12. Measuring Speech Rhythm

Amalia Arvaniti

University of Kent

Postal Address

Amalia Arvaniti
University of Kent
English Language and Linguistics, SECL
Cornwallis NW
Canterbury, Kent
CT2 7NF

Email: a.arvaniti@kent.ac.uk

12. MEASURING SPEECH RHYTHM

12.0 Abstract

This paper covers the methods of measuring rhythm and the main paradigms used to study rhythm perception. An overview of ideas about speech rhythm is provided, starting with the traditional view of isochrony and rhythm classes. Production and perception methods used to establish rhythm class differences are presented and critically reviewed, as are a number of research practices associated with them. Recent developments leading to an alternative view of rhythm are discussed, and suggestions for pedagogical practice and future research are provided.

12.1 Introduction

In the linguistics literature there is no generally acceptable definition of speech rhythm; rather, two quite distinct views prevail. According to one of these views, rhythm is the same as *timing*: it is the repetition of a specific unit at regular intervals or *isochrony*; units that have been proposed to play this role include the stress foot (a group consisting of a stressed syllable and following unstressed syllables), the syllable, and the mora (see Chapter 6, this volume). According to this view, the duration of one of these units is constantly adjusted so as to remain stable throughout a speech event. This definition of rhythm has prevailed in phonetics for some 80 years and has given rise to the rhythm class typology, according to which languages fall in one of three classes, stress-, syllable-, or mora-timing. In this view, rhythm's only phonetic exponent is duration and thus rhythm should be studied by measuring the duration of specific units of speech.

The second definition reflects the understanding of rhythm in psychology and the study of music, which see rhythm as an abstraction that relies on perceiving constituents as groups of similar and repetitive pattern. Stress feet constitute such a possible grouping with the alternation of stressed and unstressed syllables creating a pattern. The concepts of syllable- and mora-timing are incompatible with this definition of rhythm, however, since the units of each class (syllables, moras) are neither grouped, nor differentiated from each other in a way that would create a pattern.

These two very different conceptualizations have influenced the types of experimental work and measurements used to study rhythm. In the rest of the paper, I trace their history, and critically discuss the assumptions behind them and the types of measurements and experimental paradigms they have given rise to.

12.2 Historical Overview

12.2.1 The Search for Isochrony

Experimental research on rhythm has a long history dating at least as far back as Bolton (1894). The study of speech rhythm and in particular the auditory impression that English has isochronous feet have been around since the early 20th century (see Classe, 1939, for a review). Classe (1939) used the kymograph to examine these earlier claims about English. Although he recorded professional talkers, such as actors, who practiced beforehand and tapped to the beat while speaking, he concluded that stress feet (his *accentual groups*) are isochronous only if the number of syllables, grammatical structure and phonetic constitution of the feet are kept as similar as possible: isochrony was found in utterances like *There were tarts, cakes, toast and coffee* but not in *He ran, jumped over it, fell and shouted*. Classe's comments on the variability in foot duration are echoed in Jones (1918), who remarked on the 'extreme difficulty of describing or reducing to rules the innumerable rhythms heard in ordinary connected speech' (1972: 242).

Despite the caution of Classe and Jones, observations about isochrony were soon extended to other languages. Lloyd James (1940: 25) argued that English, Arabic and Persian have 'Morse code rhythm', while French and Telugu have 'machine gun rhythm'. A similar idea was expressed by Pike (1945) who coined the terms *stress-timing* and *syllable-timing* for English and Spanish respectively (in a pedagogical publication for Spanish L2 learners of English), thereby giving rise to the notion of rhythm classes. Perhaps the strongest expression of the rhythm class hypothesis is that of Abercrombie (1967: 97) who stated that '[a]s far as is known, every language in the world is spoken with one kind of rhythm or with the other', and defined each as the periodic recurrence of one of two pulse systems: chest pulses producing syllables for syllable-timing, and stress-pulses regulating the duration of stress feet. (Mora-timing did not feature in this early Western literature, though according to Warner and Arai (2001), the term was first used for Japanese in Jinbo (1927).)

In the second half of the 20th century isochrony was extensively tested, though always unsuccessfully. Studies in English showed that foot duration is proportional to the number of syllables in the foot (see Lehiste, 1977, for an early review, and Nakatani et al., 1981).

Equally, studies found no evidence that syllable duration is kept constant in languages classified as syllable-timed, including Spanish (Pointon, 1980), French (Fletcher, 1991), and Italian (Farnetani & Kori, 1990). Some studies measured both syllable and foot durations and concluded that isochrony is absent from both (Balasubramanian, 1980, on Tamil; Borzone de Manrique & Signorini, 1983, and Pointon, 1995, on Spanish). In addition, studies that compared stress- and syllable-timed languages noted more similarities than differences between them (e.g. Dauer, 1983, on English, Italian, Spanish, Greek and Thai; Bertrán, 1999, on English, Russian, Spanish, Catalan, Portuguese, French and Italian). Similarly, in their review, Warner and Arai (2001) concluded that there is little evidence in support of moras being of stable duration in Japanese.

The lack of evidence for isochrony in production led Lehiste (1977) to suggest that the percept of isochrony could reflect a tendency to underestimate durational differences between speech stimuli. Thus, she argued that Just Noticeable Differences (JNDs) established by psychophysical experiments and set at around 2.5% should be revised upwards for speech to 10% of duration for feet 300-500 ms long (see also Friberg & Sundberg, 1995). Lehiste's idea was generally supported by her findings. However, many studies (including Lehiste, 1977) report durational differences between feet that are substantially larger than Lehiste's own estimate of JND.

Other perception studies did not fare better. Scott et al. (1985) had English and French speakers listen to both French and English utterances and tap in time with word-initial consonants (which they assumed corresponded to accented beats, a task reminiscent of Classe's experiments). French listeners were more isochronous than English listeners in their tapping to stimuli of both languages, though in both groups inter-tap intervals were more evenly timed than the consonant onsets in the stimuli (giving credence to the contention of Lehiste, 1977). Miller (1984) asked English and French phoneticians and non-phoneticians to place Arabic, Finnish, Indonesian, Japanese, Polish, Spanish and Yoruba into rhythm classes. The only classifications showing some agreement were that Arabic is stress-timed (all groups), Yoruba is syllable-timed (phoneticians only), and Spanish is *stress-timed* (both French groups and English non-phoneticians). Taken together, these studies do not show that languages said to belong to distinct rhythm classes are generally perceived as rhythmically distinct.

Such results led to the eventual abandonment of the quest for strict isochrony in speech and cast doubt on the concept of rhythm classes. Roach (1982) found no differences in temporal variability between English, Russian, and Arabic (classified as stress-timed), vs. French, Telugu, and Yoruba (classified as syllable-timed), and concluded that 'the stress-timed/syllable-timed distinction […] depend[s] mainly on the intuitions of speakers of various Germanic languages all of which are said to be stress-timed' (p. 78). Bertinetto (1989: 100) noted that '[…] no other phenomenon of phonology is so widely accepted [as rhythm classes], with so little supporting evidence'; he argued that the idea of rhythm classes is based on a handful of languages in which a number of factors conspire to give a particular impression, but that such synergies cannot be expected in all languages.

Similar views were expressed in Dauer (1983, 1987), whose work had a significant impact on the study of rhythm. Dauer (1983) argued that syllable-timing is a questionable concept. She suggested instead that rhythm is stress-based and languages form a continuum, from least stress-based, like Japanese, to most stress-based, like English. In Dauer's work, placement on the continuum is determined by a set of criteria, the aim of which is to assess the extent to which stresses are acoustically *and phonologically* salient in a given linguistic system. Crucially, Dauer's criteria were not based solely on duration; they include, e.g., the function(s) of pitch and the relationship between tone and stress (Dauer, 1987). Such non-durational criteria are important for assessing salience, but they are not in line with ideas about rhythm classes and rhythm as timing.

12.2.2 Rhythm Metrics

Although Dauer (1983) advocated for rhythm as a stress-based continuum, her durational criteria were adopted by Ramus et al. (1999) to support instead the traditional classes of stress-, syllable-, and mora-timing. There authors presented rhythm classes as an uncontroversial linguistic concept and cast their own work as a way of understanding and measuring these distinctions. They reduced Dauer's criteria to just two, phonologized vowel reduction, and syllable complexity (based on permissible syllable structures in a system, not structure frequency), and argued that these two properties have measurable effects on duration that reflect rhythm class affiliation. They tested several measures and concluded – in a circular fashion – that those best capturing rhythm classes are the percentage of vocalic intervals in a stretch of speech (%V), and the standard deviation of the consonantal intervals (*Δ*C).

Based on research on Singapore English (Low et al. 2000), Grabe and Low (2002) proposed an alternative set of metrics, the pairwise variability indices (PVIs). PVIs are based on the hypothesis that in languages considered syllable-timed, consecutive intervals (either vocalic or consonantal) should differ little in duration, while differences should be magnified in stress-timed languages. Grabe and Low (2002) proposed a 'raw' PVI for consonantal intervals (12.1) and a normalized nPVI, for vocalic intervals (12.2), on the grounds that the durations of the latter are more susceptible to changes of speaking rate.[1]

(12.1)

$$rPVI = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}| \,/\, (m-1)\right]$$

(12.2)

$$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left|\frac{d_k - d_{k+1}}{(d_k + d_{k+1})/_2}\right| \,/\, (m-1)\right]$$

Grabe and Low (2002) applied PVIs to 18 languages using data from 1 speaker per language. They concluded that languages form a rhythm *continuum*, as of the 18 languages they tested only 9 could be unambiguously classified. British English, Dutch and German (all Germanic and considered stress-timed), and Spanish and French (both Romance and prototypes for syllable-timing) were classified as expected. Of the remaining thirteen languages, one was classified as stress-timed (Thai), and three as syllable-timed (Japanese, Luxembourgish, and Mandarin). Nine languages, however, were declared unclassifiable (Greek, Malay, Romanian, Singapore English, Tamil, Welsh) or of mixed rhythm (Catalan, Estonian, Polish). Thus, of the thirteen non-prototypical languages in the sample, only four were classified with some success using PVIs. This gives, at best, a success rate of 33%.

Despite these issues, rhythm metrics revived interest in rhythm classes and have been extensively used to determine rhythm class affiliation for a variety of languages. These include Bulgarian (Barry et al., 2003), Tamil (Keane, 2006), Hawaiian (Parker Jones, 2006), Czech (Dankovicová & Dellwo, 2007), Latvian (Stockmal et al., 2005), Cantonese and

[1] Grabe & Low (2002) do not offer evidence in support of this assertion. Supporting evidence can be found in Gay (1981). Other studies, however, show that duration changes related to speaking rate depend on segment identity and the role of duration in the system (Kessinger and Blumstein, 1998; Arvaniti, 1999); thus they may be language specific (Arvaniti, 2001).

Mandarin (Mok, 2009), Malaysian and Singapore English (Tan & Low, 2014), and Ontario French (Kaminskaïa et al., 2016). Despite these efforts, results and classifications have remained mostly inconclusive (see Section 12.3.2).

The inability of metrics to determine rhythm class affiliation has led to a number of responses. Nolan and Jeon (2014), following Pointon (1980), have argued that speech is *antirhythmic*, 'redolent of wilful and rebellious disregard for decent metrical principles' (p. 7) and thus that rhythm should be seen as a metaphor rather than a measurable property of speech. Some researchers have classified languages as having mixed or intermediate rhythm (e.g., Baltazani, 2007, on Greek), others have proposed that a particular language belongs to a new rhythm class (Ramus et al. 2003, on Polish; see Section 12.2.3), while still others have simply accepted a generally agreed upon classification even when metric scores contradicted it (e.g. Lin & Wang, 2007, on Mandarin). When several studies of a given language are available, the results can be even more unconvincing; e.g. Grabe and Low (2002) concluded that Greek is unclassifiable, Tsiartsioni (2003) that it is syllable-timed, and Baltazani (2007) that it has mixed rhythm. Reasons for such discrepancies are discussed in Section 12.3.2.

A frequent response to issues with PVIs and %V-*Δ*C has been to propose alternative metrics said to provide better normalization for speaking-rate effects on duration, which is generally seen as a problem. Frota and Vigário (2001) used standard deviations of normalized percentages for vocalic and consonantal intervals. Wagner and Dellwo (2004) proposed YARD (Yet Another Rhythm Determination), a measure similar to the PVIs but using z-transformed syllable durations. Dellwo (2006) proposed VarCo (standard deviation divided by the mean). Nolan and Asu (2009) used nSPVI and nFPVI which measure the (normalized) duration of the syllable and the foot respectively. Metrics that do not rely on duration have been proposed as well (Lee & Todd, 2004; Rouas et al., 2005). What all these measures have in common is that they assume that the rhythm class typology is essentially correct and reflected in segment durations; we simply need to find the correct set of measures to capture it. These alternative metrics, however, suffer from the same issues as the originals, as discussed in Section 12.3.2.

12.2.3 Perceptual Experiments on Rhythm Classes

The wholesale acceptance of the rhythm class typology is also reflected in perception research. Some of this research has focused on speech processing by adults which, it is argued, relies on syllables, mora or feet depending on the rhythm class of the listeners' native language (Cutler et al., 1986; Cutler & Otake, 1994; Murty et al., 2007). Other research has focused on discrimination among infants and adults on the grounds that if languages can be discriminated they must belong to different rhythm classes (among many, Nazzi et al., 2000; Nazzi & Ramus, 2003).

The premise behind both lines of research discussed immediately above is questionable however. Several discrimination experiments have shown that varieties of the same language can also be discriminated from each other both by infants (Nazzi et al., 2000) and adults (White et al., 2012). In addition, some languages can be discriminated from both stress- and syllable-timed prototypes; e.g. Polish can be discriminated from both English and Spanish (Ramus et al., 2003). Such results indicate that putative rhythm class is not a good explanation for discrimination. Similar arguments can be made about studies in processing: the use of a particular unit during processing does not mean that production is based on it, or that other units are not useful during either processing or production (cf. Mattys & Melhorn, 2005). I return to this point in Section 12.3.4.

12.2.4 Cycling

An alternative paradigm to using metrics is *speech cycling*. In its simplest form, cycling involves speakers repeating a short phrase such as *take a pack of cards* in time with an accelerating metronome (Tajima & Port, 2003; for more complex forms, see Cummins & Port, 1998). Cycling rests on the notion of rhythm classes but defines isochrony in relative rather than absolute terms. Specifically, the assumption behind cycling is that speech – and by extension rhythm – is produced on the basis of a number of nested cycles; a foot constitutes such a cycle, with syllable cycles nested inside it.

Cycling has been tested in languages with different prosodic systems and classifications in terms of rhythm class (Cummins & Port, 1998 on English; Tajima & Port, 2003, on Japanese; Zawaydeh et al., 2002, on Arabic; Chung & Arvaniti, 2013, on Korean). The results are comparable across languages even though English and Arabic are said to be stress-timed, Japanese is said to be mora-timed, and Korean has been impossible to classify (Arvaniti, 2012a). Overall, cycling shows that when speakers have to fit short phrases within the metronome cycle, they keep metrically prominent elements in stable phase (where *phase* is the expression of the location of these elements within the cycle defined by the onset of each successive phrase). What constitutes a prominent element, however, is language-specific: in English, the prominent elements in stable phase are stressed syllables; in Korean they are the initial syllables of accentual phrases (Chung & Arvaniti, 2013). Overall, the results indicate that speech is rhythmically organized along the same principles, independently of putative rhythm class affiliation.

12.3 Critical Issues

12.3.1 Issues with the Rhythm Class Typology

Despite the lack of evidence, the rhythm class typology remains quite popular so it is worth taking a critical look at rhythm classes as a concept (see also Dauer, 1983; Bertinetto, 1989; Arvaniti, 2009; 2012a).

A significant problem is the implausibility of syllable- and mora-timing as rhythm mechanisms (the discussion here focuses on syllable-timing but the same arguments apply to mora-timing as well). The regular repetition of syllables that is the essence of syllable-timing is rhythmically a *cadence*, the simplest form of rhythm 'produced by the simple repetition of the same stimulus at a constant frequency' (Fraisse, 1982: 151); an example would be a dripping faucet or a ticking clock. For a cadence to be perceived as such, however, the stimuli must be sufficiently separated in time to be experienced as distinct, though close enough that a number of them are perceived as occurring in the *perceptual present*, which is estimated to span 3-8 s (Woodrow, 1951; Fraise, 1963, 1982; Clarke, 1999). If the temporal spacing of stimuli is less then 200 ms, *fusion* ensues, i.e. the stimuli are not perceived as distinct (Fraisse, 1982).[2] However, the typical speaking rate of languages classified as syllable- or mora-timed is much faster than 200 ms per syllable (or 5 syllables per second), with syllable duration ranging from 128–143 ms (Dauer, 1983, on Spanish, Greek and Italian; Pellegrino et al., 2011, on Italian, French, Spanish, and Japanese). At these rates, it would be practically impossible for each syllable to be reliably perceived as a distinct beat (London, 2012, Ch.2). Morae, which by definition are shorter than syllables, are an even more unlikely base for a cadence.

Additionally, when listeners are presented with cadences they tend to impose a rhythmic pattern on them (Woodrow, 1951; Fraisse, 1963, 1982), typically grouping stimuli into trochees (strong-weak) oriambs (weak strong). This tendency is known as *subjective*

[2] Fusion is evident in the impression that the noise of machine guns is a cadence (as the metaphor about 'machine-gun' rhythm implies; Lloyd James, 1940). If slowed down by a factor of four, however, each machine gun beat is a complex sequence of sounds close to that of a beating heart.

rhythmization and has been noted since Bolton (1894). This is why the ticking of a clock is often interpreted and imitated as a series of trochees. It is difficult to reconcile the perceptual interpretation of cadences with the idea that all syllables in a syllable-timed language are equally prominent: even if they were all acoustically equal (and all evidence suggests they are not), they would not be perceived as such.

Further, research on rhythm perception shows that listeners can impose or maintain a rhythm without always having overt clues (London, 2012, Ch.1). This is because listeners pay selective attention to auditory events, a phenomenon known as *dynamic attending* (Jones, 1981). Dynamic attending rests on the idea that humans cannot attend to all events (James, 1890, cited in London, 2012, Ch. 1): rather, they focus their attention to periodically occurring events and engage in *anticipation*, i.e. they continue hearing a rhythmic pattern, in the absence of regularity, once the pattern is established (Fraisse, 1982). As an illustration, dynamic attending explains the evenly spaced tapping of the participants in the study of Scott et al. (1985).

Dynamic attending means that for syllable- and mora-timing to work, speakers of languages said to be syllable- or mora-timed – in contrast to speakers of so-called stress-timed languages – must make no selection, and must be able to attend to all events in a rapidly paced series. The idea that speakers of some languages do not engage in dynamic attending is improbable and unsupported by evidence from relevant languages. Vaissière (1991) concluded that French is a ‘boundary language’ because of the importance of phrasing and phrasal boundaries in French; in her analysis syllables are not even considered a possible determiner of rhythm, or particularly important for processing. Research on Korean also shows that phrasing is paramount for processing and production (e.g. Chung & Arvaniti, 2013; Jeon & Arvaniti, 2017). In Spanish, stressed syllables are critical for acquisition (Skoruppa et al., 2009) and processing (Sebastian & Costa, 1997); the same applies to Greek (Tzakosta, 2004; Arvaniti & Rathcke, 2015), indicating that speakers of these two languages do not treat all syllables as the same. In short, syllable- and mora-timing are psychologically implausible notions, while research on so-called syllable- and mora-timed languages shows that their speakers focus either on phrasal boundaries or stresses; the latter strategy is no different from that adopted by speakers of so-called stress-timed languages.

12.3.2 Theoretical Issues with Rhythm Metrics

Despite the problems with metrics-based rhythm classification the use of metrics remains popular in phonetics, as noted in Section 12.2.2. Metrics have also been used in psycholinguistics and acquisition (Hannon et al., 2016; Harris & Gries, 2011), second language learning (Stockmal et al., 2005; White & Mattys, 2007), speech pathology (Lowit, 2014), and forensic linguistics (Harris et al., 2014). Because of this enduring popularity, a critical appraisal of metrics is needed (for additional arguments and evidence, see Arvaniti, 2009, and 2012a).

A major issue with metrics is what Renwick (2013) termed their *volatility*: metric scores vary substantially from study to study. Arvaniti (2012a) tested metrics with (American) English, (Mexican) Spanish, and (Standard) German, Italian, Greek, and Korean data elicited from eight speakers per language in three styles: scripted isolated sentences, scripted running speech, and spontaneous speech. In addition, Arvaniti (2012a) manipulated the structure of the scripted sentences to vary syllable complexity; for instance, the English materials contained both sentences like *I just called Trent to confirm the appointment we had scheduled last Monday*, which includes a variety of syllable structures, and sentences like *Two-year-old Lucy has macaroni and cheese every day for dinner*, in which syllable structure variability is limited. Arvaniti (2012a) found that in all six languages results differed depending on syllable composition and elicitation method, and also exhibited extensive inter-speaker variability. In fact, Arvaniti (2012a) showed that the effect sizes of elicitation and syllable composition were larger than the effect of language, i.e. there is bigger variation within than across languages in terms of rhythm metric scores.

Similar results from small corpora have been reported by Wiget et al. (2010) and Prieto et al. (2012) who also found effects of syllable complexity on metrics, even though they excluded from measurement liquids, glides, and phrase-final segments to make the measured intervals more uniform. Renwick (2013) and Horton and Arvaniti (2013) support these findings: Renwick (2013) found that %V correlates most strongly with the percentage of open syllables in the sample, independently of the language tested; Horton and Arvaniti (2013) found that groupings in unsupervised clustering were based on syllable composition not purported rhythm class.

Further, different metrics give widely different results. Arvaniti (2012a) showed that even metrics said to measure the same dimension do not correlate with one another; this applies, e.g., to *Δ*C, VarcoC, and PVI, all of which purport to measure consonantal interval variability.

Horton and Arvaniti (2013) found that groupings from unsupervised clustering differed depending on the metrics used (%V-ΔC, PVIs, or Varcos). Similarly, Loukina et al. (2011) noted that languages in their sample would be distinguished using one metric but not another, while Knight (2011) reports lack of correlation between metric scores obtained from the same speakers on different occasions. Harris and Gries (2011) found that metric scores are affected even by word frequencies. In practice all these findings mean that variability in metrics is unpredictable and results cannot be replicated: metrics are affected by a large number of factors and a researcher cannot predict which factors will affect their sample or how.

As mentioned in Section 12.2.2, a response to these persistent problems has been to propose new metrics or new combinations of existing metrics. However, this approach is not an improvement because the root of the problem is the circular relationship between rhythm classes and metrics: since there is no independent evidence of rhythm class membership, metric scores are predicted based on putative rhythm class, but the same scores are also used to classify languages for rhythm class. This circularity makes it impossible to test the validity of metrics. Thus, one can argue that %V and ΔC are more accurate than PVIs because they classify Japanese as mora-timed (Ramus et al., 1999), while PVIs group it with syllable-timed languages (Grabe & Low, 2002), but this conclusion is meaningless because independent evidence for the assertion that Japanese is mora-timed is lacking.

In the face of these issues, an approach recently taken by some metric proponents is to argue that although metrics cannot be used for rhythm classification, and that such classification may indeed be ill-advised, metrics are still useful because they provide us with reliable information about timing (e.g. Post & Payne, 2018). The problem with this view is that the results obtained from metrics are still volatile and influenced by a host of factors as detailed above. Given that timing is highly complex and language- and speaker-specific (see Klatt, 1976, Turk & Shattuck-Hufnagel, 2000, and Arvaniti, 2009, for reviews), it is unrealistic to expect that segmental durations will vary in some uniform manner, across languages, corpora, and even recording sessions.

An additional issue is that all *metrics are global measures of local effects*: this means that different timing patterns can have similar effects on metric scores but for entirely different reasons. For instance, Arvaniti (2009) reports that Korean and Spanish L2 speakers of English can achieve metric scores comparable to those of L1 English speakers, but due to very different strategies: Korean speakers show extreme phrase-final vowel lengthening, while

Spanish speakers show elision of intervocalic consonants which results in long vocalic stretches. Although both practices contribute to vocalic variability, their auditory effects are very different and could hardly be said to help speakers achieve a rhythm similar to that of L1 English. Lowit (2014) reached the same conclusions about metrics examining disordered speech.

Given these findings, the use of metrics can at best lead to claims that a particular sample is more or less syllable- or stress-timed than another. However, on their own, such assertions are uninterpretable because unless one knows the source of variability, it is impossible to identify it from metric scores: metrics are global measures of local effects and as such they cannot tell us if durational variability is a global characteristic of the speech measured, or a local but substantial event, as the examples discussed immediately above indicate.

In conclusion, results based on metrics cannot be used to support the rhythm class typology and should not be used to determine the rhythm class of a given language as the process is circular. Further, there is abundant evidence that results obtained using metrics are non-replicable, and susceptible to a great deal of variability due to a large number of factors the exact effects of which on a given sample cannot be predicted. Further, these effects are opaque: they do not and cannot reflect the origins of the variation they measure. In short, metrics are a poor, unreliable and not particularly illuminating measure of both timing and rhythm.

12.3.3 Issues with Practice when Using Metrics

The issues discussed in Sections 12.3.1 and 12.3.2 clearly show that metrics should be abandoned s they are not a reliable measure. If this is not possible, however, at a minimum the most egregious practices associated with metric use should be avoided.

One such practice is to compare metric scores from a test language to scores from a small set of languages, such as English and French, seen as arbiters of rhythm class standards and used to determine the rhythm class of the test language; e.g. the PVIs of English and French reported in Grabe and Low (2000) were used by Lin and Wang (2007) to determine the rhythm class of Mandarin; metrics for German, English, French and Italian data from the BonnTempo corpus (Dellwo et al., 2004) were used by Dankovicová and Dellwo (2007) to determine the rhythm class of Czech, and by Mok (2009) to determine the rhythm class of Cantonese and Mandarin. This practice assumes that metric scores are immutable and represent something essential about languages. However, as noted in Section 12.3.2., metrics are volatile. Thus, a score obtained in a given study is nothing more than one point in a large range (Arvaniti, 2012a). Therefore, the only way to ascertain that metric scores from different languages are comparable is to obtain both samples in exactly the same way, to control for syllable composition (e.g. by ensuring that the samples contain all syllable structures permissible in each language and do so with the appropriate frequency), and to carefully examine inter-speaker variability within each language's corpus. If these conditions cannot be met, a comparison should not be made. For the same reasons, the practice of relying on one speaker (or a small number of speakers) is also inadvisable.

Even if the above advice is followed, an additional problem arises from the above practice: the circularity of metrics makes it impossible to determine what counts as 'similar' in terms of metric scores. Arvaniti (2009) compared the scores of the languages studied by Grabe and Low (2002) and showed that differences between languages supposedly in the same rhythm class are as large as differences between languages said to be in different classes; see Table 12.1. This is one more reason why such similarity comparisons for the purpose of rhythm classification should be abandoned.

Table 12.1. Comparison of nPVI scores between languages said to belong to different classes (left) and the same class (right); based on Grabe and Low (2002).

Different purported rhythm class	**Same purported rhythm class**
$\text{nPVI}_{\text{British English}} - \text{nPVI}_{\text{French}} = 13.7$	$\text{rPVI}_{\text{British English}} - \text{rPVI}_{\text{German}} = 8.8$
$\text{rPVI}_{\text{German}} - \text{rPVI}_{\text{Spanish}} = -2.4$	$\text{nPVI}_{\text{French}} - \text{nPVI}_{\text{Mandarin}} = 16.5$
$\text{rPVI}_{\text{Thai}} - \text{rPVI}_{\text{Spanish}} = 1.2$	$\text{nPVI}_{\text{French}} - \text{nPVI}_{\text{Spanish}} = 13.8$
$\text{nPVI}_{\text{Singapore English}} - \text{nPVI}_{\text{Dutch}} = 13.2$	$\text{nPVI}_{\text{Singapore English}} - \text{nPVI}_{\text{Spanish}} = 22.6$

A final risky practice is the use of several metrics in the same study followed by the reporting of results from those metrics that showed statistically significant differences according to some factor of interest (e.g. Ramus et al., 1999; White & Mattys, 2007; Li & Post, 2014; Kaminskaïa et al., 2016). The problem with this approach is statistical, as the large number of measures used for the same purposes increases the chances of Type 1 error (falsely rejecting the null hypothesis; Bohannon et al., 2015). This practice should be avoided.

12.3.4. Issues with Perception Paradigms

As mentioned in Section 12.2.3, many perception studies are based on the idea that languages from different rhythm classes can be discriminated while languages from the same class cannot (e.g. Nazzi & Ramus, 2003). In order for listeners to focus on timing, these studies have often used flat *sasasa* stimuli, a type of modified speech in which F0 is flat (slightly falling throughout an utterance), and all vocalic intervals are replaced by [a] and all consonantal intervals by [s]. For example, *dinner's ready* would be rendered as *sasasasa* with the following intervals: [d][ɪ].[n][ə].[zɹ][e].[d][i]. Flat *sasasa* has been used with the AAX (odd-ball) task, in which listeners hear two stimuli from the same language (context) followed by a third stimulus either from the same or a different language, and have to respond whether the third stimulus is from the same language as the context or not.

The idea that discrimination is only possible between rhythm classes is not supported by the results of related studies and has led to some improbable classifications. Moon-Hwan (2004) concluded that Korean is mora-timed because Korean and Italian listeners discriminated Korean from both English and Italian but not from Japanese. His conclusion, however, is unsupported by evidence from Korean phonology and processing. Further, many experiments have shown that discrimination is possible not only between but also within rhythm classes, including discrimination by infants of dialects of the same language (among many, Nazzi et al., 2000; Ramus et al., 2003; White et al. 2012; Molnar et al., 2014). Since these experiments are widely used, it is worth considering why their results are inconsistent.

One issue is that *sasasa* stimuli are extremely impoverished. This makes discrimination difficult, leading to any features remaining in the signal acquiring exceptional importance as study participants are forced to use any information they can to complete the task (cf. Hawkins, 2003, who commented that hair length is critical in determining gender in stick figures but not in real humans). For example, it is likely that listeners in Arvaniti and Rodriquez (2013) used the extreme final lengthening of Korean to discriminate it from English. Discrimination, however, does not mean that listeners rely on rhythm class, or that the feature they relied on is the one they would use when listening to real speech.

Further, there are reasons to think that *sasasa* is not an ecologically valid manipulation of the speech signal. The *sasasa* transform is based on the assumption that listeners not only estimate the duration of consonantal and vocalic intervals (and derive complex calculations from this information during online speech processing) but that they also do so independently of other prosodic information. It is well known, however, that listeners integrate prosodic information during processing: duration and amplitude interact so that a longer sound sounds louder than a shorter sound with the same average intensity (Beckman, 1986), while pitch changes influence the perception of rhythm (Dilley & McAuley, 2008; Kohler, 2009). These interactions are reflected in the fact that *sasasa* stimuli are responded to differently from richer signals, such as low-pass filtered speech (Arvaniti, 2012b). Even the information retained in *sasasa* can affect responses. Arvaniti and Rodriquez (2013) ran a series of AAX experiments, using four types of *sasasa*: flat *sasasa*, *sasasa* in which stimuli retained the F0 of the utterances from which they were derived, and versions of both in which all stimuli had the same speaking rate but proportional differences in the duration of vocalic and consonant intervals were retained. Discrimination rates differed depending on the type of *sasasa* used: retaining F0 helped discriminate English from Korean, while retaining speaking rate helped discriminate English from Greek. What all these results show is that listeners do not compute some global timing profile for each utterance, and do not process timing patterns independently of other prosodic parameters (see White et al., 2012, for similar conclusions). If they did, they would give consistent responses to stimuli with the same timing profiles, such as the four types of stimuli in Arvaniti and Rodriquez (2013), independently of the presence of other prosodic information.

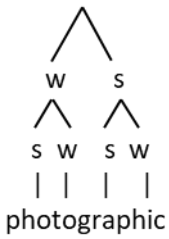
In short, experiments may lead to discrimination between languages for any number of reasons, while lack of discrimination does not necessarily mean that the languages involved are rhythmically similar. The same applies to studies on processing (see Section 12.2.3): although these studies do show that a particular constituent in the phonological organization of a given language is salient and useful to native listeners during processing, this finding neither entails membership to a rhythm class nor precludes the usefulness of other prosodic units (e.g. Mattys & Melhorn, 2005, on syllables in English; Skoruppa et al., 2009, and Arvaniti & Rathcke, 2015, on stress in Spanish and Greek respectively).

12.4. Recent Research

The issues with categorizing languages by rhythm class, coupled with the problems with metrics have led to alternative proposals. Arvaniti (2009) argued for the principled separation of rhythm from timing in phonetic research, with *timing* being used to refer to the organization of segmental duration in the speech signal, and *rhythm* seen as being based on 'the perception of series of stimuli as series of groups of similar and repetitive pattern' (Arvaniti, 2009: 57). This definition is based on the psychological understanding of rhythm (as expressed, e.g., in Fraisse (1963, 1982), London (2012)), and echoes aspects of Dauer (1983).

In this view, rhythm is an abstract representation of prominence relationships within groups: listeners extract the rhythmic pattern of an utterance by noting such differences and the temporal spacing of elements that allows them to perceive grouping (Woodrow, 1951; Fraise, 1963, 1982). Which units speakers of a given language focus on as the main grouping element in rhythm (e.g. feet or phrases) and which parameters make some elements (such as syllables or moras) more salient than others is language specific. This view of rhythm is in line with psychological definitions, as noted, and also closer to the conception of rhythm used in phonology (e.g. Hayes, 1995), in which rhythm is also seen as relying on the relative salience of a number of constituents. Specifically, the phonological approach to rhythm assumes a hierarchical relationship between elements: e.g. a word like *photographic* consists of two feet, each with one strong and one weak syllable, but the last foot is also stronger than the first, as illustrated in (12.3). This hierarchical view of rhythm is also in line with work on the psychology of rhythm in music (cf. Lerdahl & Jackendoff, 1981; London, 2012, Ch. 1), and is supported by evidence from articulation (Tilsen, 2016).

(12.3)



This approach requires that we recognize that durational variability plays only a small role in the creation of speech rhythm and that, therefore, the importance attributed to it in phonetic studies is overestimated. Dauer's (1983) finding that inter-stress interval duration is very similar across widely different languages supports this conclusion (see Table 12.2). In all these languages, prominent syllables appear approximately every half second, a pace that is conducive to perceiving rhythm, in that it allows for 6-10 beats within the perceptual present. It is also a pace close to what is known as *preferred* or *natural tempo* – the spacing of events every 0.5–0.6 s; this pace is considered neither too slow nor too fast and is likely to induce entrainment between aural stimuli and human tapping (Fraisse, 1982; Dowling & Harwood, 1986; Clarke, 1999). This rate of stress occurrence is not expected to be constant in terms of segments or syllables, given the different speaking rates and patterns of elision of different languages (Barry & Andreeva, 2001; Pellegrino et al., 2011). Equally, these inter-stress intervals need not always be regular for listeners to construe a rhythmic pattern, thanks to subjective rhythmization and anticipation (see Section 12.3.1).[3]

Table 12.2. Average duration of interstress intervals by language (after Dauer, 1983).

Language	Purported rhythm class	Interstress interval duration
Thai	Stress-timed	308 ms
Italian	Syllable-timed	468 ms
Spanish	Syllable-timed	477 ms
Greek	Syllable-timed or mixed	483 ms
English	Stress-timed	493 ms

The fact that duration is not the only exponent of rhythm is supported by recent results in a number of languages. Jeon and Arvaniti (2017) carried out a fragment monitoring experiment in Korean and found that strictly regular rhythm (operationalized as number of syllables per accentual phrase) was not helpful to listeners, who relied primarily on F0 cues for processing. These results indicate that the repetitive pattern that creates rhythm in Korean may be F0-based (since accentual phrases tend have a stable LH pitch contour), and address the issue of how rhythm is created in languages like Korean which do not have stress or foot structure (Jun, 2005).

Rhythm may also be based on alternations in amplitude. Tilsen and Arvaniti (2013) used empirical mode decomposition (EMD) to uncover regularities in the amplitude envelop. They found that the mean instantaneous frequencies for the first two modes are comparable in their corpus (that of Arvaniti, 2012a) which included English, German, Greek, Italian, Korean and

[3] English has mechanisms to bring about more regular rhythm, or *eurhythmy*. These are BEAT ADDITION, used to avoid *lapses* i.e. long stretches of unstressed syllables, and BEAT DELETION (also known as RHYTHM RULE OR IAMBIC REVERSAL) which alters the relative prominence of stresses to avoid *clashes*, i.e. adjacent syllables of equal prominence (among many, Nespor & Vogel, 1989; Hayes, 1995). It is unclear whether many other languages have similar mechanisms.

Spanish: the mean for the instantaneous frequency of the first mode was 5.7–6.7 Hz, and that of the second mode was 2.3–2.6 Hz. These frequencies correspond well to expected frequencies for vowels and heads of feet (or units of phrasing comparable to feet, such as the accentual phrase of Korean; Jun, 2005). These findings suggest that languages share many similarities in terms of rhythmic structure: these general patterns apply both to English, the prototypical stress-timed language, and Korean, a language without stress. These similarities are likely to come from motor control and articulatory restrictions, but may also have neurological underpinnings: they are within the Theta and Delta bands (4–10 Hz and 1.5–4 Hz respectively) of oscillating networks of neurons which have been shown to be crucial for speech processing and rhythm entrainment (Luo & Poeppel 2007; Goswami 2011; Goswami & Leong, 2013). In short, the studies briefly reviewed here provide *prima facie* evidence that basic periodicities creating prominence-based groupings are present in speech and likely relevant for processing given their frequency ranges.

12.5 Best Practice for Teaching and Learning

Teaching rhythm classes should be presented as a relic students should be familiar with, but study should focus on understanding the issues with rhythm classes, and on exposing students to newer approaches. The latter could include teaching about speech timing and rhythm, presenting phonological and psychological perspectives, and familiarizing students with the variety of prosodic parameters that can be used to create rhythm in speech (including the variable realization of stress cross-linguistically). To this end, mini-perceptual experiments can be conducted by modifying the speech signal to induce changes in grouping and prominence.

Since the idea of rhythm classes is still quite prevalent, bringing the issues with rhythm classes to student attention is essential. This could be done by having students listen to unfamiliar languages and try to classify them for rhythm class. This is particularly effective if the languages used have an accepted classification, as listening exposes students to the vast differences between languages said to be in the same rhythmic class; the audio files of Illustrations of the IPA published in the *Journal of the International Phonetic Association* are a good source (and openly available from https://www.cambridge.org/core/journals/journal-of-the-international-phonetic-association/illustrations-of-the-ipa-free-content). Further, using varieties of the same language can highlight the fact that rhythmic impressions can vary quite significantly even within a language.

Tasks may also involve the use of metrics to expose the issues with reliability. Students may record sentences, calculate a number of metrics, and compare scores across the class. They may also record sentences at different speaking rates, or do a repeat recording after a short break and compare their own scores across speaking rates and repetitions. Finally students can calculate metrics using different measurement criteria, so as to observe the difference that excluding some intervals (such as phrase-final segments) can make. Such exercises should be followed by an examination of the speech signal so students understand the possible sources of variation.

12.6 Future Directions

The review presented above has several implications for future research on rhythm. First, it shows that focusing exclusively on measuring duration, in whatever form, is misguided: local durational effects are unlikely to be as important for rhythm as the induction of abstract patterns of periodicity that are the basis of the mental representation of rhythm (Lerdahl & Jackendoff, 1981; Clarke, 1999, and references therein). Such abstract patterns may be created by a variety of means, including phrasing, and changes in amplitude and F0 (Tilsen & Arvaniti, 2013; Jeon & Arvaniti, 2017). These means of inducing rhythm should be investigated further, using production, entrainment, and perception studies. Investigation should delve into the role of rhythm in spoken communication (Auer et al., 1999), and in entrainment both between speech and gesture (Loehr, 2007) and across speakers (Cummins, 2009), and into the function of occasional rhythmic regularity in speech (Hawkins, 2014). The same applies to established interactions between percepts of rhythm, grouping and intonation (Dilley & McAuley, 2008; Kohler, 2009), which should also be further explored. Since it is clearly not the case that the same parameters create rhythm in all languages, a larger gamut of languages should be investigated, and the focus on Japanese and a handful or Romance and Germanic languages should be shifted to an in-depth study of languages with a variety of prosodic systems. Further, phonetics should engage more with the perception of rhythm rather than simply focusing on its acoustic manifestation, precisely because rhythm percepts cannot be determined from acoustic measurements alone. Finally, native speakers should not be ignored, as is often done; rather, the ways they experience rhythm in their language should be taken seriously into consideration when designing experiments (Arvaniti, 2009). Moving away from a system that relies on a repeatedly falsified rhythm typology and designing studies that do not rely on a dubious classificatory system is paramount for future success in the study of rhythm.

12.7 References

Abercrombie, D. (1967). *Elements of General Phonetic*s, Edinburgh: Edinburgh University Press.

Arvaniti, A. (1999). Effects of speaking rate on the timing of single and geminate sonorants. *Proceedings of the XIVth International Congress of Phonetic Sciences*. San Francisco, CA, pp. 599-602.

Arvaniti, A. (2001). Comparing the phonetics of single and geminate consonants in Cypriot and Standard Greek. *Proceedings of the 4th International Conference on Greek Linguistics*. Thessaloniki: University Studio Press, pp. 37-44.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica* 66, 46–63.

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

Arvaniti, A. (2012*b)*. Rhythm classes and speech perception. In O. Niebuhr and H. Pfitzinger, eds., *Prosodies: Context, Function, and Communication*. Walter de Gruyter, pp. 75–92.

Arvaniti, A. & Rathcke, T. (2015). The role of stress in syllable monitoring. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow. http://www.icphs2015.info/pdfs/Papers/ICPHS0212.pdf.

Arvaniti, A. & Rodriquez, T. (2013). The role of rhythm class, speaking rate, and *F0* in language discrimination. *Laboratory Phonology,* 4(1), 7–38.

Auer, P., Couper-Kuhlen, E. & Müller, F. (1999). *Language in Time: The Rhythm and Tempo of Spoken Interaction*. New York: Oxford University Press.

Balasubramanian, T. (1980). Timing in Tamil. *Journal of Phonetics*, 8, 449–467.

Baltazani, M. (2007). Prosodic rhythm and the status of vowel reduction in Greek. *Selected papers on Theoretical and Applied Linguistics from the 17th international symposium on Theoretical and Applied Linguistics*, vol. 1. Thessaloniki: Department of Theoretical and Applied Linguistics, pp. 31–43.

Barry, W. & Andreeva, B. (2001). Cross-Language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association*, 31, 51–66.

Barry, W.J., Andreeva, B., Russo, M., Dimitrova, S. & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? *Proceedings of 15th ICPhS*, Barcelona, pp. 2693–2696.

Beckman, M. E. (1986). *Stress and Non-Stress Accent*, Dordrecht: Foris.

Bertinetto, P. M. (1989). Reflections on the dichotomy 'stress' vs. 'syllable-timing'. *Revue de Phonétique Appliquée*, 91-92-93: 99–130.

Bertrán, A. P. (1999). Prosodic typology: On the dichotomy between *stress*-timed and *syllable*-timed languages. *Language Design*, 2, 103–130.

Bohannon, J., Koch, D., Homm, P., & Driehaus, A. (2015). Chocolate with high cocoa content as a weight-loss accelerator. *International Archives of Medicine, Section: Endocrinology* 8(55). https://doi.org/10.3823/1654.

Bolton, T. L. (1894). Rhythm. *The American Journal of Psychology*, 6(2), 145–238.

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

Chung, Y. & Arvaniti, A. (2013). Speech rhythm in Korean: Experiments in speech cycling. *Proceedings of Meetings on Acoustics* (*POMA*): *Proceedings of 21st International Congress of Acoustics, Montréal, 2–7 June 2013.* Available from http://scitation.aip.org/content/asa/journal/poma

Clarke, E. F. (1999). Rhythm and timing in music. In D. Deutsch, ed., *The Psychology of Music*. New York: Academic Press, pp. 473–500.

Classe, A. (1939). *The Rhythm of English Prose*, Oxford: Basil Blackwell.

Cummins, F. & Port, R. F. (1998). Rhythmic constraints on stress-timing in English. *Journal of* Phonetics, 31, 139–148.

Cummins, F. (2009). Rhythm as an affordance for the entrainment of movement. *Phonetica*, 66(1-2), 15–28.

Cutler, A., Mehler, J., Norris, D. & Seguí, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385–400.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33, 824–844.

Dankovicová, J. & Dellwo, V. (2007). Czech speech rhythm and the rhythm class hypothesis. *Proceedings of 16th ICPhS*. Saarbrücken, Germany, pp. 1241–1244.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51–62.

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of 11th ICPhS*. Tallinn, pp. 447–449.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski and I. Szigeti, eds., *Language and Language-processing: Proceedings of the 38th Linguistic Colloquium*. Frankfurt: Peter Lang, pp. 231–241.

Dellwo, V., Aschenberner, B., Dancovicova, J. & Wagner, P. (2004). The BonnTempo-Corpus and Tools: A database for the combined study of speech rhythm and rate. In *Proceedings of the 8th ICSLP*, Jeju Island, Korea, pp. 777-780.

Dilley, L. C. & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59, 294–311.

Dowling, W. J. & Harwood, D. L. (1986). *Music Cognition*. Orlando: Academic Press.

Farnetani, E. & Kori, S. (1990). Rhythmic structure in Italian noun phrases: A study of vowel durations. *Phonetica*, 47, 50–65.

Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics*, 19(2), 193–212.

Fraisse, P. (1963). *The Psychology of Time*, New York: Harper and Row.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (ed.), *The Psychology of Music*. New York: Academic Press, pp. 149–180.

Friberg, A. & Sunberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America*, 98(5), 2524–2531.

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

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, 38, 148–158.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, 15, 3–10.

Goswami, U. & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology*, 4(1), 67–92.

Grabe, E. & Low, E. L. (2002). Acoustic correlates of rhythm class. In C. Gussenhoven and N. Warner, eds., *Laboratory Phonology* 7. Berlin/New York: Mouton de Gruyter, pp. 515–546.

Hannon, E. E., Lévêque, Y., Nave, K. M. & Trehub, S. E. (2016). Exaggeration of language-specific rhythms in English and French children's songs. *Frontiers of Psychology* 2016, 7, 939. https://doi.org/10.3389/fpsyg.2016.00939

Harris, M. J. & Gries, S. T. (2011). Measures of speech rhythm and the role of corpus-based word frequency: a multifactorial comparison of Spanish(-English) speakers. *International Journal of English Studies*, 11(2), 1–22.

Harris, M. J., Gries, S. T. & Miglio, V. G. (2014). Prosody and its applications to forensic linguistics. *Linguistic Evidence in Security, Law and Intelligence*, 2. https://doi.org/10.5195/lesli.2014.12

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of* Phonetics, 31, 373–405.

Hawkins, S. (2014). Situational influences on rhythmicity in speech, music, and their interaction. *Philosophical Transactions of the Royal Society of London B*. 369: 20130398. doi: https://dx.doi.org/10.1098/rstb.2013.0398

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*, Chicago: University of Chicago Press.

Horton, R. & Arvaniti, A. (2013). Cluster and classes in the rhythm metrics. *San Diego Linguistic Papers*, 4. http://escholarship.org/uc/item/0tt1j553.

James, W. (1950). *The Principles of Psychology*, Ney York: Dover Reprint. (Originally published 1890)

Jeon, H. & Arvaniti, A. (2017). The effects of prosodic context on word segmentation: rhythmic irregularity and localised lengthening in Korean. *Journal of the Acoustical Society of America*, 141, 4251–4263.

Jinbo, K. (1980). Kokugo no onseijou no tokushitsu [The top phonetic characteristics of Japanese]. In T. Shibata, H. Kitamura and H. Kindaichi (eds.), *Nihon no gengogaku* [*Linguistics of Japan*]. Tokyo: Taishukan, pp. 5–15. (Originally published 1927)

Jones, D. (1972). *An Outline of English Phonetics*, 9th edn, Cambridge: Cambridge University Press. (Originally published 1918)

Jones, M. R. (1981). Only time can tell: on the topology of mental space and time. *Critical Inquiry*, 7, 557–576.

Jun, S. (2005). Korean intonational phonology and prosodic transcription. In S. Jun, ed., *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, pp. 201–229.

Kaminskaïa, S., Tennant, J. & Russell, A. (2016). Prosodic rhythm in Ontario French. *Journal of French Language Studies*, 26(2), 183–208.

Keane, E. (2006). Rhythmic characteristics of colloquial and formal Tamil. *Language and Speech*, 49, 299–332.

Kessinger, R. H. & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, 26(2), 117–128.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208–1221.

Knight, R. (2011). Assessing the temporal reliability of rhythm metrics. *Journal of the International Phonetic Association*, 41(3), 271–281.

Kohler, K. (2009). Rhythm in speech and language. A new research paradigm. *Phonetica*, 66, 29–45.

Lee, C. S. & Todd, N. P. M. A. (2004). Towards an auditory account of speech rhythm: Application of a model of the auditory 'primal-sketch' to two multi-language corpora. *Cognition*, 9, 225–254.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5, 253–263.

Lerdahl, F. & Jackendoff, R. (1981). *A Generative Theory of Tonal Music*. Cambridge, MA: The MIT Press.

Li, A. & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm. *Studies in Second Language Acquisition*, 36(2), 223–255.

Lin, H. & Wang, Q. (2007). Mandarin rhythm: An acoustic study. *Journal of Chinese Language and Computing*, 17(3), 127–140.

Lloyd James, A. (1940). *Speech Signals in Telephony*, London: Pitman & Sons.

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, 72, 179–214.

London, J. (2012). *Hearing in Time: Psychological Aspects of Musical Meter*, Oxford: Oxford University Press.

Loukina A., Kochanski, G., Rosner, B., Keane, E. & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *Journal of the Acoustical Society of America*, 129(5), 3258–70.

Low, E. L., Grabe, E. & Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'syllable-timing' in Singapore English. *Language and Speech*, 43, 377–401.

Lowit, A. (2014). Quantification of rhythm problems in disordered speech: a re-evaluation. *Philosophical Transactions of the Royal Society B*, 369 (1658). https://doi.org/10.1098/rstb.2013.0404

Luo, H. & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54, 1001–1010.

Mattys, S. L. & Melhorn, J. F. (2005). How do syllables contribute to the perception of spoken English? Insight from the migration paradigm. *Language and Speech*, 48(2), 223–253.

Miller, M. (1984). On the perception of rhythm. *Journal of Phonetics*, 12, 75–83.

Mok, P. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics*, 2, 148–154.

Molnar, M., Gervain, J. & 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.

Moon-Hwan, C. (2004). Rhythm typology of Korean speech. *Cognitive Processing*, 5, 249–253.

Murty, L., Otake, T. & Cutler, A. (2007). Perceptual tests of rhythmic similarity: I. Mora rhythm. *Language and Speech*, 50, 77–99.

Nakatani, L. H., O'Connor, K. D. & Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica*, 38, 84–106.

Nazzi, T., Jusczyk, P.W. & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, 43, 1–19.

Nazzi, T. & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by Infants. *Speech Communication*, 41, 233–243.

Nespor, M. & Vogel, I. (1989). On clashes and lapses. *Phonology*, 6, 69–116.

Nolan, F. & Asu, E. L. (2009). The Pairwise Variability Index and coexisting rhythms in language. *Phonetica*, 66, 64–77.

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

Parker Jones, O. (2006). Durational variability and stress-timing in Hawaiian. In P. Warren and C. I. Watson, eds., *Proceedings of the 11th Australian International Conference on Speech & Science Technology*, pp. 417–420.

Pellegrino, F., Coupé, C. & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 87, 539–558.

Pike, K. (1945). *The Intonation of American English*, Ann-Arbor: University of Michigan Press.

Pointon, G. E. (1980). Is Spanish really syllable-timed? *Journal of Phonetics, 8*, 293–304.

Pointon, G. E. (1995). Rhythm and duration in Spanish. In J. W. Lewis, ed., *Studies in General and English Phonetics: Essays in Honour of Professor J. D. O'Connor*. New York: Routledge, pp. 266–269.

Post, B. & Payne, E. (2018). Speech rhythm in development: What is the child acquiring? In P. Prieto and N. Esteve-Gibert, eds., *The Development of Prosody in First Language Acquisition*. John Benjamins, pp. 125–144.

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

Ramus, F., Dupoux, E. & Mehler, J. (2003). The psychological reality of rhythm class: Perceptual studies. *Proceedings of the XVth ICPhS*. Barcelona, pp. 337–340.

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

Renwick, M. E. L. (2013). Quantifying rhythm: Interspeaker variation in %V. *Proceedings of Meetings on Acoustics* (*POMA*), 14: 060011. http://dx.doi.org/10.1121/1.4854657

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal, ed., *Linguistic Controversies: Essays in Linguistic Theory and Practice in Honour of F. R. Palmer*. London: Edward Arnold, pp. 73–79.

Rouas, J., Farinas, J., Pellegrino, F. & André-Obrecht, R. (2005). Rhythmic unit extraction and modelling for automatic language identification. *Speech Communication*, 47, 436–456.

Scott, D., Isard, S. D. & de Boysson-Bardies, B. (1985). Perceptual isochrony in English and French. *Journal of Phonetics*, 13, 155–162.

Sebastian, N. & Costa, A. (1997). Metrical information in speech segmentation in Spanish. *Language and Cognitive Processes*, 12 (5-6), 883–887.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Limissuri, R. A., & Peperkamp, S. (2009). Language-specific stress perception by nine-month-old French and Spanish infants. *Developmental Science*, 12(6), 914–919.

Stockmal, V., Markus, D. & Bond, D. (2005). Measures of native and non-native rhythm in a quantity language. *Language and Speech*, 48, 55–63.

Tajima, K. & Port, R. F. (2003). Speech rhythm in English and Japanese. In J. Local, R. Ogden and R. Temple, eds., *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, pp. 322-339.

Tan, R. S. K. & Low, E. L. (2014). Rhythmic patterning in Malaysian and Singapore English. *Language and Speech*, 57(2), 196–214.

Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, 55, 53–77.

Tilsen, S. & 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.

Tsiartsioni E. (2003). *The Acquisition of Features of Rhythm and Stop Voicing in Greek and English L2*. Unpublished M.Phil. Dissertation, Trinity College Dublin.

Turk, A. E. & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, 28, 397–440.

Tzakosta, M. (2004). Acquiring variable stress in Greek: An Optimality-Theoretic approach. *Journal of Greek Linguistics*, 5, 97–125.

Vaissière, J. (1991). Rhythm, accentuation and final lengthening in French. In J. Sundberg, L. Nord and R. Carlson, eds., *Music, Language, Speech and Brain*. London: Palgrave, pp. 108–120.

Wanger, P. S. & Dellwo, V. (2004). Introducing YARD (Yet Another Rhythm Determination) and re-introducing isochrony to rhythm research. *Proceedings of Speech Prosody*, Nara, Japan, 2004. http://www.isca-speech.org/iscaweb/index.php/archive/online-archive.

Warner, N. & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica*, 58, 1–25.

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

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

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

Woodrow, H. (1951). Time perception. In S. S. Stevens, ed., *Handbook of Experimental Psychology*. New York: Wiley, pp. 1224–1236.

Zawaydeh, B. A., Tajima, K. & Kitahara, M. (2002). Discovering Arabic rhythm through a speech cycling task. In D. B. Parkinson and E. Benmamoun, eds., *Perspectives on Arabic Linguistics XIII-XIV*. Amsterdam/Philadelphia: John Benjamins, pp. 39–58.