30 Duration-Based and Acoustic Speech Rhythm Metrics

ROBERT FUCHS

30.1 Introduction

The scholarly debate around speech rhythm over the last decades has produced a rich literature on its theoretical underpinnings (e.g., Arvaniti, 2009; Nolan and Jeon, 2014; Gibbon, 2021), its measurement (using so-called rhythm metrics, for example, Low et al., 2000; Deterding, 2001), as well as its applications to a great number of languages and dialects (Szakay, 2006; White and Mattys, 2007; Behrman et al., 2019).[1]

Speech rhythm can be defined as 'the production, for a listener, of a regular recurrence of waxing and waning prominence profiles across syllable chains over time' (Kohler, 2009, p. 41). While many scholars of speech rhythm might overall agree with this definition, it is in fact much broader in scope than the actual, practical definition and operationalisation of speech rhythm that many studies have used and continue to use. Most research on speech rhythm focuses on duration; that is, Kohler's 'waxing and waning' is operationalised as the alternation of long and short syllables or vowels. However, Kohler refers to alternation in prominence, which might be realised acoustically by duration, but also by pitch, loudness, and potentially other acoustic features. In Section 30.3, the discussion will return to this point and explore ways of measuring rhythmic alternation involving features other than duration.

Speech rhythm was originally conceived as a suprasegmental phenomenon covering three classes of languages (e.g., Pike, 1945; Abercrombie, 1967, p. 97):

(1) syllable-timed languages, with syllables of equal duration (i.e., isochronous), such as in Spanish (see top panel of Figure 30.1, where syllables are of equal duration and feet vary in duration)
(2) stress-timed languages, with feet, that is, a stressed syllable followed by one or more unstressed syllables, of equal duration (i.e., isochronous), such as in British English (see

[1] This and the following paragraphs rely partially on Fuchs (2023c).

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

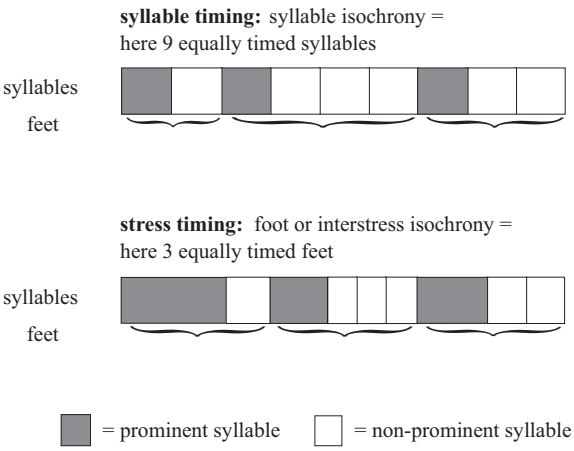


Figure 30.1 Idealised syllable-timing and stress-timing.
Idealised syllable-timing involves syllables of equal duration and feet of unequal duration (top), while idealised stress-timing involves syllables of unequal duration and feet of equal duration (bottom).
Figure from Fuchs (2016, p. 36).

bottom panel of Figure 30.1, where feet are of equal duration and syllables vary in duration)

(3) mora-timed languages, with morae (a unit smaller than a syllable but often comprising more than one phoneme) of equal duration (i.e., isochronous), such as in Japanese.

Despite attempts to test this rhythm class hypothesis in terms of speech production (e.g., Dankovičová and Dellwo, 2007) and perception (e.g., Ramus et al., 2003), there is substantial evidence that it lacks empirical support. In particular, two (interlinked) claims of the rhythm class hypothesis have been criticised: (i) the isochrony of specific prosodic units and (ii) the existence of discrete rhythm classes. Note that these two claims are independent – it is possible for the languages of the world to fall into discrete rhythm classes without isochrony.

However, evidence from speech production indicates that the claim of isochrony of syllables in syllable-timed languages, feet in stress-timed languages, and morae in mora-timed languages, respectively, is inaccurate (Dauer, 1983). Moreover, the second claim, that is, that there are distinct rhythm classes, also turns out to be problematic. Evidence from speech production indicates that languages do not fit neatly into rhythm classes (Grabe and Low, 2002), but that there are gradual instead of categorical differences in timing between languages. Both adult listeners (White et al., 2012) and infants (Molnar et al., 2014; Gasparini et al., 2021) are sensitive to temporal information rather than to rhythm class in language discrimination.

This evidence prompted a reconceptualisation of speech rhythm as a gradable phenomenon (see White et al., 2012), with some languages involving greater variability in the duration of syllables (stress-timed languages) and other languages lesser variability (syllable-timed languages). In addition to or instead of syllables, many studies focused on the nuclei of syllables, that is, vowels, similarly identifying greater variability in duration with stress-timing and lesser variability with syllable-timing. Meanwhile, the concept of mora-timing was all but abandoned.

Nevertheless, the original terminology often shines through, with languages or varieties regularly described as syllable- or stress-timed (as in the contributions to Kortmann and Schneider, 2004). Alternatively, in order to reflect the gradable nature of speech rhythm, other

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

terms can be used, such as 'more stress-timed', 'relatively stress-timed' or 'stress-based', and mutatis mutandis for syllable-timing (Dauer, 1983; Braun and Geiselmann, 2011). Nevertheless, possibly because these newer terms are somewhat unwieldy, the older ones are still in use as a shorthand. Moreover, the expression 'rhythm classes' can also still be found (e.g., White and Mattys, 2007; Gasparini et al., 2021), though it is now used more rarely and should be entirely avoided (White et al., 2012) since it inaccurately invokes clear-cut categories instead of reflecting the above-mentioned gradable nature of speech rhythm.

A considerable body of current research attempting to quantify and compare the speech rhythm of languages and dialects, or involving sociolinguistic variation more generally, relies on this notion of greater or lesser variability in the duration of consecutive syllables or vowels, with lesser variability identified with (relative) syllable-timing and greater variability with stress-timing (thus, in Figure 30.1, syllables would not be identical in duration for syllable-timing but just relatively similar, whereas in stress-timing, syllables vary greatly in duration). Returning to Kohler's definition above, the basic idea behind this approach is that speech with a tendency towards stress-timing involves great variations in prominence (including duration) between prominent (or stressed) and non-prominent (or unstressed) syllables, while a tendency towards syllable-timing involves small variations between prominent and non-prominent syllables.

Overall, in current practice more studies measure speech rhythm with regard to the durational variability of vowels rather than syllables, while other metrics rely on the durations of consonants or yet other phonetic or phonological units. Technically, this method is more complex than discussed so far, since it is not the durational variability of vowel phonemes that is measured but the durational variability of vocalic intervals, that is, one or more consecutive vowels not interrupted by any consonants, possibly spanning word boundaries.

30.2 Duration-Based Metrics

Duration-based rhythm metrics were initially conceived around the aim of quantifying the notion that there are reduced vowels in unstressed syllables in stress-timed languages, while syllable-timed languages have no or little vowel reduction. In addition, stress-timed languages tend to allow consonant clusters with several consonant phonemes, while syllable-timed languages rarely do (Dauer, 1983, pp. 55–58; Ramus et al., 1999, p. 270; Schiering, 2007).

The considerable number of duration-based rhythm metrics can be classified along three broad distinctions, that is, (1) whether they rely on the durations of vowels, consonants, or syllables, (2) whether they normalise for variation in speech rate, and (3) how they quantify variability (see Figure 30.2). Thus, a distinction can be made between (1) vocalic, consonantal, and syllabic metrics, (2) rhythm metrics normalised for speech rate or not, and (3) global and local metrics.[2]

Of these criteria, the global/local distinction requires further explanation. Global rhythm metrics compute a measure of variability of duration of all vocalic, consonantal, or syllabic

[2] This and the following paragraphs rely partially on Fuchs (2023a).

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

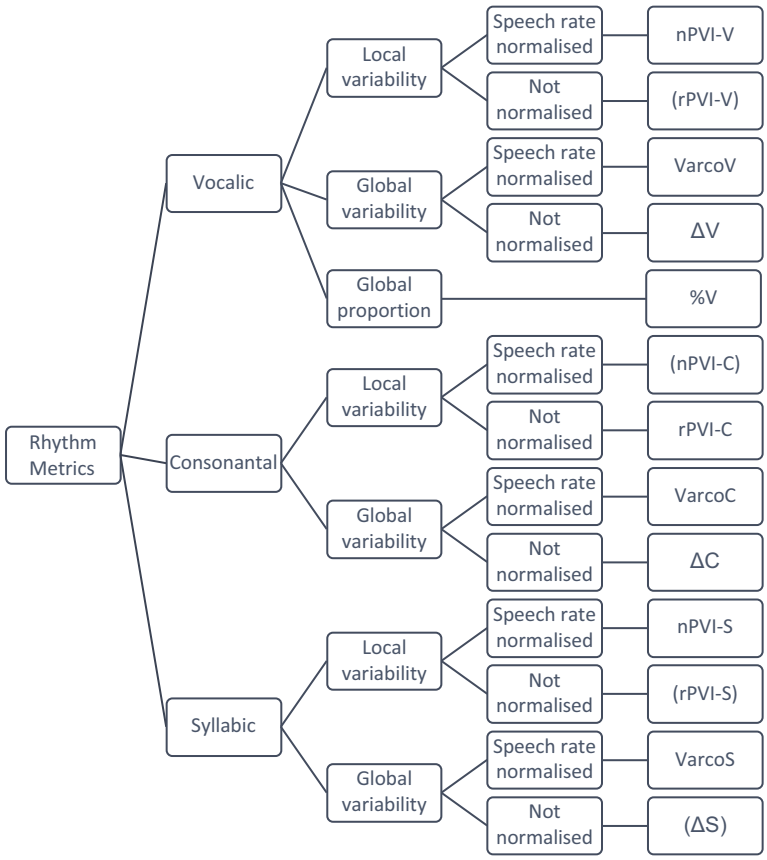


Figure 30.2 Taxonomy of common duration-based rhythm metrics.
Duration-based rhythm metrics are classified here according to the segmental unit of measurement (vocalic, consonantal, or syllabic), the operationalisation of variability (local or global), and speech rate normalisation. Theoretically possible but uncommon metrics are shown in brackets.
Figure adapted and expanded from Fuchs (2023a).

intervals regardless of their position in the utterance. This can be realised by computing the standard deviation, which yields the measures ΔV (read: 'Delta V') and ΔC for vocalic and consonantal intervals, respectively (Ramus et al., 1999). These metrics can in turn be normalised for speech rate by taking the standard deviation, divided by the mean, multiplied by 100, resulting in metrics known as VarcoV, VarcoC, and VarcoS, respectively (also known as coefficients of variation for vocalic, consonantal, and syllable durations, respectively; Dellwo, 2006; White and Mattys, 2007; Rathcke and Smith, 2011).

While global rhythm metrics are computed without regard to the temporal order of the vocal, consonantal, or syllabic units, local metrics are based on differences between adjacent pairs of vocalic intervals (for vocalic metrics), consonantal intervals (for consonantal metrics), or syllables (for syllabic metrics). Finally, the mean of all pairwise comparisons is computed. These metrics are commonly referred to as pairwise variability indices (PVI), where an initial lower case 'n' indicates the speech-rate-normalised version and 'r' the raw or non-normalised version (Low et al., 2000). Of the six theoretically possible PVIs, three, i.e. rPVI-C, nPVI-V, and nPVI-S (Gibbon and Gut, 2001),[3] have been applied in research, while the raw vocalic and

[3] While Gibbon and Gut (2001) used a different term for this metric, their definition equated to an nPVI-S index.

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

syllabic (rPVI-V, rPVI-S) and the normalised consonantal index (nPVI-C, shown in brackets in Figure 30.1) are rarely or never used. Moreover, a rhythm metric that goes beyond this taxonomy, but is often used, accounts for the proportion of vocalic durations relative to total utterance duration, known as %V (Ramus et al., 1999; %V is technically the inverse of %C, and the convention is to refer to %V). In addition to the duration-based rhythm metrics discussed here, there are others that are less widely used; a comprehensive overview can be found in Fuchs (2016, pp. 35–52).

The considerable number of duration-based speech rhythm metrics prompts the question of whether they are all equally reliable or, alternatively, if one or a few of them are superior to the rest. Empirical validity tests indicate that nPVI-V, VarcoV, and %V are the most reliable and should therefore be preferred (White and Mattys, 2007; Wiget et al., 2010; note, however, that these tests did not include syllabic metrics).[4] Additional evidence on the validity of these metrics comes from speech perception and indicates that these speech production metrics appear to at least partially capture the human perception of rhythmicity (Fuchs, 2023a). Finally, given that nPVI-V and VarcoV both measure variability in vocalic durations and are speech-rate-normalised, while %V accounts for the proportion of vocal durations over total utterance duration, it is recommended that studies on speech rhythm relying on duration-based metrics rely on %V as well as, at a minimum, either nPVI-V or VarcoV. With this choice of metrics, two potentially distinct aspects of rhythmicity are captured.

Current practice in the field does not completely follow this advice. Of all duration-based rhythm metrics, the one probably by far most widely used is nPVI-V. For example, a synthesis of previous research on variation in speech rhythm among varieties of English identified 18 relevant studies on 23 varieties of English applying the nPVI-V to samples of read speech (Fuchs, 2023b). Taken together, these studies support the widely held assumption that so-called Outer Circle varieties (where English is widely used as a second language and local lingua franca, for example, Indian and Nigerian English) tend to be more syllable-timed than Inner Circle varieties (where English is mainly used as a first language, for exmaple, British, American, and Australian English). A major factor that likely accounts for this result is prosodic transfer from relatively syllable-timed languages widely used in countries where Outer Circle varieties of English are spoken. Moreover, so-called Expanding Circle varieties, where English is mainly used as a foreign language but does not enjoy official status or play a large local role (e.g., in Japan and Germany), also tend to be more syllable-timed than Inner Circle varieties. This tendency may in some cases be explained by prosodic transfer as well. However, other reasons are conceivable, too. These include a possible tendency towards syllable-timing in second-language acquisition, and a selection as well as publication bias in the extant publication record. Selection bias may involve researchers selecting research questions that are likely to confirm assumptions they and the field as a whole would consider to be in keeping with the extant research record. Moreover, publication bias may play a role when researchers engaging in such research are more likely to consider manuscripts worth publishing,

[4] Arvaniti (2009) presented data that, according to the author, indicate that these metrics cannot differentiate between languages. However, this finding can arguably be explained by a failure to systematically account for variation in speech style (read versus spontaneous) in the statistical analysis.

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

and more likely to get manuscripts accepted in research outlets, when the results indicate a non-null finding.

The duration-based metrics discussed here have been criticised in several ways (for a more extensive discussion, see Fuchs, 2016, pp. 57–69). One objection is that duration is only one of several relevant acoustic correlates of prominence, and a more holistic assessment of rhythm should consider additional acoustic measures (see Section 30.3). Another point of contention focuses on a lack of explanatory power, arguing that rhythm metrics quantify rhythm as a surface phenomenon and are influenced by several phonological parameters simultaneously (Turk and Shattuck-Hufnagel, 2013, p. 109).

Rhythm metrics have also been criticised as empirically inadequate by Arvaniti (2009, 2012) because they are influenced by elicitation method and syllable complexity. Arvaniti used sets of sentences that were explicitly designed to elicit a more stress- or more syllable-timed rhythm and showed that this result applies across several languages. However, this approach can in turn be criticised for its circularity. Speech material with specific properties was selected, and the analysis in turn revealed exactly these properties. Moreover, speech rhythm metrics were in fact explicitly designed so as to capture differences in syllable complexity. Finally, Arvaniti's (2012) results generally go in the expected direction, for example with English and German showing higher nPVI-V values, and lower %V values, than Spanish. In fact, across the three metrics nPVI-V, VarcoV, and %V (shown in previous research to be more reliable than other durational metrics; White and Mattys, 2007; Wiget et al., 2010) and across five different conditions, the difference between English and Spanish is of a non-negligible size and goes in the expected direction in 12 out of 15 cases. Finally, further empirical evidence supporting duration-based metrics comes from a meta-analysis indicating that language discrimination by infants can be explained by durational rhythm metrics (Gasparini et al., 2021).

30.3 Acoustic Metrics

While duration-based metrics are widely used in research on speech rhythm, what they actually measure can only partially account for Kohler's (2009, p. 41) 'waxing and waning prominence profiles' (referred to in Section 30.1). These metrics merely measure variability in duration and neglect any other acoustic correlates of prosodic prominence, the most important of which are intensity and fundamental frequency/f0, alongside their psychoacoustic counterparts loudness and pitch, respectively.

Moreover, range of other proposals do not (directly) rely on a particular acoustic correlate of prosodic prominence but are based on complex transformations of spectral energy or other acoustic information in the speech signal (e.g., Galves et al., 2002; Tilson and Johnson, 2008; Goswami and Leong, 2013; Tilsen and Arvaniti, 2013; Ravignani and Norton, 2017; Gibbon and Li, 2019; Davis and Jeesun, 2023). Some of these approaches to measuring rhythm are difficult to reproduce for other researchers because the code is not publicly available. In the following, the discussion focuses on acoustic rhythm metrics that apply the Varco or PVI concept to acoustic correlates of rhythm beyond duration. The advantage of these approaches is arguably that they present a way of accounting for speech rhythm as a multidimensional

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

acoustic phenomenon (Fuchs, 2016), realised on multiple acoustic 'channels', while at the same time the Varco and PVI indices are mathematically relatively straightforward, and therefore easily interpretable, indices of variability (see Table 30.1 for an overview).

These metrics can be classified according to:

(1) the variability index they use (Varco or PVI)
(2) the phonological unit from which acoustic information is extracted (vowels or syllables)
(3) the acoustic correlate of prominence (f0, intensity, loudness, as well as combinations thereof with duration).

Moreover, variability in the acoustic correlates of prominence can be quantified in different ways:

(4a) For intensity and loudness, there are proposals to quantify variability either in their average, or in their peak value, in the syllable or vocalic interval.

Thus, VarcoS(avgInt) captures variability in the average intensity in syllables. Instead of calculating the standard deviation of syllable *duration* divided by the mean, multiplied by 100, which is the way the regular VarcoS for syllable duration is calculated (see Section 30.2), for VarcoS(avgInt), average intensity over each syllable is entered into this equation. Thus, VarcoS(avgInt) is calculated as the standard deviation of *average intensity* for each syllable, divided by the mean, multiplied by 100. Following the same approach, an nPVI-S(avgInt) can be calculated by entering average intensity for each syllable into the nPVI equation. Furthermore, instead of average intensity over a syllable or vocalic interval, peak intensity can be used as a measure, for example, in the vocalic PVI for peak intensity nPVI-V(peakInt). Finally, all these metrics can also be applied to loudness as the psychoacoustic correlate of intensity, for example, the vocalic PVI for average loudness nPVI-V(avgLoud)) and for peak loudness (nPVI-V(peakLoud)).

Based on these proposals, researchers can trace rhythmic variability independently for distinct acoustic correlates of prominence. For example, the regular nPVI-V based on durations (nPVI-V(dur) can be used to quantify the degree of variability in *durations* between vocalic intervals, while nPVI-V(loud) can be used in parallel to quantify the degree of variability in *loudness* between vocalic intervals.

(4b) A further refinement of these measures consists in possible combinations of variability in duration with either variability in intensity or loudness. More specifically, it is conceivable that variability in duration and intensity/loudness are tightly correlated with each other; that is, the longer a vowel is, the louder it is at the same time. In other cases, variability in duration and intensity/loudness might not be tightly linked; that is, there might be particularly long vowels that are not necessarily also louder than vowels of average duration.

There are two proposals to account for the potential linkage between variability in duration and intensity/loudness: applying the nPVI formula to both duration and intensity/loudness simultaneously (e.g., nPVI-V(dur+avgLoud), nPVI-V(dur+peakLoud)) or calculating the so-called Amplitude Integral or Loudness Integral (e.g., nPVI-V(AI), nPVI-V(LI)).

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

Table 30.1 Rhythm metrics

Rhythm metrics based on f0, intensity, and loudness (modified from Fuchs, 2016: *pp. 78–79)*

Metric	*Description*	*Main reference*
nPVI-V(avgInt)	Pairwise variability index for intensity variation between vocalic intervals. Mean of the differences between root mean square amplitude of successive vocalic intervals.	Low (1998)
nPVI-V(AI)	Pairwise variability index for intensity and duration variation between vocalic intervals. Mean of the differences between the Amplitude Integral of successive vocalic intervals.	Low (1998)
VarcoS(avgInt)	Coefficient of variation of average intensity in syllables (i.e., standard deviation of average intensity divided by the mean), multiplied by 100.	He (2012)
nPVI-S(avgInt)	Pairwise variability index for variation in average intensity. Mean of the differences between average intensity of adjacent syllables, divided by their sum, multiplied by 100.	He (2012)
nPVI-V(peakInt)	Pairwise variability index for intensity variation between vocalic intervals. Mean of the differences between peak amplitude of successive vocalic intervals.	Fuchs (2016)
nPVI-V(avgLoud)	Pairwise variability index for variation in average loudness between vocalic intervals. Mean of the differences between average loudness of successive vocalic intervals.	Fuchs (2016)
nPVI-V(peakLoud)	Pairwise variability index for variation in peak loudness between vocalic intervals. Mean of the differences between peak loudness of successive vocalic intervals.	Fuchs (2016)
nPVI-V (dur+avgLoud)	Pairwise variability index for combined variation in duration and mean loudness between vocalic intervals. Mean square of the normalised differences between duration and loudness of successive vocalic intervals.	Fuchs (2014b, 2016)
nPVI-V(dur +peakLoud)	Pairwise variability index for combined variation in duration and peak loudness between vocalic intervals. Mean square of the normalised differences between duration and peak loudness of successive vocalic intervals.	Fuchs (2016)
nPVI-V(LI)	Pairwise variability index for loudness and duration variation between vocalic intervals. Mean of the differences between the Loudness Integral of successive vocalic intervals.	Fuchs (2016)
nPVI-V(f0)	Pairwise variability index for f0 variation between vocalic intervals. Mean of the differences between the pitch excursion of successive vocalic intervals.	Cumming (2010, 2011)
nPVI-V(dur*f0)	Pairwise variability index for variation in duration between vocalic intervals, adjusted for the influence of f0 on duration. Mean of the differences in adjusted duration between successive vocalic intervals.	Fuchs (2014a)

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

Finally, for f0, the existing proposals are not analogous to the rhythm metrics capturing variability in intensity and loudness discussed so far. Instead, the nPVI-V(f0) proposed by Cumming (2010, 2011) calculated the mean of the differences between the pitch excursion of successive vocalic intervals. Another proposal, by Fuchs (2014a), presents an nPVI for vocalic durations that takes into account that differences in f0 also influence perceived duration. In fact, given two vowels of the same duration but different f0, the vowel with higher f0 will be perceived as longer. The proposed nPVI-V(dur*f0) takes this effect into account and thus offers a psychologically more realistic measure of rhythmic variability in perceived duration.

The rhythm metrics proposed for acoustic correlates of prominence do not cover all possible permutations of the various means of computing variability (Varco and nPVI), acoustic correlate (intensity, loudness, f0), and indices of joint variability in duration and another acoustic correlate. Other combinations, such as a joint index of variability in loudness, f0, and duration, might be useful as well.

The rhythm metrics discussed in this section have the advantage of accounting for multiple acoustic correlates of prosodic prominence, thus presenting a more holistic analysis of speech rhythm as a multidimensional phenomenon. For example, in a study comparing read speech in Singapore English and British English, Low (1998, pp. 49, 53) was able to confirm the assumption that Singapore English has a more syllable-timed rhythm than British English. In terms of the variability of vocalic durations (nPVI-V(dur)), the analysis indicated that British English had about 1.63 times the durational variability of Singapore English. By contrast, the variability in the Amplitude Integral as a joint measure of amplitude and duration indicated only 1.15 times the variability.[5]

Furthermore, in another study on read speech from Indian English and British English, several rhythm metrics confirmed the assumption that Indian English has a relatively more syllable-timed rhythm than British English (Fuchs, 2016, pp. 114, 147). However, the magnitude of the differences between the two dialects varied depending on which acoustic correlate of prominence was investigated. Specifically, the PVI for vocalic durations (nPVI-V(dur)) indicated that British English has 1.10 times the rhythmic variability of Indian English in terms of the variability of vocalic durations. By contrast, the difference in rhythmic variability for average loudness (nPVI-V(avgLoud)) was greater – 1.23 times. Finally, the combined index accounting for simultaneous variability in duration and average loudness (nPVI-V(dur +avgLoud)) showed that the rhythmic differences between British and Indian English is in fact even greater, with the former having on average 1.45 times the rhythmic variability of the latter, indicating that duration and loudness may have a compounding effect on perceived rhythmic variability in British English, compared to Indian English.

Finally, turning to the role of f0 in rhythmic variation, an analysis of read speech by Fuchs (2014a) found that, in British English, the effect of f0 on perceived duration has a noticeable impact on the degree of rhythmic variability, with a significantly higher nPVI-V(dur*f0) compared to the nPVI-V(dur) that is based on duration only. Importantly, this effect appears to be dialect-specific and was not present for Indian English. In effect, using the nPVI-V

[5] These numbers refer to Low's 'full and reduced vowel sets'.

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

(dur*f0) allowed this study to show how the effect of f0 on perceived duration further enhances the differences between relatively more stress-timed British English and relatively more syllable-timed Indian English.

In addition to these metrics, Tilsen and Arvaniti (2013) proposed several rhythm metrics based on 'empirical mode decomposition of the speech amplitude envelope', a computationally intensive method from which the authors derive syllabic and supra-syllabic measures. Their approach relies on empirical mode decomposition (Huang et al., 1998) yielding intrinsic mode functions (IMFs), where IMF_1 represents the fastest oscillation in the spectral envelope, IMF_2 the next fastest, and so on. From these, a total of seven power distribution metrics, rate metrics, and rhythm stability metrics are derived.

These studies clearly indicate that a reduction of speech rhythm to the measurement of variability in duration limits the analysis of speech rhythm to just a single acoustic correlate of prominence and may also underestimate the true degree of rhythmic variation between languages and dialects. For a more comprehensive analysis of speech rhythm as a prosodic phenomenon, variability in multiple acoustic correlates of prominence should be taken into account.

30.4 Conclusion: Applying Acoustic Rhythm Metrics

This chapter started out with a reference to Kohler's (2009) definition of speech rhythm in terms of 'waxing and waning prominence profiles across syllable chains over time'. It then presented several duration-based rhythm metrics. While some of these metrics have been widely used to account for variation in speech rhythm between languages, dialects, as well as sociolinguistic variation, they arguably neglect acoustic correlates of prominence other than duration. The discussion then turned to another class of rhythm metrics, which aim to account for rhythmic variation by quantifying variability in other acoustic correlates of prominence, that is, intensity, loudness, and f0, as well as the interaction between these and their interaction with duration. Several examples from the literature illustrate how these acoustic rhythm metrics can be fruitfully applied in order to provide a more holistic analysis of speech rhythm and to capture its multidimensional nature.

In addition to the theoretical desideratum of a more holistic analysis of speech rhythm, studies in this area should also try to adhere to a set of guidelines in order to provide valid and reproducible results. Such studies should try to compare like with like, that is, carefully select speakers and speech material that are constrained and clearly described. This requirement also includes that variation in speech style needs to be accounted for, for instance by using either read or spontaneous speech or by including both, but including speech style as a variable in the statistical analysis. The annotation and segmentation of the speech data requires clear guidelines in order to enhance comparability between and the reproducibility of studies. Further information on these guidelines as well as a guide towards the statistical analysis of speech rhythm data can be found in Fuchs (2023c).

A final point to consider concerns the computation of rhythm metrics. In order to simplify their application in empirical research and to enhance reproducibility, this chapter is accompanied by a Praat script that computes a large number of the duration-based and acoustic metrics presented in this chapter (available online at https://osf.io/79qyg/).

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

Box 30.1 Chapter Overview

Summary

The chapter introduces several duration-based and acoustic metrics of speech rhythm. While duration-based metrics are used widely, understanding speech rhythm as relating to variability in prominence opens the door to newer metrics that focus on variability in intensity, loudness, and pitch, in addition and in conjunction with duration.

Implications

Acoustic rhythm metrics have been rarely used in previous research. Their wider application will contribute to a better understanding of cross-linguistic and sociolinguistic variation in speech rhythm. In addition to production studies, research on the perception of speech rhythm promises to reveal additional evidence elucidating the complex nature of speech rhythm.

Gains

The multidimensional analysis of the acoustics of speech rhythm provides important insights for cognitive science and psycholinguistics. By incorporating both duration-based and acoustic metrics, researchers gain a nuanced understanding of the complex nature of speech rhythm and its variation within and across languages.

REFERENCES

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

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66(1–2), 46–63. https://doi.org/10.1159/000208930

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003

Behrman, A., Ferguson, S. H., Akhund, A., and Moeyaert, M. (2019). The effect of clear speech on temporal metrics of rhythm in Spanish-accented speakers of English. *Language and Speech*, 62(1), 5–29. https://doi.org/10.1177/0023830917737109

Braun, B., and Geiselmann, S. (2011). Italian in the no-man's land between stress-timing and syllable-timing? Speakers are more stress-timed than listeners. Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011), Florence, Italy, pp. 2697–2700. https://doi.org/10.21437/Interspeech.2011-689

Cumming, R. E. (2010). The language-specific integration of pitch and duration. PhD thesis, University of Cambridge.

Cumming, R. E. (2011). Perceptually informed quantification of speech rhythm in pairwise variability indices. *Phonetica*, 68(4), 256–277.

Dankovičová , J., and Dellwo, V. (2007). Czech speech rhythm and the rhythm class hypothesis. 16th International Congress of Phonetic Sciences (ICPhS), Saarbrücken, Germany, pp. 1241–1244.

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11(1), 51–62. https://doi.org/10.1016/S0095-4470(19)30776-4

Davis, C., and Jeesun, K. (2023). Novel methods for characterising L2 speech rhythm. In R. Fuchs, ed., *Speech rhythm in learner and second language varieties of English*, pp. 211–229. Springer. https://doi.org/10.1007/978-981-19-8940-7_9

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski and I. Szigeti, eds., *Language and language-processing*, pp. 231–241. Peter Lang.

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, 29, 217–230.

Fuchs, R. (2014a). Towards a perceptual model of speech rhythm: Integrating the influence of f0 on perceived duration. Proceedings of Interspeech 2014, Singapore, pp. 1949–1953.

Fuchs, R. (2014b). Integrating variability in loudness and duration in a multidimensional model of speech rhythm: Evidence from Indian English and British English. Social and Linguistic Speech Prosody. Proceedings of Seventh International Conference on Speech Prosody, Dublin, Ireland, pp. 290–294.

Fuchs, R. (2016). *Speech rhythm in varieties of English: Evidence from educated Indian English and British English*. Springer.

Fuchs, R. (2023a). Rhythm metrics and the perception of rhythmicity in varieties of English as a second language. In R. Fuchs, ed., *Speech rhythm in learner and second language varieties of English*, pp. 187–210. Springer.

Fuchs, R. (2023b). A synthesis of research on speech rhythm in native, learner and second language varieties of English: Introduction to the volume. In R. Fuchs, ed., *Speech rhythm in learner and second language varieties of English*, pp. 1–14. Springer.

Fuchs, R. (2023c). Analysing the speech rhythm of New Englishes: A guide to researchers and a case study on Pakistani, Philippine, Nigerian and British English. In G. Wilson and M. Westphal, eds., *New Englishes, new methods*, pp. 132–155. Benjamins. https://doi.org/10.1075/veaw.g68.07fuc

Galves, A., Garcia, J., Duarte, D., and Galves, C. (2002). Sonority as a basis for rhythmic class discrimination. Proceedings of Speech Prosody 2002, pp. 323–326. https://doi.org/10.21437/SpeechProsody.2002-66

Gasparini, L., Langus, A., Tsuji, S., and Boll-Avetisyan, N. (2021). Quantifying the role of rhythm in infants' language discrimination abilities: A meta-analysis. *Cognition*, 213, 104757. https://doi.org/10.1016/j.cognition.2021.104757

Gibbon, D. (2021). The rhythms of rhythm. *Journal of the International Phonetic Association*, 53(1), 233–265. https://doi.org/10.1017/S0025100321000086

Gibbon, D., and Gut, U. (2001). Measuring speech rhythm. Proceedings of the Seventh European Conference on Speech Communication and Technology (Eurospeech 2001), Aalborg, Denmark, pp. 95–98. https://doi.org/10.21437/Eurospeech.2001-36

Gibbon, D., and Li, P. (2019). Quantifying and correlating rhythm formants in speech. Proceedings of the Third International Symposium on Linguistic Patterns in Spontaneous Speech. Academia Sinica. https://doi.org/10.48550/arXiv.1909.05639

Goswami, U., and Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology*, 4(1), 67–92. https://doi.org/10.1515/lp-2013-0004

Grabe, E., and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and N. Warner, eds., *Laboratory phonology VII*, pp. 515–546. De Gruyter Mouton. https://doi.org/10.1515/9783110197105.2.515

He, L. (2012). Syllabic intensity variations as quantification of speech rhythm: Evidence from both L1 and L2. Proceedings of the Sixth International Conference on Speech Prosody 2012, Shanghai, pp. 466–469.

Huang, N. E., Shen, Z., Long, S. R., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of*

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press

London. Series A: Mathematical, Physical and Engineering Sciences, 454(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

Kohler, K. (2009). Rhythm in speech and language. *Phonetica*, 66(1–2), 29–45. https://doi.org/10.1159/000208929

Kortmann, B., and Schneider, E. W., eds. (2004). *A handbook of varieties of English. Volume 1: Phonology*. De Gruyter Mouton.

Low, E. L. (1998). Prosodic prominence in Singapore English. PhD thesis, University of Cambridge.

Low, E. L., Grabe, E., and Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43(4), 377–401. https://doi.org/10.1177/00238309000430040301

Molnar, M., Gervain, J., and Carreiras, M. (2014). Within-rhythm class native language discrimination abilities of Basque-Spanish monolingual and bilingual infants at 3.5 months of age. *Infancy*, 19(3), 326–337.

Nolan, F., and Jeon, H. S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130396. https://doi.org/10.1098/rstb.2013.0396

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

Ramus, F., Dupoux, E., and Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. *Proceedings of the 15th International Congress of Phonetic Sciences*, 3, 337–342.

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292. https://doi.org/10.1016/S0010-0277(00)00101-3

Rathcke, T., and Smith, R. (2011). Exploring timing in accents of British English. Online Proceedings of the 17th International Congress of Phonetic Sciences 2011, City University of Hong Kong, Hong Kong, pp. 1666–1669.

Ravignani, A., and Norton, P. (2017). Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. *Journal of Language Evolution*, 2(1), 4–19. https://doi.org/10.1093/jole/lzx002

Schiering, R. (2007). The phonological basis of linguistic rhythm: Cross-linguistic data and diachronic interpretation. *Sprachtypologie und Universalienforschung*, 60, 337–359. https://doi.org/10.1524/stuf.2007.60.4.337

Szakay, A. (2006). Rhythm and pitch as markers of ethnicity in New Zealand English. Proceedings of the 11th Australian International Conference on Speech Science & Technology, University of Auckland, pp. 421–426.

Tilsen, S., and Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America*, 134(1), 628–639. https://doi.org/10.1121/1.4807565

Tilsen, S., and Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America*, 124(2), EL34–EL39. https://doi.org/10.1121/1.2947626

Turk, A., and 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. https://doi.org/10.1515/lp-2013-0005

White, L., and Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35(4), 501–522. https://doi.org/10.1016/j.wocn.2007.02.003

White, L., Mattys, S. L., and Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, 66(4), 665–679. https://doi.org/10.1016/j.jml.2011.12.010

Wiget, K., White, L., Schuppler, B., et al. (2010). How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America*, 127(3), 1559–1569. https://doi.org/10.1121/1.3293004

https://doi.org/10.1017/9781009295888.036 Published online by Cambridge University Press