Chebyshev digital filters, with center frequencies of 125, 250, 500, 1,000, 2,000, 4,000, and 8,000 Hz. Next, the amplitude envelope is extracted (half-wave rectified, then low-pass filtered at 30 Hz using a fourth-order Butterworth filter), and down-sampled to 80 Hz (mean subtracted) from the full signal and each of the seven octave bands.

For each down-sampled envelope, a power spectrum analysis is performed using a 512-point fast Fourier transform, applying a Tukey window. The results are converted to decibels for frequencies up to 10 Hz

(normalized to maximum autocorrelation). Consequently, seven EMS metrics are computed for each band and one metric from the full signal, resulting in a total of 56 metrics [(7 octave bands + 1 full signal) × 7 metrics]. All amplitude measures are normalized to the average amplitude across the spectrum. Table 2 illustrates the seven metrics and their descriptions.

TABLE 2. AMPLITUDE ENVELOPE MODULATION SPECTRUM METRICS
Description of Amplitude Envelope Modulation Spectrum (AEMS) Metrics.

Metric	*Description*
Centroid	The frequency at which the amplitude of the spectrum is balanced. The period of this frequency corresponds to the duration of the dominant repetitive amplitude pattern.
Peak frequency	The frequency of the peak in the spectrum with the greatest amplitude.
Peak amplitude	The amplitude of the peak frequency described above (normalized by the overall amplitude of energy in the spectrum). This measurement indicates the extent to which the rhythm is influenced by a single frequency.
E3 – 6	Energy in the region of 3–6 Hz (normalized). This corresponds to the approximate spectral range around 4 Hz, which has shown correlations with intelligibility (Houtgast & Steeneken, 1985) and an inverse correlation with segmental deletions (Tilson & Johnson, 2008).
Below4	Energy in spectrum from 0–4 Hz (normalized). The spectrum was divided at 4 Hz, as it has been suggested that the energy below and above 4 Hz exhibited relatively low correlation across diverse speakers and sentences.
Above4	Energy in spectrum from 4–10 Hz (normalized).
Ratio4	Energy below 4 Hz/energy above 4 Hz (normalized)

2.4. DATA PROCESSING AND ANALYSIS

The values of each AEMS for each sentence and frequency band underwent outlier examination (± 2SD) based on group and speaking rate. Outliers were excluded before proceeding with statistical analyses. In total, 4.1% of the data (N=93,967) were eliminated. All statistical analyses were conducted using SPSS (Version 29).

2.5. ANALYSIS: STEPWISE DISCRIMINATION ANALYSIS

Stepwise discriminant function analyses were carried out to evaluate the categorization of German-accented and French-accented English as native English utterances; English-accented German and French-accented German as native German utterances; and native German-accented French utterances as native French. These analyses were performed for each of the five speaking rates as well as for all rates combined.

Following the methodologies outlined by Liss et al. (2010) and Wayland and Nozawa (2019), in each step of the analysis, the parameter that minimized Wilks' lambda was incorporated if the change's F statistic was statistically significant at $p < 0.05$. Furthermore, any parameter that ceased to significantly decrease Wilks' lambda ($p > 0.10$) upon adding a new variable was excluded from the discriminant function analysis. The outcome of this analysis yielded canonical functions, which signify linear combinations of the chosen predictor variables. These functions were subsequently utilized to establish classification rules for determining group membership, encompassing categories such as native English, German-accented English, French-accented English, native German, English-accented German, and French-accented German. The accuracy rate was expressed as a percentage.

To assess the robustness of the classification rules, leave-one-out cross-validation was employed. This involved classifying the excluded utterances based on the functions derived from all other utterances.

Finally, positive predictive values (PPV) were calculated for the L2 accented utterances. These values represent the percentage of correctly predicted cases with the observed characteristic compared to the total number of cases predicted as having the characteristic. For example, positive predictive values for German-accented English indicate the percentage of German-accented English utterances that were correctly predicted to be native English, as a percentage of all utterances in the analysis classified as native English.

3. RESULTS

Table 3 displays the cross-validated positive predictive values (PPVs) for German-accented English and French-accented English across the five speaking rates, as well as when all rates were considered together. The PPVs for German-accented English are consistently higher than those for French-accented English across all five rates. This suggested that a larger proportion of German-accented English utterances were categorized as English. A two-tailed independent T-test was conducted comparing PPVs across the five rates and confirmed that the difference was statistically significant [$t(8) = 5.94$, $p < .001$].

11 out of the 56 predictors were found to be statistically significant in the combined-rate DFA model. The primary predictor among these was Ratio4_125, denoting the normalized energy below 4 Hz to the energy above 4 Hz in the 125 Hz frequency band. In the DFA models for each of the five rates, the number of significant predictors varied: 1 for the normal rate, 3 for the fast rate, 4 for both the slow and fastest rates, and 5 for the slowest rate, with no overlap in the top predictor.

TABLE 3. POSITIVE PREDICTIVE VALUES FOR GERMAN AND FRENCH-ACCENTED ENGLISH
Positive predictive values (PPV) for German-accented English and French-accented English in terms of their classification as English based on EMS metrics

Metric	*Accented type*	*Speaking rate*	*PPV (%)*
EMS	German-accented English	slowest	34.0
		slow	36.1
		normal	37.5
		fast	42.4
		fastest	51.2
		all rate combined	32.2
	French-accented English	slowest	6.0
		slow	5.6
		normal	25.0
		fast	12.1
		fastest	2.4
		all rate combined	9.1

PPVs for English-accented German and French-accented German are shown in Table 4. English-accented German was classified as German at a higher percentage than French-accented German for each rate and when all the rates were combined. The difference was statistically significant [$t(8) = 6.02$, $p<.001$].

In the combined-rate DFA model, 9 significant predictors emerged, with E3-6_2000, which represents energy in the range of 3–6 Hz (normalized by overall spectrum amplitude) from the 2000 Hz band, being the top predictor. In the individual rate models, 3 to 9 significant predictors were identified.

TABLE 4. POSITIVE PREDICTIVE VALUES FOR ENGLISH- AND FRENCH-ACCENTED GERMAN
Positive predictive values (PPV) for English-accented German and French-accented German in terms of their classification as German based on EMS metrics

Metric	*Accented type*	*Speaking rate*	*PPV (%)*
EMS	English-accented German	slowest	12.2
		slow	10.0
		normal	18.4
		fast	16.7
		fastest	14.7
		all rate combined	11.7
	French-accented German	slowest	1.1
		slow	1.4
		normal	5.7
		fast	4.9
		fastest	4.9
		all rate combined	6.1

Table 5 shows PPVs for German-accented English as English and German-accented French as French. The difference was statistically significant [$t(8) = 3.46$, $p = .009$] indicating that German-accented English was classified as English significantly more frequently than German-accented French as French.

The combined-rate DFA model resulted in 14 significant predictors, with Peak amplitude-4000 being the top predictor. This predictor represents the amplitude of the frequency peak in the spectrum from the 4,000 Hz band. In the separate rate models, a varying combination of 6 to 9 significant predictors was identified for the five different rates.

TABLE 5. POSITIVE PREDICTIVE VALUES FOR GERMAN-ACCENTED ENGLISH AND FRENCH
Positive predictive values for German-accented English and German-accented French in terms of their classification as English and French, respectively, based on EMS metrics.

Metric	*Accented type*	*Speaking rate*	*PPV (%)*
EMS	German-accented English	slowest	28.9
		slow	39.0
		normal	44.4
		fast	36.6
		fastest	42.6
		all rate combined	41.5
	German-accented French	slowest	12.0
		slow	18.2
		normal	28.9
		fast	24.1
		fastest	31.3
		all rate combined	20.6

4. DISCUSSION

The aim of the study was to examine the potential influence of shared linguistic rhythm on the acquisition of rhythm in a second language (L2). The employed rhythm metrics analyzed temporal regularities extracted from the amplitude envelope modulation spectrum. These metrics capture low-rate temporal variations in spectral envelope amplitude, corresponding to prosodic units such as syllables, and regular durational variations like stressed–unstressed intervals. Both different L1-L2 language combinations (German-accented vs. French-accented English and English-accented German

vs. French-accented German) and the same L1-L2 combination (German-accented English vs. German-accented French) were explored.

The findings strongly support the advantage of the shared-L1 rhythm hypothesis, demonstrating that German-accented English is consistently more likely to be classified as English compared to French-accented English. Furthermore, German-accented English is classified as being closer to native English than German-accented French is to native French.

Interestingly, the results align with the word-based acoustic distance estimations derived from self-supervised neural models. These suggest that the word-level pronunciation of English by German speakers is closer to native English than that of French speakers. Similarly, English speakers show a closer word-based pronunciation to native German than to French. Furthermore, German speakers exhibit a closer pronunciation to English than to French. Together, the findings suggest that rhythm planning may be influenced by the words and their segmental makeup in the utterance (Myers & Watson, 2021).

The significance of various predictors identified in the Discriminant Function Analysis (DFA) models offers valuable insights into the acoustic features that contribute to the observed rhythmic classification patterns. For example, energy below 4 Hz was the primary predictor for differentiating between German-accented and French-accented English. Notably, predictor values were 1.9 for German-accented English and 2.8 for French-accented English. This indicates that in the 125-Hz octave band frequency (ranging from 88 to 177 Hz), the spectral envelope amplitude modulation rates below 4 Hz are more pronounced (relative to those above 4 Hz) in English spoken by French speakers than in that spoken by German speakers. An amplitude modulation rate of 4 Hz is typically associated with syllable-pattern information in speech, as noted by Greenberg et al. (2003) and Greenberg (2006). These findings suggest that French-accented English exhibits a stronger presence of regular temporal patterns associated with prosodic units of or closer to a syllable size, reflecting a possible influence from French's traditionally classified syllable-timed rhythm.

On the other hand, energy in the range of 3–6 Hz emerged as the top predictor for English-accented vs. French-accented German. The 3–6 Hz range roughly corresponds to the spectral region around 3–4 Hz, which has been shown to correlate with vowel deletions, particularly in English (Tilsen & Johnson, 2008). Crucially, the predictor value was higher for English-accented German compared to French-accented German (4.5 vs. 3.8). The higher values for English-accented German may thus be a greater amount of vowel deletion in German produced by English speakers compared to French-accented German, due possibly to vowel reduction in unstressed syllables in English.

Lastly, it is worth noting that top predictors and various combinations of significant predictors were identified for different speaking rates, indicating potential variations in rhythm articulation adjustments across varying rates of speech production. Further research is necessary to fully elucidate the relationship between these predictor patterns and the dynamic nature of speech rhythm under different speech tempos.

In conclusion, despite its extensive history of progress, research on speech rhythm continues to be exploratory, due to the complexity of the underlying phenomena and the lack of an effective tool that bridges the gap between linguists' intuition and tangible statistical patterns in the speech signal (Deloche et al., 2024). Our findings not only support the facilitating roles of shared linguistic rhythm in L2 speech learning but also underscore the AEMS's significant potential as a powerful tool for analyzing speech rhythms, both within and across languages. Its ability to capture regular patterns across various speech unit sizes uniquely positions it to reveal nuanced rhythmic differences overlooked by traditional methods

focused solely on segmental intervals. In addition, the AEMS approach is automated, thus avoiding the labor-intensive and error-prone process required for segmenting speech into vocalic and consonantal intervals.

Acknowledgment

We express our gratitude to Professor Volker Dellwo for his generosity in sharing the BonnTempo corpus. We also extend our appreciation to Professor Andrew Lotto for providing the MatLab codes for the EMS metrics, and to Professor Yonghee Oh for providing Figure 1.

Summary

Using metrics extracted from the amplitude envelope modulation spectrum (AEMS), this study demonstrated the facilitating roles of shared linguistic rhythm and established the AEMS as a powerful tool for analyzing speech rhythms, both within and across languages.

Implications

To fully understand the complexity of linguistic rhythm, it is crucial to employ metrics capable of quantifying temporal patterns across various speech unit sizes. Automated tools like the AEMS significantly enhance our ability to examine rhythm within and across languages, facilitating a more nuanced understanding of the connection between linguists' intuition and tangible statistical patterns in the speech signal.

Gains

The acquisition of second language (L2) rhythm is facilitated by a shared first language (L1) rhythm, with improved L2 rhythm production correlating to enhanced word/sub-word production in L2. This observation supports the notion that planning metrical representations for rhythm also depends on the words and their segmental composition in the spoken utterance.

Index terms

Cross-linguistics, Rhythm class, Acquisition, Amplitude Envelope Modulation

5. REFERENCES

Abercrombie, D. (1967). Elements of General Phonetics. Edinburgh: Edinburgh University Press.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. Phonetica, 66(1-2), 46-63.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. Journal of Phonetics, 40(3), 351-373.

Bartelds, M., de Vries, W., Richter, C., Liberman, M., & Wieling, M. (2021). Measuring foreign accent strength using an acoustic distance measure. In 12th International Seminar on Speech Production (pp. 17-20). Haskins Press.

Bartelds, M., de Vries, W., Sanal, F., Richter, C., Liberman, M., & Wieling, M. (2022). Neural representations for modeling variation in speech. Journal of Phonetics, 92, 101137.

Boersma, P., Weenink, D. 2001. Praat, a system for doing phonetics by computer. Glot International 5, 341-345.

Bertinetto, P. M. (1989). Reflections on the dichotomy 'stress' vs.'syllable-timing'. Revue de phonétique appliquée, 91(93), 99-130.

Dauer, R. 1983. Stress-timing and syllable-timing reanalysed. Journal of Phonetics 11, 51–62.

Cumming, R. E. (2010). Speech rhythm: The language-specific integration of pitch and duration (Doctoral dissertation, University of Cambridge).

Dellwo, V., Aschenberner, B., Wagner, P., Dancovicova, J., & Steiner, I. (2004). BonnTempo-Corpus and BonnTempo-Tools: A database for the study of speech rhythm and rate. In Eighth International Conference on Spoken Language Processing.

Deloche, F., Bonnasse-Gahot, L., & Gervain, J. (2024). Acoustic characterization of speech rhythm: going beyond metrics with recurrent neural networks. arXiv preprint arXiv:2401.14416.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. Papers in laboratory phonology, 7(515-546), 1-16.

Greenberg, S. (2006). A multi-tier framework for understanding spoken language. Listening to speech: An auditory perspective, 411-433.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. Journal of Phonetics, 31(3-4), 465-485.

Kohler, K. J. (2009). Rhythm in speech and language: A new research paradigm. Phonetica, 66(1-2), 29-45.

Langus, A., Mehler, J., & Nespor, M. (2017). Rhythm in language acquisition. Neuroscience & Biobehavioral Reviews, 81, 158-166.

Liss, J. M., LeGendre, S., & Lotto, A. J. (2010). Discriminating dysarthria type from envelope modulation spectra.

Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. The Journal of the Acoustical Society of America, 129(5), 3258-3270.

Myers, B. R., & Watson, D. G. (2021). Evidence of Absence: Abstract Metrical Structure in Speech Planning. Cognitive Science, 45(8), e13017.

Ordin, M. & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. J. Acoust. Soc. Am.138 (2), 533-545.

Pike, K.L. The Intonation of American English. University of Michigan (Ann Arbor), July 1945. doi:10.2307/409880

Ramus, F., Nespor, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal, Cognition 73, 265- 292.

Roach, P. (1982). On the distinction between 'stress-timed'and 'syllable-timed'languages. Linguistic controversies, 73, 79.

Terken, J., & Hermes, D. (2000, October). The perception of prosodic prominence. In Prosody: Theory and experiment: Studies presented to Gösta Bruce (pp. 89-127). Dordrecht: Springer Netherlands.

Tilsen, S., & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. The Journal of the Acoustical Society of America, 124(2), EL34-EL39.

Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. Laboratory Phonology, 4(1), 93-118.

Van Maastricht, L., Krahmer, E., Swerts, M., & Prieto, P. (2019). Learning direction matters: A study on L2 rhythm acquisition by Dutch learners of Spanish and Spanish learners of Dutch. Studies in Second Language Acquisition, 41(1), 87-121.

Wayland, R., & Nozawa, T. (2019, December). Calibrating rhythms in L1 Japanese and Japanese accented English. In Proceedings of Meetings on Acoustics 178ASA (Vol. 39, No. 1, pp. 2844-2844). Acoustical Society of America.

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm?. The Journal of the Acoustical Society of America, 127(3), 1559-1569.