Summary and Keywords

Prosody is an umbrella term used to cover a variety of interconnected and interacting phenomena, namely stress, rhythm, phrasing, and intonation. The phonetic expression of prosody relies on a number of parameters, including duration, amplitude, and fundamental frequency (F0). The same parameters are also used to encode lexical contrasts (such as tone), as well as paralinguistic phenomena (such as anger, boredom, and excitement). Further, the exact function and organization of the phonetic parameters used for prosody differ across languages. These considerations make it imperative to distinguish the linguistic phenomena that make up prosody from their phonetic exponents, and similarly to distinguish between the linguistic and paralinguistic uses of the latter. A comprehensive understanding of prosody relies on the idea that speech is prosodically organized into phrasal constituents, the edges of which are phonetically marked in a number of ways, for example, by articulatory strengthening in the beginning and lengthening at the end. Phrases are also internally organized either by stress, that is around syllables that are more salient relative to others (as in English and Spanish), or by the repetition of a relatively stable tonal pattern over short phrases (as in Korean, Japanese, and French). Both types of organization give rise to rhythm, the perception of speech as consisting of groups of a similar and repetitive pattern. Tonal specification over phrases is also used for intonation purposes, that is, to mark phrasal boundaries, and express information structure and pragmatic meaning. Taken together, the components of prosody help with the organization and planning of speech, while prosodic cues are used by listeners during both language acquisition and speech processing. Importantly, prosody does not operate independently of segments; rather, it profoundly affects segment realization, making the incorporation of an understanding of prosody into experimental design essential for most phonetic research.

Keywords: stress, rhythm, intonation, phrasing, timing, prosody, suprasegmentals, duration, pitch, amplitude

1. Introduction

Prosody is an umbrella term used to cover a variety of interconnected and interacting phenomena, namely stress, rhythm, phrasing, and intonation. A term that was extensively used in the past and remains popular today is the term *suprasegmentals*; it is the title of

Lehiste's classic monograph on the topic (*Suprasegmentals*, 1970), and also used in Ladd (2008, chap. 1). The term *suprasegmentals* will be avoided here as it alludes to a two-layered view of speech, whereby consonants and vowels constitute one layer and prosody is seen as the icing on a cake, a decorative and optional component that does not interfere with the integrity of the main segmental layer. This metaphor is evident in descriptions of speech as being produced "without prosody" (e.g., Conderman & Strobel, 2010; Wingfield, Lahar, & Stine, 1989; Witteman, van Heuven, & Schiller, 2012). Although for the purposes of analysis the principled distinction between segments and various components of prosody is desirable, the idea that segments are independent of prosody does not hold at the phonetic level: not only is it impossible to produce speech without prosody, but segments are strongly influenced by all aspects of prosodic structure, as discussed in some detail in this article.

What is often meant by speech without prosody is the absence of certain marked patterns associated with emotion and affect (also known as *affective*, *emotive*, or *emotional prosody*). This use of the term *prosody* is not covered here, as it refers to *paralinguistic* functions of the phonetic parameters that also encode (linguistic) prosody. Though paralinguistic phenomena are beyond the scope of the present article, it is worth considering why this confusion has arisen and how one can make a principled distinction between (linguistic) prosody and paralinguistics. In the words of Ladd (2008, p. 34) "paralinguistic messages deal primarily with basic aspects of interpersonal interaction—such as aggression, appeasement, solidarity, condescension—and with the speaker's current emotional state—such as fear, surprise, anger, joy, boredom." As such, paralinguistic information can often be conveyed even in the absence of a linguistic signal; for example, anger can be detected even when listening to low-pass filtered speech or a language unknown to the listener (Ladd, 2008, chap. 1; but see Chen, Gussenhoven, & Rietveld, 2004, on language-specific aspects of such interpretations). Arvaniti (2007) argues that a possible diagnostic criterion for paralinguistic phenomena is the gradience of both the acoustic parameters used to express them and of what they signify; for example, greater pitch range expansion indicating a greater degree of surprise. This definition is very close to what Bolinger (1961) has referred to as *gradience* (for a discussion, see Ladd, 2014, chap. 4).

One reason for the conflation between prosody and paralinguistics is that the acoustic parameters used to encode linguistic prosodic distinctions are also used to convey paralinguistic information: for example, pitch is the main exponent of intonation but is also used paralinguistically to express excitement, boredom, and anger (see also section 5). In addition, however, the conflation of paralinguistics and prosody is related to the ubiquitous confounding of the linguistic phenomena that are part of prosody with their phonetic exponents; as an example, it is often the case that the term *intonation*, one of the components of prosody, is used as a synonym for *fundamental frequency* (F0), the main phonetic exponent of intonation. In order to avoid this confusion, here the linguistic components of prosody will be kept distinct from the phonetic parameters used for their realization.

A final issue to consider is that many of the phonetic exponents of prosody are also used to encode lexical contrast. Thus, F0 is the prime exponent of both intonation and lexical meaning in languages with tone, such as Cantonese, Thai, or Igbo. This in itself is not an issue, but it should be borne in mind that speakers unfamiliar with a given language are prone to interpreting the use of prosodic parameters according to how they are organized and used in their own linguistic system. This is amply demonstrated in studies of L2 prosody, and the development of creoles (Gooden, Drayton, & Beckman, 2009; Ortega-Llebaria, Hong, & Fan, 2013; Ortega-Llebaria, Nemogá, & Presson, 2017; Qin, Chien, & Tremblay, 2017; Skoruppa, Cristià, Peperkamp, & Seidl, 2011; Tremblay, Broersma, & Coughlin, 2018). For instance, a native speaker of English may interpret as stress the high falling pitch of a lexically accented syllable in Japanese or the pitch rise on Korean phrase-initial syllables because in English stressed syllables are often associated with high or rising pitch (Beckman, 1986; de Jong, 1994). In Japanese, however, pitch accent does not cue prominence (Beckman, 1986), while in Korean the pitch rise is a phrasal, not a stress-related, phenomenon (Jun, 2005A). Taking into consideration the possibility of such cross-linguistic differences is essential when studying prosody.

2. Stress

Stress is a phenomenon that straddles the divide between lexical and postlexical levels: while word stress is a lexical property, stress applies to entire utterances as well (and is sometimes referred to as *sentence stre*ss; for a discussion, see Ladd, 2008, chap. 6). Here, stress is included both because it operates at the phrasal level, and because many of its phonetic exponents are traditionally seen as part of prosody. Stress is not a phonetic property as such; rather, the term refers to the fact that in many languages one or more syllables in a word stand out relative to the rest, with the differences leading to alternations in prominence at the phrasal level as well. For instance, native speakers of English are likely to agree that in *subject* (n.) the first syllable is more prominent than the last, that is, stressed, while the reverse is the case with *subject* (v.). Such differences in relative salience can be phonetically achieved in a number of ways that vary substantially across languages that have stress.

The primary function of stress is *culminative*, that is, stress makes syllables stand out relative to others, a function that has repercussions for rhythm as detailed in section 3. In addition, in languages in which the location of stress can vary, stress may also have *contrastive function*, that is, lead to a change in meaning, as in *subject* (n.) versus *subject* (v.). In some languages, such as Spanish and Greek, the functional load of stress is significant, while in others, such as English, it is limited to a small set of lexical items. Finally, stress has *delimitative function* in languages in which its position is fixed, as in Hungarian and Finnish in which stress always falls on the first syllable of a word. Something similar applies in English, in which 85% of content words start with a stressed syllable (Cutler & Carter, 1987). Statistical probabilities of this sort aid speech segmentation, processing, and acquisition (among many, Cutler, 2015; Skoruppa et al., 2011).

The Phonetics of Prosody

The fact that from a linguistic perspective stress expresses relations of relative salience (e.g., Hayes, 1995; Ladd, 2008, chap. 6; Liberman & Prince, 1977) has led to assumptions that stressed syllables must necessarily be acoustically prominent as well. Phonetically, however, stress is not a uniform phenomenon as the view of stress as acoustic prominence implies. This is evident if one considers the findings on the connection between stress and duration. On the one hand, many studies show that stressed vowels are longer than unstressed ones (e.g., Beckman, 1986, on English; Sluijter & van Heuven, 1996, on Dutch; Arvaniti, 2000, on Greek; Ortega-Llebaria & Prieto, 2011, on Catalan and Spanish; Farnetani & Kori, 1990, and D'Imperio & Rosenthal, 1999, on Italian; Garellek & White, 2015, on Tongan; Yakup & Sereno, 2016, on Uyghur). On the other hand, stressed vowel duration is also affected by a number of additional parameters, such as the position of the stressed syllable in the word (D'Imperio & Rosenthal, 1999, on Italian), the presence of pitch accent and focus (e.g., Botinis, 1989, on Greek; Sluijter & van Heuven, 1996, on Dutch), the interaction of stress, accent, and boundary lengthening (Katsika, 2016, on Greek), and the level of stress involved (e.g., Farnetani & Kori, 1990, Arvaniti, 1992, 1994, and Garellek & White, 2015, found no evidence for durational effects of secondary stress in Italian, Greek, and Tongan, respectively). Finally, there are languages like Welsh in which stressed syllables are shorter than unstressed ones (Williams, 1985). In short, although durational differences associated with stress are present in many languages, stress alone cannot explain all durational distinctions.

In addition to duration, stress is associated with greater amplitude, a view that harks back to Stetson (1951) and the connection between stress and chest pulses. This view is not strongly supported by studies measuring average intensity, in that consistent differences are found in some languages (e.g., Garellek & White, 2015, on Tongan) but not others (e.g., Arvaniti, 2000, on Greek). However, when intensity differences are combined with duration, they often lead to consistently greater *amplitude integral* (Beckman, 1986, on English; Arvaniti, 2000, on Greek; Ortega-Llebaria & Prieto, 2007, on Spanish). Amplitude integral combines the average intensity of a signal with its duration to give a measure of loudness that integrates the effect of duration on loudness (Beckman, 1986; Lieberman, 1960). This measurement is based on the fact that a longer sound will sound louder than a shorter sound with the same average intensity (Moore, 2012).

An alternative measure of loudness was proposed by Sluijter and van Heuven (1996) in their study of Dutch stress: they found that stressed vowels have greater spectral balance, that is, they show a smaller reduction in the amplitude of higher frequencies or less spectral tilt. Sluijter and van Heuven (1996) associated this difference with greater vocal effort. Their findings about Dutch stress were replicated for some languages (e.g., Polish, Macedonian, and Bulgarian; Crosswhite, 2003), but not consistently so: Campbell and Beckman (1997) found spectral tilt differences only between *accented* and unstressed vowels in American English, while Garellek and White (2015) found no spectral tilt effect in Tongan.

Stress-related differences also pertain to vowel quality: in English, for example, the vowel of the second syllable of *subject* (n.) is reduced to a schwa, [ˈsʌbdʒəkt], while the reverse obtains in *subject* (v.), [səbˈdʒɛkt]. Vowel quality differences are essential in determining stress in English, as indicated by vowel alternations like those found in *photograph* [ˈfəʊtəgrɑːf] versus *photography* [fəˈtɒgrəfi] (e.g., Beckman & Edwards, 1994; see Cutler, 2015, for a review). Similarly to English, the distinction between stressed and unstressed vowels is phonologized in Italian, albeit to a much smaller extent: Italian has a seven vowel system, [i e ɛ a ɔ o u], but the distinction between open-mid and close-mid vowels is neutralized in unstressed position (Rogers & d'Arcangeli, 2004). In other languages, however, changes in quality, though evident, are not substantial and not always consistent across speakers and vowels (Sluijter & van Heuven, 1996, on Dutch; Fourakis, Botinis, & Katsaiti, 1999, on Greek; Ortega-Llebaria & Prieto, 2011, on Catalan and Spanish; Adamou & Arvaniti, 2014, on Romani). On the other hand, Garellek and White (2015, p. 23), who also found no significant changes in vowel quality between stressed and unstressed vowels in Tongan, report differences in *voicing* quality, which indicate that stressed vowels are clearer (i.e., less noisy or breathy) than unstressed vowels.

Although the investigation of stress has often focused on vowels, there is evidence that stress affects consonants as well. Suomi and Ylitalo (2004) report that stress leads to longer consonant durations in Finnish. In German, VOT for voiceless stops is longer in stressed syllables (Haag, 1979). In English, lenited forms of some consonants appear in unstressed syllables: simplifying somewhat, in American English, /t/ and /d/ are flapped intervocalically when they are onsets of unstressed syllables, as in *city* > [ˈsɪɾi], while in British English, intervocalic /t/ is realized as a glottal stop in the same context; for example, *city* > [ˈsɪʔi].

A way to unify these observations from a number of different languages is offered by de Jong's analysis of stress as *localized hyperarticulation* (de Jong, 1995). De Jong borrows the term *hyperarticulation* from Lindblom's *H&H theory* (Lindblom, 1990), according to which variation in speech can be accounted for by positing a continuum from hypo- to hyperarticulation. The ends of the continuum reflect two competing forces on articulation, economy of effort (which leads to hypoarticulation), and the need to be understood (which leads to hyperarticulation). De Jong (1995, p. 491) posits that "stress involves a localized shift toward hyperarticulate speech." Although de Jong's data come from English, hyperarticulation can unify the results of cross linguistic studies like those discussed in this section, in that in all languages stressed syllables are hyperarticulated in some way. Hyperarticulation may be manifested as increases in duration, amplitude, or both, changes in phonation that may lead to changes in spectral characteristics, and changes in vowel quality. Viewing stress as localized hyperarticulation is also consistent with the results of articulatory studies (among many, Beckman, Edwards, & Fletcher, 1992; Cho & Keating, 2009; Harrington, Fletcher, & Roberts, 1995), and offers an explanation for various types of phonologized vowel reduction in unstressed syllables, as in English and Italian.

The cross-linguistic variation in the realization of stress is illustrated in Figure 1, which shows the English word *banana* as pronounced by a speaker of American English, namely, [bəˈnænə], and Figure 2, which show its Greek cognate [bɐˈnɐnɐ]. The differences in duration and spectral changes are significant between the two renditions and extend to the function words as well: *a* in English is produced as a schwa, while [mɲɐ] 'a/one' in Greek shows no such reduction. Nevertheless, to native speakers of each language, the middle syllable of *banana* stands out and is considered stressed. To speakers of English, the stressed syllable of the Greek rendition may not sound very prominent, but it is so to native speakers of Greek (cf. Arvaniti & Rathcke, 2015; Protopapas, Panagaki, Andrikopoulou,Gutiérrez Palma, & Arvaniti, 2016, among many). It is equally important to recognize that to speakers of Greek the change of vowel quality in the American English version does not make the middle syllable more prominent, because they do not associate stress with changes in vowel quality.
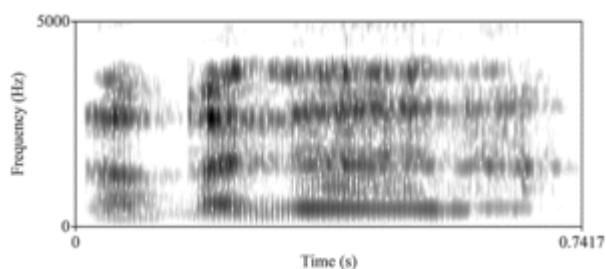

Figure 1. Spectrogram of *a banana*, as pronounced by a speaker of American English: [ə bəˈnænə].
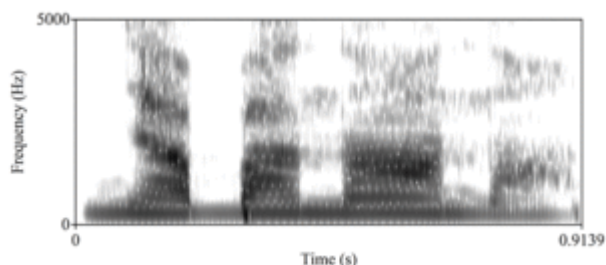
Source: Author.

Figure 2. Spectrogram of [mɲɐ bɐˈnɐnɐ] *'a banana'*, as pronounced by a speaker of Standard Greek.

Source: Author.

A common misconception about the phonetic correlates of stress relates to F0. Specifically, it is often said that stressed syllables have high or rising pitch, and many studies of stress include an investigation of F0 along these lines (e.g., Ortega-Llebaria & Prieto, 2011, on Spanish and Catalan; Gordon & Applebaum, 2010, on Turkish Kabardian; Garellek & White, 2015, on Tongan). These claims can be traced back to Fry (1958). Fry manipulated a number of acoustic parameters in pairs of English words like *subject* (n.) and *subject* (v.) and showed that changes in F0 outweighed those of duration and intensi-

ty in inducing a change in the perceived location of stress. Fry (1958) interpreted this result as evidence that F0 is the most important correlate of stress.

The problem with Fry's experiments is that they confounded stress with intonation. Fry (1958) assumed that he was testing word stress, but his stimuli were one-word utterances; thus, his tests conflated word stress with accentuation (or *sentence stress*), which is expressed primarily by means of F0 (see section 5). Specifically, in a language like English, certain pitch movements, known as *pitch accents*, are expected to co-occur with stressed syllables: when listeners hear a word like *subject* accented on the first syllable, they assume the accent is there because that syllable is stressed, even if its vowel quality, duration, or intensity are not ideal. In the words of Francis Nolan (cited in Ladd, 2008, p. 54), pitch is *prominence cueing* not *prominence lending.* In short, the relationship between stress and F0 is an indirect one: stressed syllables are docking sites for pitch movements, but whether these pitch movements will occur at all and of what type they will be is not determined by stress but by intonation (see also section 5, and Gordon, 2014, for a review).

The following examples further illustrate this point. Figures 3 and 4 show the same phrase, uttered in two different ways: in Figure 3, both the stressed syllable of *Isabel* and that of *Dunham* show high pitch, while in Figure 4 the first (and stressed) syllable of *Isabel* has low pitch instead. The difference in pitch between the two utterances is one of intonation. The utterance in Figure 3 is a typical response to a run-of-the-mill question, such as *who was on the phone?* The utterance in Figure 4 queries the interlocutor's contribution to the common ground; it would be used, for instance, if the speaker had just been told that clumsy Isabel Dunham, and not her gifted sister Mary, won gold in gymnastics; it could be followed by *are you sure you don't mean MARY Dunham?* As these two examples illustrate, there is no direct connection between stress and high pitch, and syllables remain stressed even when they have low pitch (as their duration, amplitude, and spectral characteristics in Figures 3 and 4 indicate). This last point is further illustrated in Figures 5 and 6 with words that form a minimal pair based on the location of stress: in both Figure 5, *MARY's the óbject of inquiry*, and Figure 6, *MARY will objéct to the inquiry*, F0 is flat on the word *object*; however, differences in the duration and quality of the vowels are evident and sufficient to indicate the difference in stress between the noun in Figure 5 and the verb in Figure 6 (cf. Huss, 1978).
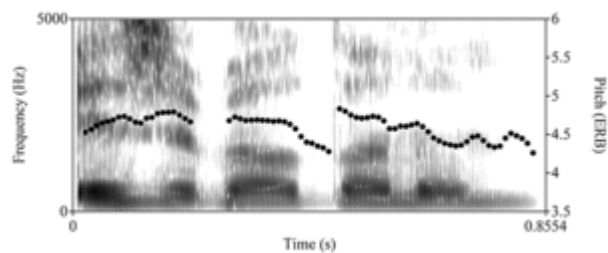
Figure 3. Spectrogram and F0 contour of *Isabel Dunham*, uttered as a statement (e.g., as an answer to the question *who was on the phone?*).
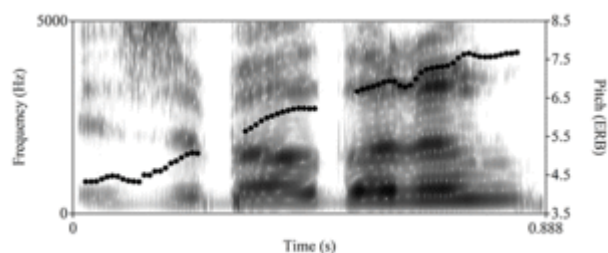
Source: Author.

Figure 4. Spectrogram and F0 contour of *Isabel Dunham*, uttered as a question (e.g., as a response to a piece of news about Isabel Dunham that the speaker considers to more likely apply to her sister, Mary).

Source: Author.

In conclusion, there is no direct connection between stress and objective acoustic measures of prominence, in that different languages may indicate stress in a number of ways (though F0 is unlikely to be a direct correlate of stress). These cross-linguistic findings have implications for phonetic research: they suggest that it is not possible to determine whether a language has stress by simply measuring the duration, intensity, or spectral characteristics of segments. This is both because parameters that encode stress differ across languages, and also because not all languages have stress, so acoustic prominence may be the outcome of phrasal processes instead. For instance, syllables initial to the accentual phrase in Korean are articulatorily strengthened and show resistance to coarticulation (Cho & Keating, 2009). Neither of these phenomena is related to stress, however, as Korean does not have stress (Jun, 2005A). Avoiding the temptation to interpret such acoustic effects as exponents of stress is important. Further, in order to understand the contribution of F0 and disentangle intonation from stress, it is essential to consider data from utterances with different tunes, rather than base conclusions on declarative utterances only (in which the connection between stress and high F0 is most likely to be manifested). The exploration of whether a system has stress should start with phonological observations, taking into consideration morphophonological alternations that may lead to alternations in vowel quality, processes like blending and hypocoristic formation, and the potential role of stress in acquisition and processing (for criteria, see Gordon, 2011). The

answer may be that, like Ambonese Malay, a language does not have stress or other phenomena with primarily culminative function (Maskikit-Essed & Gussenhoven, 2016).
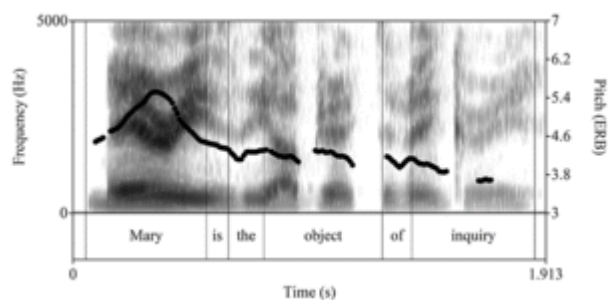

Figure 5. Spectrogram and F0 contour of *MARY is the object of inquiry* (with focus on *Mary*).
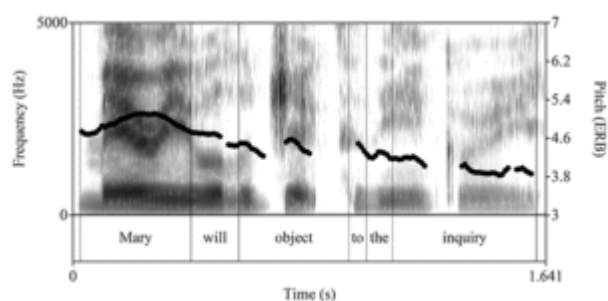
Source: Author.

Figure 6. Spectrogram and F0 contour of *MARY will object to the inquiry* (with focus on *Mary*).

Source: Author.

3. Rhythm

There is no good and generally accepted definition of speech rhythm. A definition based on the psychology of rhythm is adopted here, namely that rhythm is an abstraction that relies on perceiving constituents in speech as groups of a similar and repetitive pattern. This definition, however, is not generally accepted. In much of the phonetic literature rhythm has been confounded with timing, that is, with duration patterns, and specifically with the idea that languages fall into distinct categories based on keeping some constituent constant in duration. This idea can be traced back to impressionistic work on English from the early 20th century, which eventually gave rise to the notion of *isochrony* and rhythm classes: languages are said to be stress-, syllable-, or mora-timed depending on whether the unit that is supposed to show stable (i.e., isochronous) duration is the stress foot, the syllable, or the mora, respectively. Experimental research starting with Classe (1939) and continuing to at least the 1980s has failed time and again to find evidence of isochrony in production, leading some authors to advocate that rhythm classes are a perceptual illusion of speakers of various Germanic languages (Roach, 1982; for reviews see Arvaniti, 2009, 2012A, IN PRESS-A).

Along similar lines to Roach (1982), Dauer (1983) argued that syllable-timing (and by extension mora-timing) is not a plausible basis for rhythm and proposed instead that languages form a rhythm continuum *from least to most stressed-based*. According to Dauer (1983), a language's placement on the continuum is determined by the prominence of its stress exponents. Although Dauer's equating of acoustic prominence with stress is problematic, as discussed in section 2, her conceptualization of rhythm is closer to the understanding of rhythm in psychology and musicology.

3.1. Rhythm Classes and Rhythm Metrics

Dauer's main point of a stress-based rhythm continuum was largely ignored in subsequent research, but a small subset of her criteria for determining stress salience formed the basis of rhythm quantification in Ramus, Nespor, and Mehler (1999). The aim of Ramus et al. (1999) was to quantify timing differences related to rhythm class, a concept they considered uncontroversial in linguistics. They argued that so-called stress-timed languages like English have more varied syllable structures and greater vowel reduction than syllable- and mora-timed languages, and that this is reflected in differences in the duration of vocalic and consonantal stretches of speech. In addition, Ramus et al. argued that these differences can be used during acquisition to resolve the bootstrapping problem by allowing infants to pay selective attention to one of the timing units (for arguments against this view see later in this section). Ramus et al. (1999) tested a number of measures and concluded that %V, the percentage of vocalic intervals, and ΔC, the standard deviation of consonantal intervals, best capture the differences they argued exist between rhythm classes.

Since Ramus et al. (1999), a number of additional metrics have been proposed, for example the pairwise variability indices or PVIs (Grabe & Low, 2002), and Varcos, standard deviations divided by the mean (Dellwo, 2006). Variations on these metrics and several additional measures have also been proposed. Frota and Vigário (2001) proposed the use of standard deviations of normalized percentages. Wagner and Dellwo (2004) proposed a measure similar to the PVIs but based on z-transformed syllable durations. Nolan and Asu (2009) used PVIs on syllable and foot durations. Many of these metrics have been widely applied in fields ranging from forensic work and L2 phonetics to the study of acquisition and atypical speech (e.g., Hannon, Lévêque, Nave, & Trehub, 2016, on acquisition; White & Mattys, 2007, on L2; Liss et al., 2009, on atypical speech; Harris, Gries, & Miglio, 2014, on forensic phonetics).

Despite their popularity, metrics are fraught with problems, both theoretical and methodological. First, metrics are implausible as measures of rhythm during acquisition: they require that infants retain in short-term memory chunks of speech in order to compute global statistical trends (which cannot be computed on the fly) with the aim of recognizing the rhythm class of the language they are learning. Second, while infants learning languages categorized as stress-timed can focus on stressed syllables, it is unclear what infants learning syllable- or mora-timed languages could focus on as, by definition, all syllables (or moras) should be equally plausible word onsets. Empirical evidence does not

support the inevitable conclusion stemming from this view, namely that infants learning syllable- or mora-timed languages face greater challenges (among many, Tzakosta, 2004, on Greek; Pons & Bosch, 2010, on French and Spanish). Third, metrics are circular: as there is no independent evidence for rhythm classes, metrics are used both to determine class affiliation and to support the notion that rhythm classes exist (for a detailed critique, see Arvaniti, 2009).

In addition, metrics are problematic as measures, even if circularity and implausibility are ignored. First, they are volatile and affected by a large number of factors. Renwick (2013), and Horton and Arvaniti (2013) have independently shown that %V, the measure often said to be the most stable and accurate predictor or rhythm class, strongly correlates with the number of closed syllables in the speech sample, independently of the language tested. Additionally, metric scores show significant interspeaker variation, and are also affected by the overall segmental composition of the speech sample used (Wiget et al., 2010; Arvaniti, 2012A), and the method of eliciting it (Arvaniti, 2012A). The effect size of these factors is larger than that of language, indicating larger variability within than across languages (Arvaniti, 2012A). Second, the exact effects of such factors on metrics are unpredictable; this is reflected in the fact that metrics said to capture the same phenomenon (e.g., consonantal variability), do not correlate with one another (Loukina, Kochanski, Rosner, Keane, & Shih, 2011; Arvaniti, 2012A; Horton & Arvaniti, 2013). This is because metrics are strongly influenced by local effects, such as phrase final lengthening or the irregularities present in atypical speech (e.g., Arvaniti, 2009; Lowit, 2014). This effectively means that two speech samples can yield similar scores in some metric but for entirely different reasons (Arvaniti, 2009). Consequently, metric scores are uninterpretable on their own; they can be interpreted only with close scrutiny of timing relationships between segments in a given speech sample, but such scrutiny is not aided by metric scores (Arvaniti, 2009). Finally, metrics can be problematic from a statistical perspective because they are often used in bundles, with researchers selecting results that turn out to be statistically significant according to some factor relevant to the study, a practice that increases type I error (among others, Li & Post, 2014; Kaminskaïa, Tennant, & Russell, 2016). For all these reasons, metrics provide no strong evidence in favor of rhythm classes and are not reliable or informative measures of timing in general, as has recently been argued, for example, by Post & Payne (2018).

3.2. Rhythm Classes and Perception

Perception experiments have not provided greater support for rhythm classes than production research. Many experiments are based on discrimination among either infants or adults (among many, Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Jusczyk, & Johnson, 2000; Nazzi & Ramus, 2003). Another set of studies is based on processing (in the form of spotting or monitoring), which, it is argued, relies on syllables, morae, or feet depending on the rhythmic class of the listeners' native language (Cutler, Mehler, Norris, & Seguí,1986, 1992; Otake, Hatano, Cutler, & Mehler, 1993; Cutler & Otake, 1994; Murty, Otake, & Cutler, 2007).

The argument behind the discrimination experiments is that if two languages can be discriminated from each other then they must belong to different rhythm classes. However, the premise behind these experiments is questionable, as several studies have shown that varieties of the same language can be discriminated from each other both by infants and adults (Nazzi, Jusczyk, & Johnson, 2000, and White, Mattys, & Wiget, 2012, on infants and adults, respectively). In addition, some discrimination experiments have led to counterintuitive results, such as Moon-Hwan (2004), who concluded that Korean is mora-timed because it was discriminated from Italian and English but not Japanese; there is no evidence, however, that the mora plays any role in Korean timing, phonology, or processing. Finally, some languages can be discriminated from both stress- and syllable-timed prototypes, a result that also sits uneasily with the premise that languages belong to distinct rhythm classes (see e.g., Ramus, Dupoux, & Mehler, 2003, who found that Polish can be discriminated from both English and Spanish). Such results indicate that putative rhythm class is not a good explanation for discrimination.

Another problem with the discrimination experiments is that in order to force listeners to focus on timing (seen as the only exponent of rhythm in this work), studies have usually relied on *flat sasasa*. This is a type of modified speech in which F0 is "flat" (slightly falling throughout an utterance), while all vocalic intervals are replaced by [a] and all consonantal intervals by [s]; for example, *it's raining again* would be rendered as *asasasasas* with the intervals corresponding to [ɪ] [tsɹ] [eɪ] [n] [ɪ] [ŋ] [ə] [g] [ɛ] [n]. There is evidence, however, that flat *sasasa* is not ecologically valid; for example, in Arvaniti (2012B) utterances rendered into flat *sasasa* yielded different responses from low-pass filtered versions of the same utterances. Since both modifications retain timing characteristics but low-pass filtering is closer to actual speech, the differences in responses indicate that the percept resulting from *sasasa* is not close to what listeners obtain from speech.

To explore the issues with *sasasa* and the discrimination paradigm, Arvaniti and Rodriquez (2013) ran a series of AAX experiments with English as the standard (AA) and Danish, Spanish, Greek, Korean, and Polish as comparisons (X). Arvaniti and Rodriquez (2013) used two *sasasa* versions, flat *sasasa*, and *sasasa* that retained the original F0 of the utterances, and additionally manipulated the speaking rate of the stimuli so as to retain or eliminate differences in speaking rate between standards and comparison. The results showed that both speaking rate and F0 play a substantial role in driving discrimination, with effects depending on the language pair but not on putative rhythm class. When F0 and speaking rate differences were eliminated, discrimination was much weaker independently of the putative rhythm class of the languages involved. Overall, the results indicate that discrimination experiments are difficult for participants, who end up latching onto any differences they can find in the signal in order to complete the task. Critically, the results of Arvaniti and Rodriquez (2013) confirm that *sasasa* is not ecologically valid, in that it does not reflect the perception of speech rhythm in natural stimuli: if that were the case, changes in F0 or speaking rate would have had no effect on responses. The fact that they do indicates that listeners do not process the timing of segments independently of the other prosodic parameters present in the speech signal but, rather, they integrate prosodic information. This conclusion is supported by experiments on distal prosody (Dil-

ley & McAuley, 2008; Dilley, Mattys, & Vinke, 2010 *inter alia*). 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 related.

Similar arguments apply to studies in processing. Such studies rely mostly on variations of the spotting paradigm, whereby listeners are asked to spot (or monitor for) fragments (such as a syllable) in a continuous speech stream. Many of these experiments do show that a particular phonological constituent is salient in each language and useful to native listeners during processing (e.g., Cutler et al., 1986, on the role of the syllable in English and French; Otake et al., 1993, on the mora in Japanese; Murty, Otake, & Cutler, 2007, on the mora in Tamil). More generally, studies have also confirmed the significance of stressed syllables for the acquisition and processing of so-called stress-timed languages like English and German (e.g., Schmidt-Kassow & Kotz, 2008; Rothermich, Schmidt-Kassow, Schwartze, & Kotz, 2010; Skoruppa et al., 2011). However, stress has also been shown to be crucial for processing in so-called syllable-timed languages, including Spanish and Greek (e.g., Soto-Faraco, Sebastián Gallés, & Cutler, 2001; Magne et al., 2007; Skoruppa et al., 2009; Arvaniti & Rathcke, 2015; Protopapas et al., 2016). In other words, the expected compartmentalization of stress versus syllable is not supported by experimental evidence. This should not be surprising. As Mattys and Melhorn (2005) point out, this body of literature often refers to "stress" when discussing speech processing of English; however, recognizing stress during processing requires that listeners can recognize syllables as well. In short, these findings neither prove membership to a rhythm class nor preclude the usefulness of other prosodic units during speech planning and processing.

The traditional idea of rhythm classes is not problematic only because it is unsupported by studies in production or perception. It is important to recognize that the conceptualization of rhythm as timing is problematic on cognitive grounds as well (see Arvaniti, 2009, 2012A; see Arvaniti, IN PRESS-A, for detailed arguments). A major problem is the implausibility of syllable-and mora-timing as rhythm mechanisms. Both represent a *cadence*, the simplest form of rhythm "produced by the simple repetition of the same stimulus at a constant frequency" (Fraisse, 1982, p. 151). For a cadence to be perceived as such, however, stimuli must be sufficiently separated in time to be experienced as distinct, that is, for fusion to be avoided. According to Fraisse (1982) this temporal spacing is at least 200 ms.[1] However, the typical speaking rate of many languages classified as syllable- or mora-timed is much faster than that, with reported rates ranging from 128 to 143 ms per syllable (Dauer, 1983, on Spanish, Greek and Italian; Pellegrino, Coupé, & Marsico, 2011, on Italian, French, Spanish, and Japanese). At these rates, it would be extremely difficult if not impossible for each syllable or mora to be reliably perceived as a distinct beat (see, e.g., London, 2012, chap. 2).

In addition, listeners exhibit *subjective rhythmization*, that is, they tend to impose a rhythmic pattern on cadences, typically grouping stimuli into trochees or iambs (Bolton, 1894; Woodrow, 1951; Fraisse, 1963, 1982). This perceptual tendency is difficult to reconcile with the idea that all syllables or moras are equally prominent: even if they were all

acoustically equal (and all evidence suggests they are not), they would not be perceived as such. This also begs the question: how would a child acquire a syllable- or mora-timed language if their perceptual system predisposes them not to perceive the language as such?

Further, research on rhythm perception shows that listeners can impose or maintain a rhythm without constant overt clues, particularly once a pattern is established (London, 2012, chap. 1). In part, this is so because of *dynamic attending*, the fact that listeners pay selective attention to auditory events, focusing on those periodically occurring (e.g., Jones, 1981). Dynamic attending rests on the idea that humans cannot attend to all events (James, 1890, cited in London, 2012, chap. 1). Again, syllable- and mora-timing cannot be reconciled with this tendency, as they would require that speakers of so-called syllable- and mora-timed languages make no selection, and are capable of attending to all events in a rapidly paced series (cf. the issue with acquisition previously discussed). This idea is implausible, and unsupported by native speaker intuitions (e.g., Vaissière, 1991, on French), processing (e.g., Jeon & Arvaniti, 2017, on Korean), speech production (e.g., Chung & Arvaniti, 2013, on Korean; Arvaniti & Rathcke, 2015, on Greek), and acquisition (e.g., Tzakosta, 2004, on Greek; Pons & Bosch, 2010, on French and Spanish). In short, the idea of syllable- and mora-timing is psychologically implausible, while research on so-called syllable- and mora-timed languages shows that speakers of such languages focus either on phrasal boundaries (French, Korean) or stresses (Greek, Spanish), and rely on rhythm groups larger than the syllable or the mora.

3.3. Alternative Views on Rhythm

If rhythm is not based exclusively on the regular timing of some unit, then how is it created? As mentioned, a possibility is to see rhythm as a perceptual phenomenon, specifically the perception of speech as a series of groups of a similar and repetitive pattern (Arvaniti, 2009). This definition is not new; it is based on the psychological understanding of rhythm (e.g., Woodrow, 1951, Fraisse, 1963, 1982; London, 2012). It is also closer to the conception of rhythm used in phonology (e.g., Hayes, 1995), in which rhythm is seen as relying on the relative salience of constituents at several levels of the prosodic hierarchy (see section 4). If such a definition is adopted, then research on rhythm should focus on what phenomena could lead to listeners perceiving speech as consisting of groups of similar and repetitive pattern. Following Dauer (1983), a plausible organizational principle would be stress and the creation of stress feet, in languages that have stress. This, however, leaves open the question of how rhythm is created in languages that do not. Some suggestions are offered here.

First, the regularity that leads to the perception of rhythm may be related to alternations in duration, but this is neither necessary nor sufficient (since, as mentioned, listeners do not process timing as a dimension of the speech signal that is distinct from other prosodic parameters). In short, duration is not the only exponent of rhythm and should not be seen as such. Thus, although segmental timing is an essential component of a language's pho-

netics, it deserves to be studied independently of the connection to rhythm (see, e.g., Turk & Shattuck-Hufnagel, 2000, 2014, and references therein).

A phonetic parameter beyond duration that may contribute to rhythm is amplitude. Tilsen and Arvaniti (2013) used *empirical mode decomposition* (EMD; Huang et al., 1998) to extract regularities from the amplitude envelope of filtered speech waveforms. This envelope displays quasi-periodic fluctuations in energy that tend to arise from (but do not completely coincide with) the alternation of vowels and consonants. Thus, for Tilsen and Arvaniti (2013) "rhythm is conceptualized as periodicity in the envelope, and greater stability of that periodicity corresponds to greater rhythmicity" (Tilsen & Arvaniti, 2013, p. 629). Simplifying considerably, EMD extracts a number of basis functions from the signal, termed *intrinsic mode functions* (IMFs). Each IMF captures oscillations on a different time-scale and can be analyzed using a Hilbert transform to obtain an instantaneous phase; the instantaneous frequency (ω) of an IMF is the time derivative of phase. Tilsen and Arvaniti (2013) argue that in speech the instantaneous frequencies of the first two IMFs correspond to periodicities at the syllable-level (ω_1) and foot-level (ω_2), respectively. Their results show that the average ω_2 in their corpus is 2.5 Hz, a frequency that corresponds—assuming the interpretation of Tilsen and Arvaniti (2013) is correct—to recurrent beats every 400 ms. This is in line with the average foot duration reported in Dauer (1983). Further, the variance of ω_2 is comparable across the languages they examined, English and German, which are classed as stress-timed, Italian and Spanish, which are classed as syllable-timed, and Greek and Korean, which remain unclassified. The fact that ω_2 variance is similar across these languages suggests similarities in rhythmicity in languages that are traditionally considered to belong to distinct rhythm classes. In particular, it would suggest the presence of a louder element every approximately half a second and comparable levels of fluctuation from this standard in all the languages examined. The fact that this pattern applies even in Korean, a language without stress, indicates that stress is not required for grouping purposes.

In addition, research on Korean indicates that F0 may also play a part in creating rhythmic groupings. Jeon and Arvaniti (2017) found that the regular F0 pattern spanning the accentual phrase in Korean (a prosodic constituent of typically 3–4 syllables long) is more important during processing than having accentual phrases of equal duration (in number of syllables). This result agrees with previous literature on the processing of Korean (see Jeon & Arvaniti, 2017, and references therein). A way to interpret this result is to recognize that in Korean, and possibly languages typologically similar to it such as French, rhythm may rely on the presence of a repetitive F0 pattern over short phrases, rather than on segmental timing. It is possible that changes associated with this F0 pattern give rise to the amplitude alternations reported by Tilsen and Arvaniti (2013) for Korean.

Although much more research is needed on these alternatives to the traditional view of rhythm as timing, it is important to reiterate that no acoustic parameter can be solely responsible for rhythm, due to perceptual integration, as previously mentioned. The importance of perceptual integration has been further demonstrated by a number of perceptual studies: Dilley and McAuley (2008), Kohler (2009), and Dilley et al. (2010), among others,
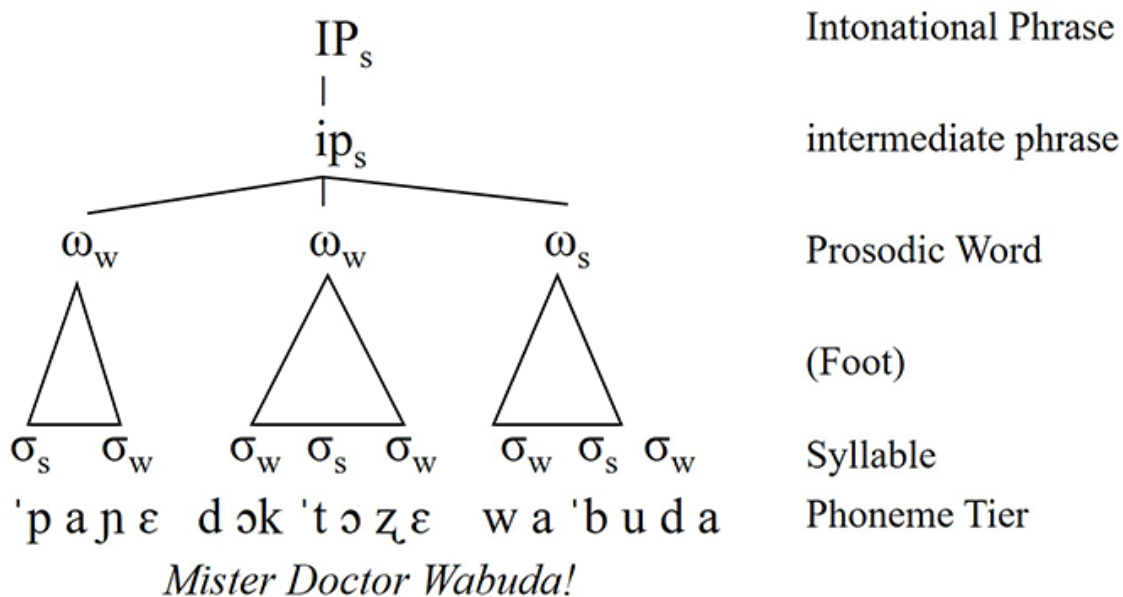
have shown that the perception of grouping and relative prominence is influenced by changes in F0 patterns. In conclusion, moving away from the traditional rhythm class typology and considering how components of prosody may contribute to the creation of rhythm in languages with typologically distinct prosodic systems may yield the insights that have not been forthcoming in the study of speech rhythm as timing, and the adherence to the rhythm class typology.

4. Phrasing

Phrasing refers to the fact that in speech words are chunked together rather than being produced as distinct and independent elements in a string. Phrasing is critical for organizing and planning speech production, and influences perception as well (among many, Krivocapić & Byrd, 2012; Katsika, Shattuck-Hufnagel, Mooshammer, Tiede, & Goldstein, 2014; see Turk & Shattuck-Hufnagel, 2014, for a review). Phrasing is also necessary to understand intonation (see section 5). Phrasing has been investigated from both a phonological and a phonetic perspective, though the two do not always agree. In order to understand the phonetic results, it is essential to understand the essential tenets of phonological accounts of phrasing.

According to phonological accounts of phrasing, words are grouped into a hierarchical prosodic structure that does not allow recursion (but see Ladd, 1988, on phonetic evidence for limited prosodic recursion). A model implicitly adopted in much work on prosody, particularly intonation, is that proposed by Pierrehumbert and Beckman (1988; see also D'Imperio, Elordieta, Frota, Prieto, & Vigário,2005, for a review). This model does not assume a direct mapping from syntax (as do the models of Selkirk, 1984, and Nespor & Vogel, 1986). Rather, phrasing is empirically determined, as it is affected by speaking rate, speech clarity, and the length of constituents. For instance, *my girlfriend's mother's sister is a heavy smoker* is more likely to be produced with a phrasal break after *sister* than is *she's a smoker*. Similarly, clear speech is likely to result in shorter phrases than otherwise (Smiljanic & Bradlow, 2008). Further, Pierrehumbert and Beckman (1988) posit different levels depending on the language. For instance, they argue that the English prosodic hierarchy has three main levels, the prosodic word (ω), intermediate phrase (ip), and intonational phrase (IP), an analysis based on Beckman and Pierrehumbert (1986). In contrast, their analysis of Japanese prosody requires an additional level, that of the *accentual phrase* (AP). The AP features prominently in the prosodic analysis of French, and Korean (Fougeron & Jun, 2002, and Jun, 2005A, respectively). An illustration of prosodic structure after Pierrehumbert and Beckman (1988) is given in (1) with a phrase from Polish (based on data from Arvaniti, Żygis, & Jaskuła, 2017).

    (1)

IP~s~ Intonational Phrase

ip~s~ intermediate phrase

ω~w~   ω~w~   ω~s~   Prosodic Word

(Foot)

σ~s~ σ~w~   σ~w~ σ~s~ σ~w~   σ~w~ σ~s~ σ~w~   Syllable

ˈpaɲɛ dɔkˈtɔʐɛ waˈbuda   Phoneme Tier

Mister Doctor Wabuda!

An issue of phonetic interest relates to empirical evidence for prosodic structure. In phonological models, prosodic structure is said to regulate many connected speech phenomena. Following Selkirk (1980) these phenomena can be classed into the following categories:

a. *Domain limit rules*: rules apply at the edge of some prosodic domain, for example, Nespor and Vogel (1986) analyze voiceless stop aspiration in English as a domain limit rule that applies at the left edge of the foot.

b. *Domain span rules*: these apply within a specific prosodic domain; for example, the rule of s-voicing in Italian is said to apply to intervocalic /s/ within the prosodic word domain; similarly, flapping in American English can be analyzed as a domain span rule applying within the foot (Nespor & Vogel, 1986).

c. *Domain juncture* rules: these rules apply at the juncture between two constituents of a specific type, provided the boundary occurs within some higher constituent; for example, Dutch has an optional s-voicing rule that applies if /s/ occurs ω-finally and the next ω begins with a vowel, provided both ωs are part of the same intonational phrase (Gussenhoven & Jacobs, 2017, chap. 12).

Phonetic research on rules like those discussed immediately prior, however, has shown that very often they are not categorical, as phonological models predict, but gradient, coarticulatory phenomena. Vowel deletion due to hiatus across a word boundary in Greek is a case in point. Arvaniti (1991) and Baltazani (2006) have shown that vowel deletion does not apply within the clitic group (as argued by Nespor & Vogel, 1986), or the *small phrase z* (as argued by Condoravdi, 1990) and is not based on vowel sonority (as argued by Malikouti-Drachman & Drachman, 1992). Rather, the reason for the divergence among these studies (and the fact that they do not agree on which vowel is deleted and in what contexts) has to do with the fact that in Greek most instances of vowel hiatus across a

word boundary lead to vowel coalescence, not deletion (Baltazani, 2006; for a detailed review, see Arvaniti, 2007).

Seminal work on the gradient nature of connected speech phenomena phonologically described as categorical was presented by Nolan (1992), who reported EPG data on English coronal assimilation and degemination (Chomsky & Halle, 1968). Phonologically, this rule can be analyzed as a domain juncture rule whereby a coronal stop at the right edge of a prosodic word assimilates to the stop at the onset of the following word provided they are both in the same intermediate phrase. Nolan (1992) compared sequences such as *make calls* and *late calls* (both embedded in longer utterances). In *make calls*, degemination should lead to the sequence being pronounced [meɪkɔːlz]; in *late calls*, complete assimilation followed by degemination should lead to an identical sequence after the initial [l], that is, [leɪkɔːlz]. Nolan (1992) found, however, that sequences like *late calls* rarely show complete deletion of the coronal gesture; this gesture may be undershot (in that it does not result in a complete alveolar closure) and may overlap substantially in time with the velar closure, but it is rarely entirely absent. In other words, this is a pattern of gradient assimilation, resulting in traces of [t] being present in the signal. These traces affect transitions from the preceding vowel and research shows they are recoverable, that is, available to listeners during processing (Gow Jr., 2002). Similar results are reported by Zsiga (1995) on palatalization in American English (e.g., the palatalization of /s/ in *miss you*). Zsiga (1997) also considered vowel harmony and assimilation in Igbo and concluded that while some connected speech processes are categorical others are gradient; she further argued that only the former should be formalized in phonology. In short, whether a particular pattern is absolute or gradient is a matter of empirical investigation.

Although phonetic studies have shown that some connected speech phenomena analyzed as phonological rules are in fact gradient, there are many other ways in which speakers demarcate prosodic structure. Initial boundaries, particularly those higher in the prosodic hierarchy, show *articulatory strengthening* (among many, Fougeron & Keating, 1997, and Byrd, 2000, on American English; Cho & Keating, 2001, and Cho, Son, & Kim, 2016, on Korean; Recasens & Espinosa, 2005, on Catalan; Fougeron, 2001, and Georgeton, Antolík, & Fougeron, 2016, on French). Such strengthening can be manifested as a longer or more robust constriction of the initial consonant, but can also take other forms; for example, Dilley, Shattuck-Hufnagel, and Ostendorf (1996) showed that word-initial vowels in English are produced with glottalization, particularly if they are also initial to the intonational phrase. In addition, prosodic boundaries are associated with durational changes, particularly at the right edge, which is often associated with lengthening (Cambier-Langeveld & Turk, 1999, on English and Dutch; Byrd & Saltzman, 2003, on English; Nakai, Kunnari, Turk, Suomi, & Ylitalo, 2009, on Finnish; Katsika, 2016, and Loutrari, Tselekidou, & Proios, 2018, on Greek; see Turk & Shattuck-Hufnagel, 2014, for a review). In addition, prosodic boundaries may be tonally specified, particularly in languages that do not have stress. This is found with respect to the accentual phrase in French (Fougeron & Jun, 2002), Korean (Jun, 2005A), Japanese (Pierrehumbert & Beckman, 1988), and Ambonese Malay (Maskikit-Essed & Gussenhoven, 2016), to mentioned but a few. The presence of tonal marking does not imply that segmental effects are lacking: Ko-

rean, for instance, is well known for its segmental changes at phrasal boundaries (Jun, 2005A). Finally, listeners rely on these cues about prosodic phrasing during speech processing and utterance disambiguation (Hirschberg & Avesani, 2000, Krivokapić & Byrd, 2012, Jeon & Arvaniti, 2017, Loutrari et al., 2018, *inter alia*).

5. Intonation

Intonation refers to the language-specific and systematic modulations of fundamental frequency (F0) that span entire utterances and have grammatical function(s), such as encoding pragmatic information and marking phrasal boundaries. As noted briefly in section 1, the terms *F0*, *pitch*, and *intonation* are often used interchangeably in the literature, a practice that has led to the confusion of linguistics and paralanguage, on the one hand, and of phonological phenomena with their phonetic exponents, on the other. To avoid this confusion, I discuss each term in some detail in section 5.1..

5.1. Intonation, F0, and Pitch

F0, measured in Hz, is a property of the speech signal directly related to the rate of vibration of the vocal folds. F0 changes throughout an utterance in ways that relate to a number of factors. These include biological factors, such as a speaker's age and gender: children have overall higher pitched voices than adults, and women have higher pitched voices than men (e.g., Daly & Warren, 2002; Warren, 2005; Clopper & Smiljanic, 2011; Graham, 2014; see Titze, 1994, for an overview). These biological differences relate to the size of the larynx and the thickness and length of the vocal folds, but they are also exploited for indexical sociolinguistic purposes, so that people of similar build and biological sex may use different pitch range and have different average pitch (e.g., van Bezooijen, 1995, on Japanese and Dutch; Yuasa, 2008, on Japanese and American English). F0 is also used to index *paralinguistic* information, such as boredom, anger, or excitement (Ladd, 2008, chap. 1; see also section 1).

In addition to socioindexical and paralinguistic functions, F0 serves two main linguistic purposes. First, at the lexical level, it is the prime exponent of lexical tonal contrasts, for languages that have them, such as Cantonese, Japanese, or Igbo; in these languages changes in F0 lead to changes in lexical meaning.[2] Second, at the postlexical (i.e., phrasal) level, F0 is used to mark prosodic boundaries and convey pragmatic meaning and information structure distinctions. It is these specific uses that will be referred to here as *intonation*, as they are part of a language's prosodic system. Intonation is specified at the phrasal level by means of a complex interplay between metrical structure (informally, the representation of patterns of prominence), prosodic phrasing, syntax, and pragmatics; these factors determine where F0 movements will occur and of what type they will be. As discussed in section 2, for instance, some changes in pitch synchronize with stressed syllables. It is important to note that intonation is used in all languages, whether they have lexical tone or not. Disentangling the contribution of lexical tone from that of intonation on F0 contours is not a trivial task, and is a topic on which more re-

search is needed (for the analysis of systems combining lexical tone and intonation, see, among others, Pierrehumbert & Beckman, 1988, and Venditti, 2005, on Japanese; Bruce, 1977, 2005, on Swedish;Peng et al., 2005, on Mandarin;Wong, Chan, & Beckman, 2005, on Cantonese; and Downing & Rialland, 2017, on a number of African tone languages).

F0 gives rise to the percept of pitch. There are several scales for measuring pitch but no strong consensus on which is best for the investigation of tone and intonation. Some studies use Hz (e.g., Rietveld & Gussenhoven, 1985; Arvaniti, Ladd, & Mennen, 1998), a practice that although occasionally frowned upon, is not aberrant in that the relationship between F0 and pitch is almost linear up to approximately 1,000 Hz, a threshold significantly above that of F0 in human speech (Stevens & Volkmann, 1940). The two *pitch* scales used most frequently in intonation studies are ERB (Equivalent Rectangular Bandwidth) and semitones. ERB reflects a semi-logarithmic relation between pitch and F0 in the frequencies used for intonation and has been shown to accurately reflect the relation between F0 and perceived pitch (Glasberg & Moore, 1990; Hermes & van Gestel, 1991). Semitones are a logarithmic transformation of the Hertz scale originally related to Western music. Although semitones are now increasingly used in intonation research, the evidence in favor of semitone use over ERB is sparse (Henton, 1989), and largely refuted (see Daly & Warren, 2001, contra Henton, 1989; see also Stevens & Volkmann, 1940). To the author's knowledge, the only research directly comparing various scales of pitch in intonation is Nolan (2003). Nolan asked 18 speakers to imitate the intonation of utterances produced by one male and one female talker and compared the imitated versions to the original intonation with both sets of pitch contours expressed in semitones, ERB, Hz, Mel, or Bark. He found that the differences between the two versions were smaller for semitones and ERB compared to Hz, Mel, and Bark, with differences in semitones being marginally smaller than ERB. This led Nolan (2003) to conclude that semitones best reflect how intonation is perceived. However, this conclusion rests on the assumption that speakers were accurate in their imitations. This assumption cannot be ascertained based on Nolan (2003). Further, semitones have the dubious advantage of minimizing differences between male and female speakers; this can be convenient for statistical analysis but it may well hide systematic differences related to sex and gender (as in Henton, 1989), which could come to light with other scales and separate by-gender analyses of data (cf. Daly & Warren, 2001). Since sex- and gender-related differences are valid and perceptible, eliminating them from analysis does not seem advisable. The same applies to other types of scaling differences as well; for example, Fujisaki and Hirose (1984) argue that by using a logarithmic scale of pitch they eliminated differences in the scaling of components of their model in different positions in an utterance. However, doing so may, once again, mask differences that are important for phonetic modeling and relevant for perception (e.g., Yuen, 2007).

Independently of the scale used, an issue faced by all researchers relates to what measurements are best for intonation research. F0 presents as a curve (with discontinuities due to voicelessness) and one of the biggest challenges is determining what elements of this curve need to be measured and accounted for. There is no consensus on this issue. Many researchers focus on straightforward measures such as measuring average F0 over

specific stretches of speech that range from a syllable to entire utterances and beyond (see, e.g., many studies on stress, such as Ortega-Llebaria & Prieto, 2011; Gordon & Applebaum, 2010; Garellek & White, 2015). From a linguistic perspective, such measures are not particularly meaningful. In addition, they are unlikely to be representative of perception: listeners (of non-tonal languages at least) tend to perceive pitch movements as level pitch (e.g., Dilley & Brown, 2007; Haung & Johnson, 2010), and to equate this level pitch to a point between the mean and end frequency of a pitch movement (Nábělek, Nábělek, & Hirsch, 1970; 't Hart, Collier, & Cohen, 1990). This means that listeners are likely to perceive rising pitch as high, and falling pitch as low. Given that rising pitch movements often tend to show overshoot (or *peak delay*), that is, to extend beyond the syllable with which they are expected to co-occur, estimates based on averages of within syllable excursions are likely to under-estimate perceived pitch. This problem with F0 averaging may be the reason why research on pitch in relation to gender and sexual orientation has not always yielded results that matched known stereotypes (e.g., Gaudio, 1994; Waksler, 2001; Munson, McDonald, DeBoe, & White, 2006).

Similar comments can be made about measuring *pitch dynamism*. The term *pitch dynamism* refers to the frequency and extent of pitch excursions in a given stretch of speech. No single method of measuring dynamism is available. Gaudio (1994, p. 46), following Eady (1982) measured "(1) the average extent of changes in F0, using the absolute value of every pitch change (i.e., If2 - fl, If3 - f2, etc.); (2) the total number of 'fluctuations,' defined as changes in the pitch track from a positive to a negative slope, or vice versa; (3) the number of 'upward' and 'downward' fluctuations, defined as changes in pitch at least as great as some predetermined minimum value; and (4) the average number of fluctuations per second." Daly and Warren (2001) used instead the first differential of the pitch curves to develop a measure of dynamism expressed in ERB/s and semitones/s. Similar measures are available in ProsodyPro (Xu, 2013). Although such measures can be informative, a possible issue is that they give equal weight to pitch movements that are deliberate (e.g., part of an accent), and others that may be incidental (e.g., transitions between accents; see section 5.2). It is unclear whether listeners attend equally to both types.

Much of the work that resorts to general descriptive measures of F0, such as average F0, is not conceived with some specific model of intonation in mind. Rather, in this work, F0 is treated as the main object of inquiry (e.g., Cooper & Sorensen, 1981). Other research on intonation has been couched in terms of a number of different models. Some of these models—such as INTSINT (International Transcription System for Intonation; Hirst & Di Cristo, 1998) and the frameworks collectively known as the British school (e.g., Crystal, 1969; O'Connor & Arnold, 1973)—aim to present idealizations of F0 curves. For example, INTSINT includes the categories **T** (Top), **H** (Higher), **U** (Upstepped), **S** (Same), **M** (mid), **D** (Downstepped), **L** (Lower), **B** (Bottom), which can be used to reconstruct pitch tracks in a way that abstracts away from phonetic detail (e.g., by replacing curves with straight lines). Such models cannot easily capture useful generalizations about intonation, or describe phonetic detail. Other models, like PENTA (Parallel Encoding and Target Approximation; Xu, 2005) and Fujisaki (1983), focus directly on modeling F0 detail instead. For

example, in PENTA, modeling success is measured based on how close the approximations remain to the original F0 curves. Such models capture the phonetics of F0 but have difficulty with intonation generalizations and some types of phonetic detail (see Arvaniti & Ladd, 2009, and Arvaniti, 2019, for discussions and illustrations). As argued by Arvaniti (2019), all these models, whether they focus on phonetic detail or rely on idealizations, essentially model F0 rather than intonation *per se*.

5.2. The Autosegmental-Metrical Model

A model that provides a principled separation between F0 curves and intonation is the *autosegmental-metrical model of intonational phonology* (henceforth AM).[3] By doing so, AM can account both for phonetic detail and allow for phonological generalizations (Arvaniti & Ladd, 2009; Arvaniti, 2019). The essential tenets of the model are largely based on Pierrehumbert's dissertation (1980), with additional refinements built on experimental research and formal analysis involving a large number of languages (see also Bruce, 1977, for an early understanding of tonal alignment and the decomposition of tunes into lexical and phrasal elements; see Ladd, 2008, for a theoretical account; see Gussenhoven, 2004, and Jun, 2005B, 2014, for language surveys; see Arvaniti, IN PRESS-B, for an overview of AM).

The term *autosegmental-metrical* was coined by Ladd (1996) and reflects the connection between two sub-systems of phonology required to adequately account for intonation structure, an autosegmental tier representing intonation's melodic part, and metrical structure representing phrasing and relative prominence. In AM, tunes are *phonologically* represented as a string of Low (L) and High (H) tones and combinations thereof (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Ladd, 2008). Tones are *autosegments*, abstract symbolic primitives that are independent of vowels and consonants. Their identity as Hs and Ls is determined by phonetic observation and defined in relative terms: H is used to represent tones deemed to be high in a melody with respect to the speaker's range and other tones in the same contour; L is used to represent tones deemed to be low by the same criteria (cf. Pierrehumbert, 1980, pp. 68–75). Tonal events (which may be composed of more than one tone) are considered morphemes with pragmatic meaning. All events in a melody contribute compositionally to the pragmatic interpretation of an utterance in tandem with propositional meaning and other pragmatic context (Pierrehumbert & Hischberg, 1990).

The relationship between tonal autosegments and the segmental string (often referred to as *tune-text association*) is mediated by metrical structure. Specifically, in AM, tones associate either with constituent heads (informally, stressed syllables), or with phrasal boundaries. The former are referred to as *pitch accents*, for example, H*. The star notation reflects the fact that this tone is meant to be phonologically associated to a stressed syllable. Tones that associate with phrasal boundaries are collectively known as *edge tones*. All AM analyses recognize *boundary tones* as a type of edge tone, for example, H%. Many analyses also recognize a second type of edge tone, the *phrase accent*, for example, H-. Following Beckman and Pierrehumbert (1986), it is by and large understood that when
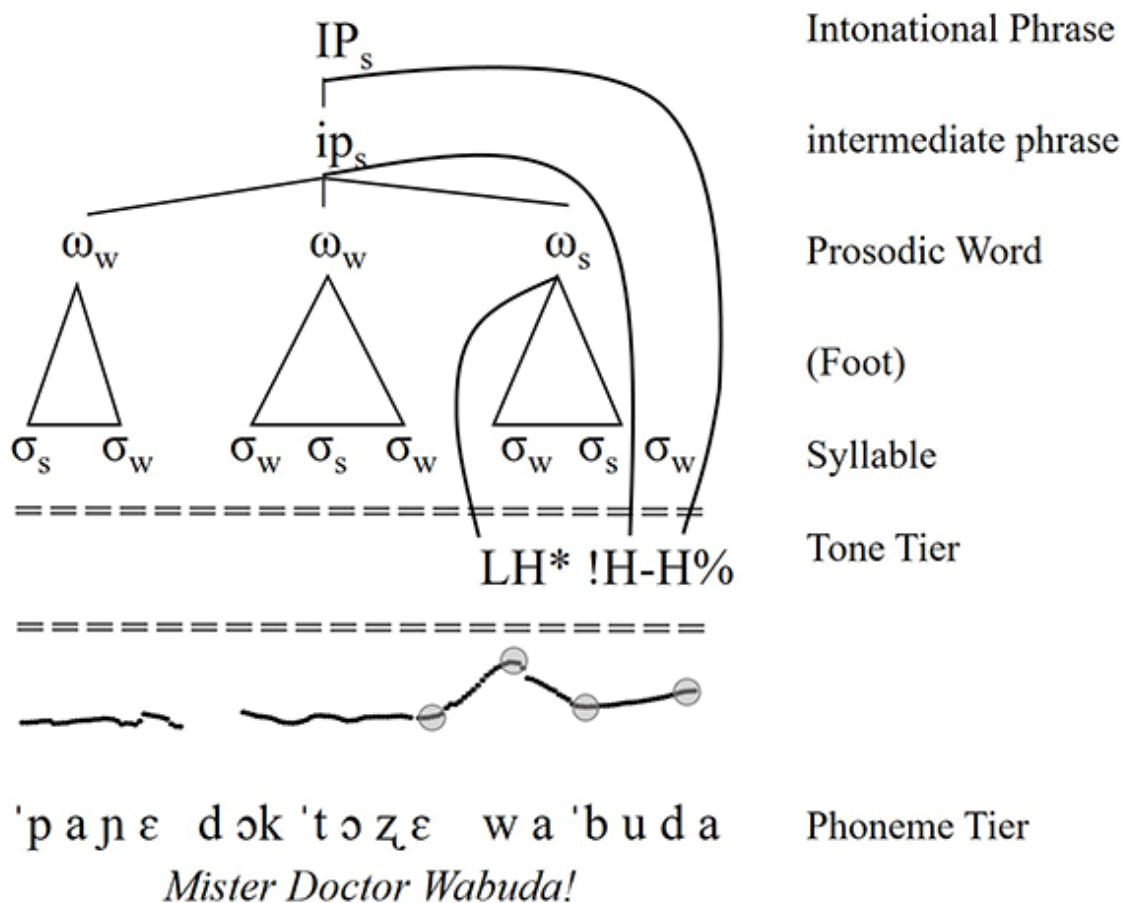
both types of edge tones are posited, phrase accents associate with intermediate phrase boundaries, and boundary tones with intonational phrase boundaries. The representation and F0 contour in (2) provide an illustration of AM, using the same example as in (1).

(2)
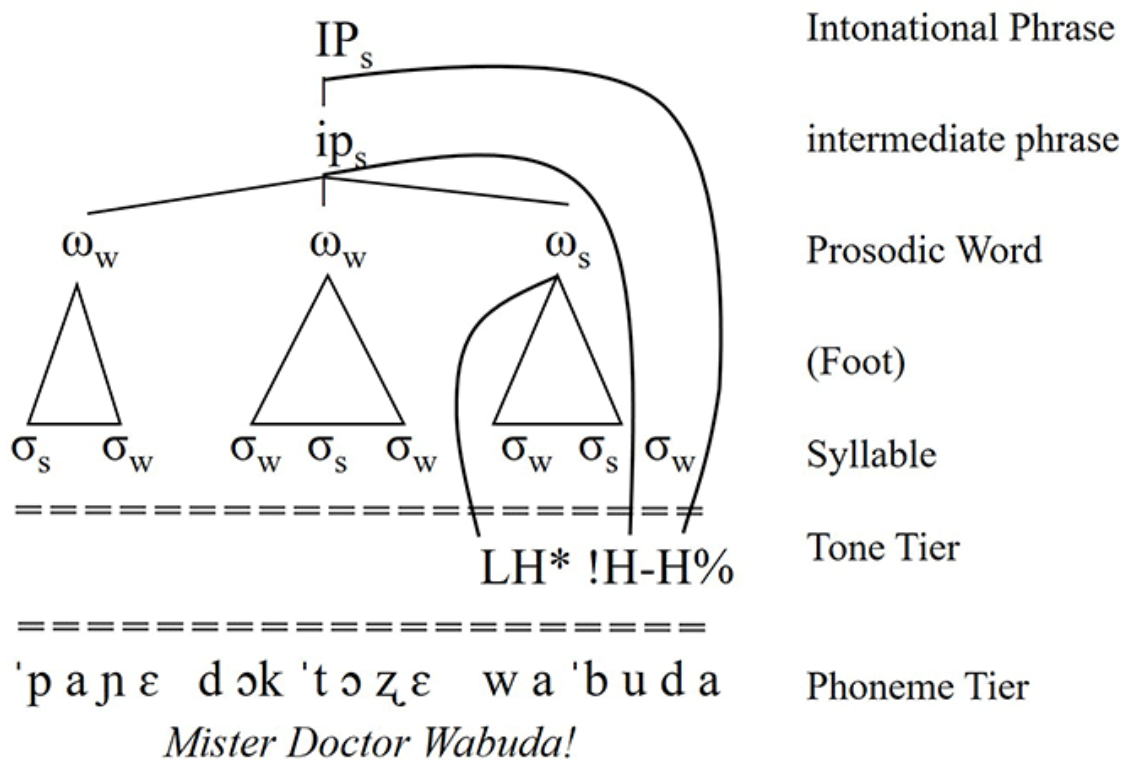
The abstract phonological primitives of intonation are phonetically realized as *tonal targets*, that is, as points in the F0 contour. Tonal targets are usually *turning points*, such as peaks, troughs, and elbows in the contour; they are defined by their *scaling* and *alignment*. Scaling refers to the value of targets on an F0 or pitch scale; alignment refers to their position relative to segments, such as the onset of a stressed vowel or a phrase-final syllable. The representation and F0 contour in (3) illustrate the connection between phonological representation and phonetic realization, using the same example as in (1) and (2), and including the F0 track of the utterance with the tonal targets corresponding to the tune's four tones marked as circles.

(3)

IP_s — Intonational Phrase

ip_s — intermediate phrase

ω_w ω_w ω_s — Prosodic Word

(Foot)

σ_s σ_w σ_w σ_s σ_w σ_w σ_s σ_w — Syllable

LH* !H-H% — Tone Tier

ˈpaɲɛ dɔkˈtɔʑɛ waˈbuda — Phoneme Tier

Mister Doctor Wabuda!

In AM, scaling is said to take place on the fly with every tone's scaling being calculated as a fraction of the scaling of the preceding tone (Liberman & Pierrehumbert, 1984). There are three main influences on tonal scaling: declination, tonal context, and tonal identity. Following Pierrehumbert (1980), it is generally understood that the scaling of tones can be modeled with reference to a declining baseline that is invariant for each speaker (at a given time). The baseline is defined by its slope and a minimum value that is assumed to represent the bottom of the speaker's range, a value that is considered stable for each speaker (Maeda, 1976; Menn & Boyce, 1982; Pierrehumbert & Beckman, 1988). The effect of declination is a systematic lowering of targets, though declination can be suspended (e.g., in questions), and is reset across phrasal boundaries (Ladd, 1988; see also Truckenbrodt, 2002). Listeners anticipate declination effects and adjust their processing of tonal targets accordingly (e.g., Yuen, 2007). L and H tones (apart from terminal L%s) are scaled above the baseline and with reference to it (cf. Liberman & Pierrehumbert, 1984). An exception is final peaks, which exhibit what Liberman and Pierrehumbert (1984) have called *final lowering*, because they are scaled lower than predicted. Final lowering has been reported in several languages with very different prosodic systems, including Japanese (Pierrehumbert & Beckman, 1988), Dutch (Gussenhoven & Rietveld, 1988), Yoruba (Connell & Ladd, 1990, and Laniran & Clements, 2003), Kipare (Herman, 1996), Spanish (Prieto, Shih, & Nibert, 1996), and Greek (Arvaniti & Godjevac, 2003).

As mentioned, tonal alignment is defined as the position of the tonal target relative to the segmental string. Alignment is closely related to phonological association: for example, pitch accents are expected to co-occur with the metrically prominent syllable with which they are associated; boundary tones are associated with phrasal boundaries and typically realized on the phrase-final syllable (but see Gussenhoven, 2000, for an alternative synchronization of boundary tones). The strict phonetic alignment observed by Arvaniti, Ladd, and Mennen (1998) for Greek pitch accents of the form L*+H gave rise to the notion of *segmental anchoring*, the idea that tonal targets anchor onto particular segments in phonetic realization. Specifically, in Greek, Arvaniti et al. found that the L* tone is synchronized with the onset of the accented syllable, while the H appears roughly 10 ms after the onset of the first post-accentual vowel. Segmental anchoring was explored in subsequent work by Ladd and colleagues (e.g., Ladd & Schepman, 2003; Atterer & Ladd, 2004; Ladd, Schepman, White, Quarmby, & Stackhouse, 2009). The idea of segmental anchoring also spurred a great deal of research in a variety of languages that largely supported it (among many, D'Imperio, 2001, for Neapolitan Italian; Myers, 2003, for Kinyarwanda; Elordieta & Calleja, 2005, for Basque Spanish; Arvaniti & Garding, 2007, for American English; Dalton & Ní Chasaide, 2007, for Irish; Gordon, 2008, for Chickasaw; Prieto, 2009, for Catalan). However, it is not the case that such anchoring is equally strict in all languages, as demonstrated, for example, by Smiljanic (2006) for Serbian and Croatian, and by Welby and Lœvenbruck (2006) for French. Alignment variability may be related to a lack of sufficient vocalic material (e.g., Baltazani & Kainada, 2015, on Epirus Greek; Grice, Ridouane, & Roettger, 2015, on Berber), pertain to a specific tonal event (Frota, 2002, Portuguese), or may simply be just a feature of a given intonational system (e.g., Arvaniti, 2016, on Romani). One consistent finding regarding alignment is that many rising accents show *peak delay*, a term that refers to the fact that accentual pitch peaks appear after the syllable with which the accent is phonologically associated. This was first documented by Silverman and Pierrehumbert (1990), who examined the phonetic realization of **prenuclear** H* accents in American English, and has since been reported for South American Spanish (Prieto, van Santen, & Hirschberg, 1995), Greek (Arvaniti et al., 1998), Kinyarwanda (Myers, 2003), Catalan (Prieto, 2005), Irish (Dalton & Ní Chasaide, 2007), Chickasaw (Gordon, 2008), Bininj Gun-wok (Bishop & Fletcher, 2005), and Romani (Arvaniti, 2016), inter alia.

In AM, tonal targets are considered to be the sole exponents of the underlying phonological representation of intonation (but see later in this section for recent developments on this point). The rest of the contour is derived by interpolation between targets. Interpolation between targets is considered to be linear, with the exception of the *sagging interpolation* between H* pitch accents in English which, according to Pierrehumbert (1981), gives rise to an F0 dip between the two accentual peaks (for an alternative analysis that posits that the sag is the reflex *of a low tone*, see Ladd & Schepman, 2003).

The fact that the phonetic implementation of an AM phonological representation relies solely on the realization of its phonological tones as tonal targets means that at the phonetic level the parts of the F0 contour that are not targets do not need to be specified in order to be realized. In other words, there is no requirement for each syllable in an utter-

ance to have some tonal specification, and in fact most syllables are not assigned a specific F0 value during production. This is referred to as **underspecification** in AM. **Underspecification** was first illustrated by Pierrehumbert and Beckman (1988, pp. 13ff.) for Tokyo Japanese accentual phrases (APs). They showed that the F0 contours of APs without an accented word could be successfully modeled by positing only one H target, associated with the AP's second mora, and one L target realized at the beginning of the following AP; the F0 slope from the H to the L target depends on the number of moras between the two. This change in F0 slope is difficult, if not impossible, to model if every mora is specified for F0, as such specifications would need to differ by AP length.

While data like those of Japanese show sparse tonal specification, AM predicts that it is also possible for an utterance to involve more tones than tone bearing units, a phenomenon known as *tonal crowding*. The Greek contour shown in Figure 7, [zi] 'is s/he alive?', is such an instance: the phonological representation of this contour is L* (L+)H- L% (Arvaniti, Ladd, & Mennen, 2006A), and all tones are associated with the single vowel in the utterance. In order for the tones to be realized, this vowel is significantly lengthened: in Figure 7, it is 350 ms long. In contrast, in [epi'zisane] 'did they survive?', shown in Figure 8, there is sufficient segmental material for each tone to be realized on a different syllable, so the stressed [i] is just 120 ms long. Tonal crowding is extremely frequent, yet AM is the only model of intonation that can successfully handle it and predict its outcomes (see Arvaniti & Ladd, 2009, and Arvaniti & Ladd, 2015, for a comparison of the treatment of tonal crowding in AM and PENTA).

Tonal crowding is phonetically resolved in a number of ways: (a) *truncation*, the elision of part of the contour (Bruce, 1977, on Swedish; Grice, 1995, on British English; Arvaniti, 1998, on Cypriot Greek; Grabe, 1998, on English and German; Grabe, Post, Nolan, & Farrar, 2000, on British English; Arvaniti & Ladd, 2009, on Standard Greek); (b) *undershoot*, the realization of all tones without them reaching their targets (Bruce, 1977, on Swedish; Arvaniti, Ladd, & Mennen, 1998, 2000, 2006A, 2006B, on Standard Greek; Prieto, 2005, on Catalan; Arvaniti & Ladd, 2009, on Standard Greek); (c) temporal realignment of tones (Silverman & Pierrehumbert, 1990, on American English); (d) segmental lengthening, as in the example in Figure 8, the aim of which is to accommodate the realization of all tones with as little undershoot as possible (Arvaniti & Ladd, 2009, on Standard Greek; Grice, Savino, & Roettger, 2019, on Bari Italian). Undershoot and temporal realignment often work synergistically giving rise to *compression* (e.g., Arvaniti, Żygis, & Jaskuła, 2017, on Polish). Empirical evidence indicates that the mechanism used is specific to elements in a tune (Ladd, 2008; Arvaniti & Ladd, 2009; Arvaniti, 2016). Arvaniti (2016) and Arvaniti et al. (2017) in particular have argued that such different responses to tonal crowding can be used as a diagnostic to determine which parts of a tonal event are optional (those that are truncated in tonal crowding) and which are required (those that are compressed under the same conditions). Arvaniti et al. (2017) further posit that phonological representations should only include required elements.
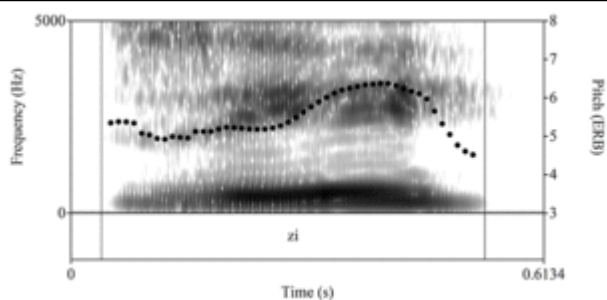
Figure 7. Spectrogram and F0 of Greek utterance [zi] 'is s/he alive?', uttered as a question with a tune represented in AM as L* (L+)H- L%.
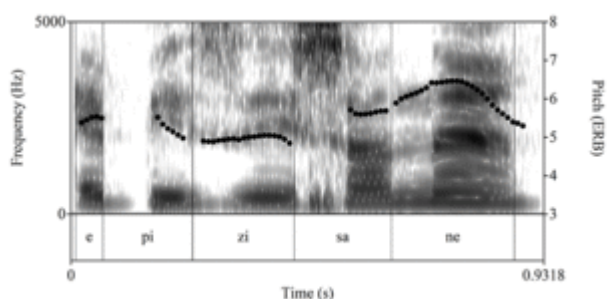
Source: Author.

Figure 8. Spectrogram and F0 of Greek utterance [epi'zisane] 'did they survive?', uttered as a question with a tune represented in AM as L* (L+)H- L%.

Source: Author.

Despite the success of AM, several studies indicate that seeing tonal targets as points connected by linear interpolation may not provide a sufficiently accurate phonetic model of intonation, in the sense that such a model could be missing information that is critical for perception and the encoding of contrasts within an intonation system. Barnes and colleagues have shown that the pitch accents of English represented as L*+H and L+H* differ in terms of shape, the former being concave and the latter convex (Barnes, Veilleux, Brugos, & Shattuck-Hufnagel, 2012; Barnes, Brugos, Shattuck-Hufnagel, & Veilleux, 2013). This difference is not captured by the **autosegmental** representations of these accents, nor anticipated by linear interpolation between the L and H tones. In order to account for this difference, Barnes et al. (2012, 2013) proposed Tonal Centre of Gravity (TCoG), a measurement that aims to capture the difference between accents perceived as predominantly low in pitch and accents perceived as predominantly high. The formula for TCoG is shown in (4).

(4)

$$T_{cog} = \frac{\sum_i f0_i \times t_i}{\sum_i f0_i}$$

An alternative way to think about the difference between English L*+H and L+H* would be to conceive of the L* tone of L*+H as having duration, that is, being a stretch of low F0, rather than being a point in the F0 contour. Similarly, H tones may be realized as **plateaux**. In some languages **plateaux** are used interchangeably with peaks (e.g., Arvaniti, 2016, on Romani); in others the two are distinct, so that the use of peaks or **plateaux** affects the interpretation of the tune (e.g., D'Imperio, 2000, and D'Imperio, Terken, & Piterman, 2000, on Neapolitan Italian), the scaling of the tones involved (e.g., Knight & Nolan, 2006, and Knight, 2008, on British English), or both (Barnes et al., 2013, on American English).

Data like those from **plateaux**, low F0 stretches, and different types of interpolation indicate that a phonetic model involving only targets as turning points and linear interpolation between them may be too simple to fully account for all phonetic detail pertaining to F0 curves. Although the perceptual relevance of additional details is at present far from clear, while there is evidence that details may not be relevant for processing tunes ('t Hart et al., 1990), recent research has focused on capturing such detail. Alternatives to measuring tonal targets include the use of *functional Principal Component Analysis*, which captures tune differences that are difficult to model in terms of targets (Gubian, Torreira, & Boves, 2015; Lohfink, Katsika, & Arvaniti, 2019), and the Synchrony approach of Cangemi, Albert, and Grice (2019), which integrates phonetic prominence and F0 information. Finally, recent studies have questioned the assumption that F0 is the sole exponent of intonation, and suggest instead that intonational categories may also be cued by additional parameters, such as changes in the duration and amplitude of segments synchronized with particular tonal events (see e.g., Niebuhr, 2012, on German; Arvaniti et al., 2017, on Polish; Gryllia, Baltazani, & Arvaniti, 2018, and Lohfink et al., 2019, on Greek).

6. Conclusion

Prosody is an important component of each language's phonology and plays a critical role in speech production, language acquisition, and speech processing and perception. For these reasons it is important to expand research on the numerous facets of prosody. Progress, however, will be hindered if prosody, its phonetic exponents, and their role in expressing paralinguistic information are confounded in research.

Further Reading

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

Arvaniti, A. (2016). Analytical decisions in intonation research and the role of representations: Lessons from Romani. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *7*(1), 6.

Arvaniti, A. (2019). Crosslinguistic variation, phonetic variability, and the formation of categories in intonation. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Canberra, Australia: Australasian Speech Science and Technology Association.

Arvaniti, A., & Ladd, D. R. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, *26*, 46–63.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in laboratory phonology* (Vol. 3, pp. 7–33). Cambridge, UK: Cambridge University Press.

Beckman, M. E., & Venditti, J. J. (2011). Intonation. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The handbook of phonological theory* (pp. 485–532). Malden, MA: Wiley-Blackwell.

Brugos, A., Shattuck-Hufnagel, S., & Veilleux, N. (2006). *Transcribing prosodic structure of spoken utterances with ToBI* (MIT Open Courseware).

Chen, A., Gussenhoven, C., & Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech*, *47*, 311–349.

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

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, *97*(1), 491–504.

Dilley, L., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63*(3), 274–294.

Downing, L. J., & Rialland, A. (Eds). (2017). *Intonation in African tone languages*. Berlin, Germany: De Gruyter Mouton.

Gordon, M. (2011). Stress: Phonotactic and phonetic evidence. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 924–948). Malden, MA: Wiley-Blackwell.

Gordon, M. (2014). Disentangling stress and pitch accent: Toward a typology of prominence at different prosodic levels. In H. van der Hulst (Ed.), *Word stress: Theoretical and typological issues* (pp. 83–118). Cambridge, UK: Cambridge University Press.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

Jun, S., & Fletcher, J. (2014). Methodology of studying intonation: From data collection to data analysis. In S. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 493–519). Oxford, UK: Oxford University Press.

Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics*, *40*, 430–442.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge, UK: Cambridge University Press.

Ladd, D. R. (2014). *Simultaneous structure in phonology*. Oxford, UK: Oxford University Press.

London, J. (2012). *Hearing in time: Psychological aspects of musical meter*. Oxford, UK: Oxford University Press.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Dissertation, MIT). Bloomington, IN: UILC. Published 1988.

Pierrehumbert, J., & Beckman, E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. L. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge MA: MIT Press.

Tremblay, A., Broersma, M., & Coughlin, C. E. (2018). The functional weight of a prosodic cue in the native language predicts speech segmentation in a second language. *Bilingualism: Language and Cognition*, *21*, 640–652.

Turk, A., & Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130395.

References

Adamou, E., & Arvaniti, A. (2014). Illustrations of the IPA: Greek Thrace Xoraxane Romane. *Journal of the International Phonetic Association*, *44*(2), 223–231.

Arvaniti, A. (1991). *The phonetics of Greek rhythm and its phonological implications* (Unpublished doctoral dissertation). University of Cambridge.

Arvaniti, A. (1992). Secondary stress: Evidence from Modern Greek. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 398–423). Cambridge, UK: Cambridge University Press.

Arvaniti, A. (1994). Acoustic features of Greek rhythmic structure. *Journal of Phonetics*, *22*, 239–268.

Arvaniti, A. (1998). Phrase accents revisited: Comparative evidence from Standard and Cypriot Greek. In R. H. Mannell & J. Robert-Ribes (Eds.), *Proceedings of the 5th International Conference on Spoken Language Processing* (Vol. 7, pp. 2883–2886). Sydney: Australian Speech Science and Technology Association, Incorporated (ASSTA).

Arvaniti, A. (2000). The phonetics of stress in Greek. *Journal of Greek Linguistics*, *1*, 9–38.

Arvaniti, A. (2007). On the relationship between phonology and phonetics (or why phonetics is not phonology). In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 19–24). Saarbrücken, Germany: University des Saarlandes.

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

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

Arvaniti, A. (2012b). Rhythm classes and speech perception. In O. Niebuhr (Ed.), *Prosodies: Context, function and communication* (pp. 75–92). Berlin, Germany: Walter de Gruyter.

Arvaniti, A. (2016). Analytical decisions in intonation research and the role of representations: Lessons from Romani. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *7*(1), 6.

Arvaniti, A. (2019). Crosslinguistic variation, phonetic variability, and the formation of categories in intonation. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Canberra, Australia: Australasian Speech Science and Technology Association.

Arvaniti, A. (in press-a). Measuring rhythm. In R. A. Knight & J. Setter (Eds.), *The Cambridge handbook of phonetics*. Cambridge, UK: Cambridge University Press.

Arvaniti, A. (in press-b). The autosegmental-metrical model of intonational phonology. In S. Shattuck-Hufnagel & J. Barnes (Eds.), *Prosodic theory and practice*. Cambridge, MA: MIT Press.

Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. In J. Cole & J. I. Hualde (Eds.), *Papers in laboratory phonology* (Vol. 9, pp. 547–576). Berlin, Germany: Mouton de Gruyter.

Arvaniti, A., & Godjevac, S. (2003). The origins and scope of final lowering in English and Greek. In M.J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona 3–9 August 2003* (pp. 1077–1080). Barcelona, Spain: ICPhS Organizing Committee.

Arvaniti, A., & Ladd, D. R. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, *26*, 46–63.

Arvaniti, A., & Ladd, D. R. (2015). Underspecification in intonation revisited: A reply to Xu, Li, Prom-on and Liu. *Phonology*, *32*(3), 537–541.

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, *26*, 3–25.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2000). What is a Starred Tone? Evidence from Greek. In M. Broe & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 119–131). Cambridge, UK: Cambridge University Press.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006a). Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication*, *48*, 667–696.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006b). Tonal association and tonal alignment: Evidence from Greek polar questions and contrastive statements. *Language and Speech*, *49*, 421–450.

Arvaniti, A., & Rathcke, T. (2015). The role of stress in syllable monitoring. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow.

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

Arvaniti, A., Żygis, M., & Jaskuła, M. (2017). The phonetics and phonology of the Polish calling melodies. *Phonetica*, *73* (3–4), 338–361.

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics*, *32*, 177–197.

Baltazani, M. (2006). Focusing, prosodic phrasing, and hiatus resolution in Greek. In L. Goldstein, D. Whalen, & C. Best (Eds.), *Laboratory phonology* (Vol. 8, pp. 473–494). Berlin, Germany: Mouton de Gruyter.

Baltazani, M., & Kainada, E. (2015). Drifting without an anchor: How pitch accents withstand vowel loss. *Language and Speech*, *58*(1), 84–115.

Barnes, J., Brugos, A., Shattuck-Hufnagel, S., & Veilleux, N. (2013). On the nature of perceptual differences between accentual peaks and plateaux. In O. Niebuhr (Ed.), *Understanding prosody: The role of context, function and communication* (pp. 93–118). Berlin, Germany: De Gruyter.

Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. (2012). Tonal center of gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology*, *3*(2), 337–383.

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht, The Netherlands: Foris.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in laboratory phonology* (Vol 3, pp. 7–33). Cambridge, UK: Cambridge University Press.

Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 68–86). Cambridge, UK: Cambridge University Press.

Beckman, M., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9–54). Oxford, UK: Oxford University Press.

Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook*, *3*, 255–310.

Beckman, M. E., & Venditti, J. J. (2011). Intonation. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The handbook of phonological theory* (pp. 485–532). Malden, MA: Wiley-Blackwell.

Bishop, J., & Fletcher, J. (2005). Intonation in six dialects of Bininj Gun-wok. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 331–361). Oxford, UK: Oxford University Press.

Bolinger, D. L. (1961). *Generality, gradience, and the all-or-none*. The Hague, The Netherlands: Mouton.

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

Botinis, A. (1989). *Stress and prosodic structure in Greek: A phonological, acoustic, physiological and perceptual study*. Lund, Sweden: Lund University Press.

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund, Sweden: Gleerup.

Bruce, G. (2005). Intonational prominence in varieties of Swedish revisited. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 410–429). Oxford, UK: Oxford University Press.

Brugos, A., Shattuck-Hufnagel, S., & Veilleux, N. (2006). *Transcribing prosodic structure of spoken utterances with ToBI* (MIT Open Courseware).

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, *57*(1), 3–16.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149–180.

Cambier-Langeveld, T., & Turk, A. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics*, *27*, 171–206.

Campbell, N., & Beckman, M. E. (1997). Stress, prominence, and spectral tilt. In A. Botinis, G. Kouroupetroglou, & G. Carayannis (Eds.), *Intonation: Theory, models and applications (Proceedings of the ESCA Workshop on Intonation)* (pp. 67–70). Athens, Greece: ESCA and University of Athens Department of Informatics.

Cangemi, F., Albert, A., & Grice, M. (2019). Modelling intonation: Beyond segments and tonal targets. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, 2019*. Canberra, Australia: Australasian Speech Science and Technology Association.

Chen, A., Gussenhoven, C., & Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech*, *47*, 311–349.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*(2), 155–190.

Cho, T., & Keating, P. A. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*(4), 466–485.

Cho, T., Son, M., & Kim, S. (2016). Articulatory reflexes of the three-way contrast in labial stops and kinematic evidence for domain-initial strengthening in Korean. *Journal of the International Phonetic Association*, *46*(2), 129–155.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York, NY: Harper & Row.

Chung, Y., & Arvaniti, A. (2013). **Speech rhythm in Korean: Experiments in speech cycling**. In *Proceedings of Meetings on Acoustics (POMA): Proceedings of 21st International Congress of Acoustics*, Montreal, 2–7 June 2013, 060216.

Classe, A. (1939). *The rhythm of English prose*. Oxford, UK: Basil Blackwell.

Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics*, *39*(2), 237–245.

Conderman, G., & Strobel, D. (2010). Fluency Flyers Club: An oral reading fluency intervention program. *Preventing School Failure: Alternative Education for Children and Youth*, *53*(1), 15–20.

Condoravdi, C. (1990). Sandhi rules of Greek and prosodic theory. In S. Inkelas & D. Zec (Eds.), *Phonology-syntax connection* (pp. 63–84). Chicago, IL: University of Chicago Press.

Connell, B., & Ladd, D. R. (1990). Aspects of pitch realization in Yoruba. *Phonology*, 7, 1–30.

Cooper, W., & Sorensen, J. (1981). *Fundamental frequency in sentence production*. Heidelberg, Germany: Springer.

Crosswhite, K. (2003). Spectral tilt as a cue to word stress in Polish, Macedonian, and Bulgarian. In M.-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona 3–9 August 2003* (pp. 767–770). Barcelona, Spain: ICPhS Organizing Committee.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge, UK: Cambridge University Press.

Cutler. A. (2015). Lexical stress in English pronunciation. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 106–124). New York, NY: John Wiley & Sons.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3–4), 133–142.

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., Mehler, J., Norris, D., & Seguí, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*, 381–410.

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

Dalton, M., & Ní Chasaide, A. (2007). Melodic alignment and micro-dialect variation in Connemara Irish. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes* (pp. 293–315). Berlin, Germany: Mouton de Gruyter.

Daly, N., & Warren, P. (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics*, *5*(1), 85–96.

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. In *Proceedings of the 11th International Congress of Phonetic Sciences* (pp. 447–450). Academy of Sciences of the Estonian S.S.R.

de Jong, K. J. (1994). Initial tones and prominence in Seoul Korean. *OSU Working Papers in Linguistics*, *43*, 1–14.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, *97*(1), 491–504.

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

Dilley, L. C., & Brown, M. (2007). Effects of pitch range variation on f_0 extrema in an imitation task. *Journal of Phonetics*, *35*(4), 523–551.

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

Dilley, L., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63*(3), 274–294.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*, 423–444.

D'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment* (Unpublished doctoral dissertation). The Ohio State University.

D'Imperio, M. (2001). Focus and tonal structure in Neapolitan Italian. *Speech Communication*, *33*(4), 339–356.

D'Imperio, M., Elordieta, G., Frota, S., Prieto, P., & Vigário, M. (2005). Intonational phrasing in Romance: The role of syntactic and prosodic structure. In S. Frota, M. Vigário, & M. J. Freitas (Eds.), *Prosodies* (pp. 59–98). The Hague, The Netherlands: Mouton de Gruyter.

D'Imperio, M., & Rosenthal, S. (1999). Phonetics and phonology of main stress in Italian. *Phonology*, *16*, 1–28.

D'Imperio, M., Terken, J., & Piterman, M. (2000). Perceived tone "targets" and pitch accent identification in Italian. In M. Barlow (Ed.), *Proceedings of Australian International Conference on Speech Science and Technology* (SST) (Vol. 8, pp. 201–211). Canberra: Australian Speech Science and Technology Association.

Downing, L. J., & Rialland, A. (Eds). (2017). *Intonation in African tone languages*. Berlin, Germany: De Gruyter Mouton.

Eady. S. J. (1982). Differences in the F0 patterns of speech: Tone language versus stress language. *Language and Speech*, *25*, 29–42.

Elordieta, G., & Calleja, N. (2005). Microvariation in accentual alignment in Basque Spanish. *Language and Speech*, *48*, 397–439.

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

Fletcher, J., Grabe, E., & Warren, P. (2004). Intonational variation in four dialects of English: The high rising tune. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 390–409). Oxford, UK: Oxford University Press.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, *29*, 109–135.

Fougeron, C., & Jun, S. (2002). Realizations of accentual phrase in French intonation. *Probus*, *14*(1), 147–172.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, *101*(6), 3728–3740.

Fourakis, M., Botinis, A., & Katsaiti, M. (1999). Acoustic characteristics of Greek vowels. *Phonetica*, *56*, 28–43.

Fraisse, P. (1963). *The psychology of time*. New York, NY: Harper and Row.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York, NY: Academic Press.

Frota, S. (2002). Tonal association and target alignment in European Portuguese nuclear falls. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (Vol. 7, pp. 387–418). Berlin, Germany: Mouton de Gruyter.

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

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*(2), 126–152.

Fujisaki, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 39–55). Heidelberg, Germany: Springer-Verlag.

Fujisaki, H., & Hirose, K. (1984). Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan (E)*, *5*(4), 233–242.

Garellek, M., & White, J. (2015). Phonetics of Tongan stress. *Journal of the International Phonetic Association*, *45*(1), 13–34.

Gaudio. R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech*, *69*(1), 30–57.

Georgeton, L., Antolík, T. K., & Fougeron, C. (2016). Effect of domain initial strengthening on vowel height and backness contrasts in French: Acoustic and ultrasound data. *Journal of Speech, Language and Hearing Research*, *59*(6), S1575–S1586.

Glasberg, B. R., & Moore, B. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*, 103–138.

Gooden, S., Drayton, K., & Beckman, M. (2009). Tone inventories and tune-text alignments: Prosodic variation in "hybrid" prosodic systems. *Studies in Language*, *33*(2), 354–394.

Gordon, M. (2008). Pitch accent timing and scaling in Chickasaw. *Journal of Phonetics*, *36*(3), 521–535.

Gordon, M. (2011). Stress: Phonotactic and phonetic evidence. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 924–948). Malden, MA: Wiley-Blackwell.

Gordon, M. (2014). Disentangling stress and pitch accent: Toward a typology of prominence at different prosodic levels. In H. van der Hulst (Ed.), *Word stress: Theoretical and typological issues* (pp. 83–118). Cambridge, UK: Cambridge University Press.

Gordon, M., & Applebaum, A. (2010). Acoustic correlates of stress in Turkish Kabardian. *Journal of the International Phonetic Association*, *40*, 35–58.

Gow, D. W., Jr. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 163–179.

Grabe, E. (1998). *Comparative intonational phonology: English and German*. MPI Series in Psycholinguistics 7. Wageningen, Germany: Ponsen en Looien.

Grabe, E., & Low, E. L. (2002). Acoustic correlates of rhythm class. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (Vol. 7, pp. 515–546). Berlin, Germany: Mouton de Gruyter.

Grabe, E., Post, B, Nolan, F., & Farrar, K. (2000). Pitch accent realization in four varieties of British English. *Journal of Phonetics*, *28*, 161–185.

Graham, C. R. (2014). Fundamental frequency range in Japanese and English: The case of simultaneous bilinguals. *Phonetica*, *71*, 271–295.

Grice, M. (1995). Leading tones and downstep in English. *Phonology*, *12*(2), 183–233.

Grice, M., Ridouane, R., & Roettger, T. (2015). Tonal association in Tashlhiyt Berber: Evidence from polar questions and contrastive statements. *Phonology*, *32*(2), 241–266.

Grice, M., Savino, M., & Roettger, T. B. (2019). Tune-text negotiation: The effect of intonation on vowel duration. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Canberra, Australia: Australasian Speech Science and Technology Association.

Gryllia, S., Baltazani, M., & Arvaniti, A. (2018). **The role of pragmatics and politeness in explaining prosodic variability**. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of the 9th International Conference on Speech Prosody 2018* (pp. 158–162). Poznań.

Gubian, M., Torreira, F., & Boves, L. (2015). Using functional data analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, *49*, 16–40.

Gussenhoven, C. (2000). The boundary tones are coming: On the non-peripheral realization of boundary tones. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 132–151). Cambridge, UK: Cambridge University Press.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

Gussenhoven, C., & Jacobs, H. (2017). *Understanding phonology*. New York, NY: Routledge.

Gussenhoven, C., & Rietveld, A. C. M. (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics*, *16*, 355–369.

Haag, W. K. (1979). An articulatory experiment on Voice Onset Time in German stop consonants. *Phonetica*, *36*(3), 169–181.

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*, *7*, 939.

Harrington, J., Fletcher, J., & Roberts, C. (1995). Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. *Journal of Phonetics*, *23*(3), 305–322.

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*, 11–29.

Hart, J. 't, Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An Experimental-phonetic approach to speech melody*. Cambridge, UK: Cambridge University Press.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago, IL: University of Chicago Press.

Haung, T., & Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica*, *67*, 243–267.

Henton, C. G. (1989). Fact and fiction in the description of female and male pitch. *Language & Communication*, *9*(4), 299–311.

Herman, R. (1996). Final lowering in Kipare. *Phonology*, *13*, 171–196.

Hermes, D. J., & van Gestel, J. C. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, *90*, 97–102.

Hirschberg, J., & Avesani, C. (2000). Prosodic disambiguation in English and Italian. In A. Botinis (Ed.), *Intonation: Analysis, modeling and technology* (pp. 87–96). Dordrecht, The Netherlands: Springer.

Hirst, D., & Di Cristo, A. (Eds.). (1998). *Intonation systems: A survey of twenty languages*. Cambridge, UK: Cambridge University Press.

Hirst, D., Di Cristo, A., & Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In M. Horne (Ed.), *Prosody: Theory and experiment, studies presented to Gösta Bruce* (pp. 51–87). Dordrecht, The Netherlands: Kluwer Academic.

Horton, R., & Arvaniti, A. (2013). **Cluster and classes in the rhythm metrics**. *San Diego Linguistic Papers*, *4*, 28–52.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., . . . Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *454*, 903–995.

Huss, V. (1978). English word stress in the post-nuclear position. *Phonetica*, *35*, 86–105.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, *23*, 225–257.

James, W. (1890). *The principles of psychology*. New York, NY: Dover Reprint.

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

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

Jun, S. (2005a). Korean intonational phonology and prosodic transcription. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 201–229). Oxford, UK: Oxford University Press.

Jun, S. (Ed). (2005b). *Prosodic typology: The phonology of intonation and phrasing*. Oxford, UK: Oxford University Press.

Jun, S. (Ed.). (2014). *Prosodic typology II: The phonology of intonation and phrasing*. Oxford, UK: Oxford University Press.

Jun, S., & Fletcher, J. (2014). Methodology of studying intonation: from data collection to data analysis. In S. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 493–519). Oxford, UK: Oxford University Press.

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

Katsika, A. (2016). The role of prominence in determining the scope of boundary lengthening in Greek. *Journal of Phonetics*, *55*, 149–181.

Katsika, A., Shattuck-Hufnagel, S., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). Compatible vs. competing rhythmic grouping and errors. *Language and Speech*, *57*, 544–562.

Knight, R. A. (2008). The shape of nuclear falls and their effect on the perception of pitch and prominence: Peaks vs. plateaux. *Language and Speech*, *51*(3), 223–244.

Knight, R. A., & Nolan, F. (2006). The effect of pitch span on intonational plateaux. *Journal of the International Phonetic Association*, *36*(1), 21–38.

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

Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics*, *40*, 430–442.

Ladd, D. R. (1988). Declination "reset" and the hierarchical organization of utterances. *Journal of the Acoustical Society of America*, *84*, 530–544.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge, UK: Cambridge University Press.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge, UK: Cambridge University Press.

Ladd, D. R. (2014). *Simultaneous structure in phonology*. Oxford, UK: Oxford University Press.

Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, *31*, 81–112.

Ladd, D. R., Schepman, A., White, L., Quarmby, L. M., & Stackhouse, R. (2009). Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics*, *37*(2), 145–161.

Laniran, Y. O., & Clements, G. N. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics*, *31*, 203–250.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

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

Liberman, M. Y., & Pierrehumbert, J. B. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure:*

Studies in phonology presented to Morris Halle (pp. 157–233). Cambridge, MA: MIT Press.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249–336.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, *32*, 451–454.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (Vol. 55, pp. 403–439). NATO ASI Series (Series D: Behavioural and Social Sciences). Dordrecht, The Netherlands: Springer.

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Spitzer, S., Lotto, A. J., & Caviness, J. N. (2009). Quantifying speech rhythm deficits in the dysarthrias. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1334–1352.

Lloyd James, A. (1940). *Speech signals in telephony*. London, UK: Pitman & Sons.

Lohfink, G., Katsika, A., & Arvaniti, A. (2019). *Variability and category overlap in the realization of intonation*. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Canberra, Australia: Australasian Speech Science and Technology Association.

London, J. (2012). *Hearing in time: Psychological aspects of musical meter*. Oxford, UK: Oxford University Press.

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

Loutrari, A., Tselekidou, F., & Proios, H. (2018). Phrase-final words in Greek storytelling speech: A study on the effect of a culturally-specific prosodic feature on short-term memory. *Journal of Psycholinguistic Research*, *47*(4), 947–957.

Lowit, A. (2014). Quantification of rhythm problems in disordered speech: A re-evaluation. *Philosophical Transactions of the Royal Society B*, *369*(1658).

Maeda, S. (1976). *A characterization of American English intonation* (Unpublished dissertation). MIT.

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cerebral Cortex*, *17*, 2659—2668.

Malikouti-Drachman, A., & Drachman, G. (1992). Greek clitics and lexical phonology. In W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer, & J. R. Rennison (Eds.), **Phonologica 1988** (pp. 197–206). Cambridge, UK: Cambridge University Press.

Maskikit-Essed, R., & Gussenhoven, C. (2016). No stress, no pitch accent, no prosodic focus: The case of Ambonese Malay. *Phonology*, *33*(2), 353–389.

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.

Menn, L., & Boyce, S. (1982). Fundamental frequency and discourse structure. *Language and Speech*, *25*, 341–383.

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

Moore, B. C. J. (2012). *An introduction to the psychology of hearing* (6th ed.). Bingley, UK: Emerald Group.

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics*, *34*, 202–240.

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

Myers, S. (2003). F0 Timing in Kinyarwanda. *Phonetica*, *60*, 71–97.

Nábělek, I. V., Nábělek, A. K., & Hirsh, I. J. (1970). Pitch of tone bursts of changing frequency. *Journal of the Acoustical Society of America*, *48*, 536–553.

Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics*, *37*, 29–45.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception Performance*, *24*, 756–766.

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. (1986). *Prosodic phonology*. Dordrecht, The Netherlands: Foris.

Niebuhr, O. (2012). At the edge of intonation: The interplay of utterance-final F0 movements and voiceless fricative sounds in German. *Phonetica*, *69*, 7–21.

Nolan, F. (1992). The descriptive role of segments: Evidence from assimilation. In G. J. Doherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 261–280). Cambridge, UK: Cambridge University Press.

Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales.In M.J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona 3–9 August 2003 (pp. 771–774). Barcelona, Spain: ICPhS Organizing Committee.

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

O'Connor, J. D., & Arnold, G. F. (1973). *Intonation of Colloquial English*. London, UK: Longman.

Ortega-Llebaria, M., Hong, G., & Fan, Y. (2013). English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception. *Journal of Phonetics*, *41*(3–4), 186–197.

Ortega-Llebaria, M., Nemogá, M., & Presson, N. (2017). Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese-English bilinguals perceive "Rose?" vs. "Rose". *Bilingualism: Language and Cognition*, *20*(2), 367–383.

Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró, & M. J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 155–176). Amsterdam, The Netherlands: John Benjamins.

Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech*, *54*(1), 1–25.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 358–378.

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

Peng, S., Chan, M. K. M., Tseng, C., Huang, T., Lee, O. K., & Beckman, M. E. (2005).Towards a pan-Mandarin system for prosodic transcription. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 230–270). Oxford, UK: Oxford University Press.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Dissertation). MIT. Published 1988, Bloomington, IN: IULC.

Pierrehumbert, J. B. (1981). Synthesizing intonation. *Journal of the Acoustical Society of America*, *70*, 985–995.

Pierrehumbert, J., & Beckman, E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.

Pierrehumbert, J., & Hirschberg, J. (1990).The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. L. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge MA: MIT Press.

Pons, F., & Bosch, L. (2010). Stress pattern preference in Spanish-learning infants: The role of syllable weight. *Infancy*, *15*(3), 223–245.

Post, B., & Payne, E. (2018). Speech rhythm in development: What is the child acquiring? In P. Prieto & N. Esteve-Gibert (Eds.), *The development of prosody in first language acquisition* (pp. 125–144). Amsterdam, The Netherlands: John Benjamins.

Prieto, P. (2005). Stability effects in tonal clash contexts in Catalan. *Journal of Phonetics*, *33*(2), 215–242.

Prieto, P. (2009). Tonal alignment patterns in Catalan nuclear falls. *Lingua*, *119*, 865–880.

Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics*, *24*, 445–473.

Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, *23*, 429–451.

Protopapas, A., Panagaki, E., Andrikopoulou, A., Gutiérrez Palma, N., & Arvaniti, A. (2016). Priming stress patterns in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(11), 1739–1760.

Qin, Z., Chien, Y.F., & Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second-language learners of English. *Applied Psycholinguistics*, *38*, 541–570.

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm class: Perceptual studies. In M.J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona 3–9 August 2003 (pp. 337–340). Barcelona, Spain: ICPhS Organizing Committee.

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

Recasens, D., & Espinosa, A. (2005). Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: Evidence from two Catalan dialects. *Journal of the International Phonetic Association*, *35*(1), 1–25.

Renwick, M. E. L. (2013). Quantifying rhythm: Interspeaker variation in %V. *Proceedings of Meetings on Acoustics (POMA), 14*, 060011.

Rietveld, A. C. M., & Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, *13*, 299–308.

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* (pp. 73–79). London, UK: Edward Arnold.

Rogers, D., & D'Arcangeli, L. (2004). Italian. *Journal of the International Phonetic Association*, *34*(1), 117–121.

Rothermich, K., Schmidt-Kassow, M., Schwartze M., & Kotz, S. A. (2010). Event-related potential responses to metric violations: Rules versus meaning. *NeuroReport*, *21*(8), 580–584.

Schmidt-Kassow, M., & Kotz, S. A. (2008). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience*, *21*(9), 1693–1708.

Selkirk, E. (1980). Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff (Ed.), *Juncture* (pp. 107–129). Saratoga, CA: Anma Libri.

Selkirk, E. O. (1984). *Phonology and syntax: The relationship between sound and structure*. Cambridge, MA: MIT Press.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., . . .Hirschberg, J. (1992). ToBI: A standard for labelling English prosody. In **Proceedings of the 1992 International Conference on Spoken Language Processing, 12–16 October, Banff**. Banff, AB: ISCA.

Silverman, K., & Pierrehumbert, J. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 72–106). Cambridge, UK: Cambridge University Press.

Skoruppa, K., Cristià, A., Peperkamp, S., & Seidl, A. (2011). English-learning infants' perception of word stress patterns. *Journal of the Acoustical Society of America*, *130*(1), EL50-55.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., . . .Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*, 914–919.

Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*, *100*, 2471–2485.

Smiljanic, R. (2006). Early vs. late focus: Pitch-peak alignment in two dialects of Serbian and Croatian. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory phonology* (Vol. 8, pp. 495–518). Berlin, Germany: Mouton de Gruyter.

Smiljanic, R., & Bradlow, A. (2008). Temporal organization of English clear and conversational speech. *Journal of the Acoustical Society of America*, *124*(5), 3171–3182.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*, 412–432.

Stetson, R. H. (1951). *Motor phonetics: A study of speech movements in action*. Amsterdam, The Netherlands: North-Holland Publishing Company.

Stevens, S. S., & Volkmann, J. (1940). The relation of pitch to frequency: A revised scale. *The American Journal of Psychology*, *53*(3), 329–353.

Suomi, K., & Ylitalo, R. (2004). On durational correlates of word Stress in Finnish. *Journal of Phonetics*, *32*(1), 35–63.

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

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice-Hall.

Tremblay, A., Broersma, M., & Coughlin, C. E. (2018). The functional weight of a prosodic cue in the native language predicts speech segmentation in a second language. *Bilingualism: Language and Cognition*, *21*, 640–652.

Truckenbrodt, H. (2002). Upstep and embedded register levels. *Phonology 19*, 77–120.

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

Turk, A., & Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130395.

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, & R. Carlson (Eds.), *Music, language, speech and brain: Proceedings of an international symposium at the Wenner-Gren Center, Stockholm, 5–8 September 1990*. London, UK: Palgrave.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*(3), 253–265.

Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 172–200). Oxford, UK: Oxford University Press.

Wagner, P. S., & Dellwo, V. (2004). Introducing YARD (Yet Another Rhythm Determination) and re-introducing isochrony to rhythm research. In B. Bel & I. Marlien (Eds.), **Proceedings of Speech Prosody 2004**, *Nara, Japan*, March 23–26, 2004.

Waksler, S. (2001). Pitch range and women's sexual orientation. *Word*, *52*, 69–77.

Warren, P. (2005). Patterns of late rising in New Zealand English: Intonational variation or intonational change? *Language Variation and Change*, *17*(2), 209–230.

Welby, P., & Loevenbruck, H. (2006). Anchored down in Anchorage: Syllable structure and segmental anchoring in French. *Italian Journal of Linguistics*, *18*, 74–124.

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? *The Journal of the Acoustical Society of America*, *127*, 1559–1569.

Williams, B. (1985). Pitch and duration in Welsh stress perception: The implications for intonation. *Journal of Phonetics*, *13*(4), 381–406.

Wingfield, A., Lahar, C. J., & Stine, E. A. L. (1989). Age and decision strategies in running memory for speech: Effects of prosody and linguistic structure. *Journal of Gerontology*, *44*(4), P106–P113.

Witteman, J., van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*, *50*(2), 2752–2763.

Wong, W. Y. P., Chan, M. K. M., & Beckman, M. E. (2005). An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 271–300). Oxford, UK: Oxford University Press.

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224–1236). New York, NY: Wiley.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, *95*, 2240–2253.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, *46*(3–4), 220–251.

Xu, Y. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis. In B. Bigi & D. Hirst (Eds.), *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France* (pp. 7–10). Aix en Provence, France: Labratoire Parole et Langage.

Yakup, M., & Sereno, J. A. (2016). Acoustic correlates of lexical stress in Uyghur. *Journal of the International Phonetic Association*, *46*(1), 61–77.

Yuasa, I. P. (2008). *Culture and gender of voice pitch: A sociophonetic comparison of the Japanese and Americans*. London, UK: Equinox.

Yuen, I. (2007). Declination and tone perception in Cantonese. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes* (pp. 63–77). Berlin, Germany: Mouton de Gruyter.

Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence* (pp. 282–302). Cambridge, UK: Cambridge University Press.

Zsiga, E. C. (1997). Features, gestures, and Igbo vowels: An approach to the phonology-phonetics interface. *Language*, *73*(2), 227–274.

Notes:

(1.) Fusion is evident in the impression that the noise of machine guns is a cadence (as implied by the oft-quoted metaphor of French *machine-gun* rhythm, first mentioned in Lloyd James, 1940). If slowed down by a factor of four, however, each machine-gun beat is in fact a complex sequence of sounds, closer to that of a beating heart, that is, an iamb.

(2.) Analyses differ regarding the typology of languages with tonal phenomena at the lexical level. For some, such languages form one category, in that they all use F0 to encode lexical meaning. They differ primarily in terms of the frequency of tonal specifications: at one end of the continuum we find languages like Cantonese, in which every syllable is specified for tone, while at the other end we find languages like Japanese, in which only one syllable per word in a subset of the lexicon is tonally specified (e.g., Beckman & Venditti, 2011). Others make a typological distinction between tonal languages in which more than one syllable is specified for tone and languages with pitch accent, such as Japanese, in which only one syllable is thus specified (e.g., Hyman, 2006). A discussion of this topic is beyond the scope of this article.

(3.) AM is often confused with ToBI (Tones and Break Indices), a framework based on AM and developed for the prosodic *annotation* of spoken corpora. ToBI was first developed for the annotation of American English (Silverman et al., 1992). It has since been revised and renamed to clarify it is designed for Mainstream American English (Brugos, Shattuck-Hufnagel, & Veilleux, 2006). Additional versions adapted to the needs of other languages have been developed in the past 25 years as well (see Jun, 2005b, and Jun, 2014). ToBI is not the prosodic equivalent of the International Phonetic Alphabet (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005): it does not provide universal off-the-shelf categories and re-

quires an AM analysis before it can be implemented for a new language. Jun and Fletcher (2014) and Arvaniti (2016) provide guidance on how to develop such an analysis.

Amalia Arvaniti

Professor of English Language and Linguistics, Faculty of Arts, Radboud University