

This is a section of [doi:10.7551/mitpress/10413.001.0001](https://doi.org/10.7551/mitpress/10413.001.0001)

# Prosodic Theory and Practice

**Edited by: Jonathan Barnes, Stefanie Shattuck-Hufnagel**

## **Citation:**

*Prosodic Theory and Practice*

**Edited by: Jonathan Barnes, Stefanie Shattuck-Hufnagel**

**DOI: 10.7551/mitpress/10413.001.0001**

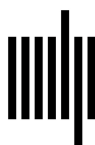
**ISBN (electronic): 9780262543194**

**Publisher: The MIT Press**

**Published: 2022**

## **OA Funding Provided By:**

OA Funding from MIT Press Direct to Open



**The MIT Press**

---

## Commentary on Chapter 4: An Enhanced Autosegmental-Metrical Theory (AM<sup>+</sup>) Facilitates Phonetically Transparent Prosodic Annotation

Laura C. Dilley and Mara Breen

*Entia non sunt multiplicanda praeter necessitate.*

(Translation from Latin: Entities should not be multiplied beyond necessity.)

—Occam's Razor, attributed to William of Ockham (ca. 1285–1349)

### Introduction

We welcome this opportunity to respond to the well-organized, thoughtful essay by Jun in this chapter and to share our perspective on the ToBI\* enterprise—where by ToBI\*, we mean all tones and break indices (ToBI)-like annotation systems, including Mainstream American English MAE\_ToBI, German GToBI, and so on—and how this enterprise fits in with the scientific study of tone and intonation in language. Much time has now passed—some forty years—since the core theoretic ideas behind ToBI\* were put forward in groundbreaking, well-cited PhD dissertations at MIT by Goldsmith (1976), Pierrehumbert (1980), and Liberman (1975), which formed the core ideas in what has come to be known as autosegmental-metrical (AM) theory (Ladd 2008). Furthermore, more than twenty-five years have passed since the original MAE-ToBI was developed (Beckman and Hirschberg 1994; Beckman, Hirschberg, and Shattuck-Hufnagel 2005); this development included the third (and most recent) ToBI workshop in Columbus, Ohio, in 1993, which the first author of this commentary attended following her first undergraduate year at MIT. This long time span provides perspective on the strengths and weaknesses of ToBI\*, as well as the theory that underlies it.

Our commentary aims to contextualize Jun's chapter by highlighting theoretical insights from forty years ago that led to the broad adoption of AM theory, an approach that has facilitated the discovery of important empirical insights about the cross-linguistic structure of intonation. We then show that several serious problems exist with traditional AM theory as it stands, leading to limitations on ToBI\*'s value as a scientific tool. We argue that these problems can be clearly traced to a theoretical failure to prioritize consistent and transparent codification of the role of syntagmatic relationships in intonational phonology. Drawing on empirical evidence about the attested cognitive representations for pitch in the world's nonlinguistic communicative tonal systems (i.e., music), we propose a theoretical clarification of syntagmatic elements in intonational phonology, leading to a proposal for an enhanced AM theory, or AM<sup>+</sup>. We show that attributing both syntagmatic and paradigmatic properties to tones provides a unifying account of multiple outstanding challenges in intonational phonological research that have not yet found a satisfactory explanation, including (i) the tonal composition of Greek prenuclear accents (Arvaniti, Ladd, and Mennen 1998), (ii) influences of contour shape and slope on perception of phonological contrasts (Barnes

et al. 2012; D’Imperio 2000; Niebuhr 2007b), (iii) evidence against a nonmonotonic interpolation function account of F0 turning points on metrically nonprominent syllables (Dilley 2005; Dilley and Heffner 2013; Ladd and Schepman 2003), (iv) the lack of invariant timing in bitonal pitch accents (Dilley, Ladd, and Schepman 2005), (v) characterization of pointed versus plateau-shaped pitch accents (Niebuhr and Hoekstra 2015), and several others. Finally, we present the rhythm and pitch (RaP) prosodic transcription system as an AM<sup>+</sup>-based empirical tool that can be extended toward the goal of developing an international prosodic alphabet (IPrA; Hualde and Prieto 2016).

### Enduring Insights from over Forty Years of Traditional AM Theory

Elaborating on Jun, we highlight some key ideas and findings from the last forty or more years that constitute contributions of ToBI\* to knowledge about tonal aspects of linguistic systems:

- *Tones are autonomous from segments:* That tones are autonomous from segmental structures but temporally coordinated with them was a foundational idea for the field of intonational phonology, as highlighted by Jun. This idea provided the basis for ToBI\*’s descriptive notations, in which entities such as H (high) and L (low) are viewed as discrete tonal events that have abstract associations with segments (e.g., Ladd 2008).
- *Surface intonation contours reflect sparse tonal representations:* Another core idea highlighted by Jun is the idea that tones are sparse; for example, they do not occur on every syllable and are connected via F0 interpolations.<sup>1</sup>
- *Prominence- and boundary-related tones have distinctive distributional properties:* The key idea of ToBI\* that tones participate in either pitch accents or edge tones has stood the test of time. As highlighted by Jun, starred tones of pitch accents associate with (and unstarred tones flank) metrically prominent positions, while phrase tones associate with constituent edges.
- *Peaks, valleys, and elbows are phonologically significant evidence of tones:* Abundant evidence that falls largely outside the scope of Jun’s essay has shown that in general, F0 peaks, valleys, and elbows—transitions from a flat region of pitch to a rise or fall—constitute phonologically significant evidence of “tones” across a wide variety of intonation languages (D’Imperio, Gili Fivela, and Niebuhr 2010; del Giudice et al. 2007; Knight and Nolan 2006; Welby 2006). These F0 points have been argued to serve as “control points” in production and to be important for perception (Gussenhoven 2004; House 1990; Ladd 2008). Furthermore, considerable evidence suggests that effects of abstract tonal structure on F0 are better conceived in terms of perceptual targets involving auditory pitch (Barnes et al. 2012; D’Imperio 2000). Jun hints at some problems with the ToBI\* framework’s handling of accounting for F0 turning points and facts about the importance of pitch for phonology, a topic we explore here.
- *Tones have paradigmatic phonological status:* A core proposal of both Goldsmith (1976) and Pierrehumbert (1980) was that tones have paradigmatic phonological status, meaning that they are defined relative to the speaker’s pitch range. This proposal is supported by the observation that in lexical tone languages, perceivers can recognize the tone of a single-syllable word spoken in isolation with a level tone (Lee 2009; Peng et al. 2012). Relatedly, perceptual studies demonstrate that in intonation languages, listeners can discern the location of a syllable in a speaker’s pitch range with reasonably good accuracy (Bishop and Keating 2012; Honorof and Whalen 2005).

- *Starred tones of pitch accents form associations with syllables that have hierarchical metrical prominence:* The theory behind ToBI\* posited a notion of *starred tones*, that is, tones that participate in pitch accentuation by associating with metrically prominent syllables. The potential influence of the hierarchical organization of stress on tones that was first worked out in Liberman (1975) was not explored in Pierrehumbert (1980). However, the explanatory value of viewing stress as hierarchical and metrical survives to the present day.

We agree with Jun that these theoretical points capture important generalizations about tonal systems made possible by the invention of ToBI\*. However, in the next section, we argue that Goldsmith's (1976) and, later, Pierrehumbert's (1980) assumption that tones have (strictly) paradigmatic phonological status provided an incomplete assessment of phonological properties of tone. We identify this theoretical choice as the source of considerable, enduring problems with ToBI\*'s phonetic transparency and consistency.

### **Strictly Paradigmatic Phonological Representations Lead to Descriptive Inadequacy and Inconsistency in Traditional AM Theory and ToBI\***

Jun alludes to theoretical problems with ToBI\* by stating, "Some of [the] challenges [of ToBI\*'s handling of intonational phenomena] stem from properties of the AM theory that ToBI adopts" (section 4.5). Jun cites, without elaboration, the lack of phonetic transparency in ToBI\* to be one of its key problems. In this section, we trace ToBI\*'s problems with phonetic transparency and consistency to inadequate treatment of syntagmatic aspects of tonal phonological representations.

It is abundantly clear that both paradigmatic aspects as well as syntagmatic aspects of representations are important for tonal systems (Cutler, Dahan, and van Donselaar 1997; Ladd 2008; Lee 2009). *Syntagmatic properties*, which involve defining tone height in relation to adjacent tones rather than to a global referent, have long been thought to be central to tonal representations across languages (Cole 2015; Jakobson, Fant, and Halle 1952; Ladd 2008; O'Connor and Arnold 1973; Odden 1995). There is considerable evidence that cognitive representations of tonal information include syntagmatic relationships in lexical tone languages (Odden 1995; Wong and Diehl 2003), intonation languages (Dilley 2005; Dilley and Brown 2007), and nonlinguistic tonal systems, such as world musical traditions (Burns 1999; Dowling and Fujitani 1971; Monelle 2014; Patel 2010).

Both Goldsmith (1976) and Pierrehumbert (1980) acknowledged the importance of syntagmatic relationships for tonal representations, but they prioritized only the capture of paradigmatic aspects in phonology. We will show that the assumption of strictly paradigmatic features in phonology was highly problematic. Still, given that Goldsmith (1976) marked the birth of the idea of true tonal autonomy from segments, it was arguably not the time to explore the specific featural representations of tones themselves.<sup>2</sup> Indeed, no linguistic theoretic notational device had yet been developed that could yield conceptual insight into how tones themselves can be dually paradigmatic *and* syntagmatic. (AM<sup>+</sup> develops such a device; see the "A Way Forward" section.)

The choice of strictly paradigmatic tonal phonological representations was viewed as a simplifying assumption, but this assumption led to an overall theory of the "grammar" that was, in practice, not simple. To justify delving into the sequelae of this theoretic choice, especially the explanatory burden put on the "phonetic component" of

the grammar by assuming a very weak phonology, we cite Pierrehumbert and Beckman (1988, 4), who state, “the division of labor between the phonology and the phonetics is an empirical question, one which can only be decided by constructing complete models in which the role of both in describing the sound structure is made explicit.” As we will discuss, the theoretic assumption that phonology lacks syntagmatic phonological restrictions on tones led to the following: (i) complex phonetic rules for tone scaling, which did not, in the end, “work” to achieve desired restrictions on relative tone heights; (ii) inconsistencies in assumed mappings of pitch accentual tones to significant F0 events (peaks and valleys); and (iii) complications in when F0 events corresponded to interpolation functions versus phonological tones (accents or phrase tones). These problems have led to difficulties in using ToBI\* for prosodic typology (Hualde and Prieto 2016).

### Complex Phonetic Rules and Mechanisms for Tone Scaling That Didn’t Work

Accepting as true the a priori premise that tonal representations lack syntagmatic restrictions required complete redefinitions of what constitutes a “phonological representation” and what is “phonetic.” That is, given the a priori premise that the phonological component of the grammar encodes only paradigmatic aspects of tones, the logical consequence was the further assumption that the phonetic component of the grammar is home to syntagmatic restrictions on relative tone heights. There is abundant evidence that syntagmatic changes—being higher or lower than another tone—are meaningful, and until Pierrehumbert (1980), meaningful contrasts were considered to be part of phonology. Suddenly, the phonetic component of the grammar, which prior to that time had been taken to refer to, for example, biomechanical forces during speech production, was endowed with the power to make meaning-based distinctions.<sup>3</sup>

To supplement this “weak” phonology, it was necessary to invent a “strong” phonetics that consisted, in Jun’s words, of “rules that map the phonological representation (abstract level tone target sequences) to the phonetic representation (the F0 contour)” (section 4.2). These rules, comprising a complex set of equations laid out in an entire chapter of Pierrehumbert (i.e., Chapter 4), were the main mechanism in the “grammar” for scaling the relative F0 heights of tones, one to another. They entailed an assumption of an abstract tone reference line necessary for phonetic scaling of tones, together with a gradient parametric value (which was termed “prominence” but was equated with F0), along with abstruse parameters  $n$  and  $k$ , which lacked a phonetic interpretation. Pierrehumbert and Beckman (1988) later proposed a version of the phonetic module that dispensed entirely with the phonetic rules, instead proposing that paradigmatic tones were scaled with respect to both a high reference line and low reference line, as a function of a parameter again termed “prominence” but which was just a proxy for F0. A variety of other proposals were put forward that varied with respect to numbers of reference lines, whether reference lines were static or dynamically changed, and whether tones were assumed to be on reference lines or could vary freely with respect to the reference lines (e.g., Ladd 1986). In many cases, the reference lines were just a proxy mechanism for imposing syntagmatic (phonological) restrictions on relative tone heights, as in Liberman and Pierrehumbert (1984). These accounts ignored the issue of how listeners could perceptually recover phonological representations from F0, or else they sidestepped the issue by assigning meaning to the “phonetic” component rather than to phonology.

There was, furthermore, a serious problem with the phonetic rules in Pierrehumbert (1980): they did not actually restrict syntagmatic relative F0 heights of tones. As demonstrated in Dilley and Brown (2007, 545–548), the rules failed to successfully

restrict scaling of L and H tones so that specific claimed F0 contours would correspond to the intended tonal entities. For example, Dilley and Brown showed that even for bitonal accents such as L+H\* and L\*+H (uniformly assumed to entail rising contours), the rules permitted H tones to fall below the adjacent L tones, allowing L+H\* and L\*+H to map onto falling contours. The revised theory of Pierrehumbert and Beckman (1988) also suffered from the same serious problem, as further shown in Dilley and Brown (548–549), so that again, rising portions of L+H\* and L\*+H contours were permitted to map onto falling contours. Dilley and Brown showed that problems of this sort are not limited to these two accents, but are instead widespread throughout the accounts for tonal sequences of a variety of types.

### Inconsistencies in Mapping Pitch Accents to F0 Events

Numerous complications and inconsistencies in the pitch accent inventory can be traced to the piecemeal way in which syntagmatic restrictions were handled in Pierrehumbert (1980). The theoretical distinction between bitonal accents such as L+H\* and single-tone accents such as H\* was itself motivated in part as a means of capturing syntagmatic relations. Specifically, Pierrehumbert states, “A pitch accent *can impose a particular relationship between the f0 on the accented syllable and the immediately preceding or following f0 value*, independent of the existence of other accents. . . . In our theory, the bitonal accents [H\*+L, H+L\*, L\*+H, L+H\*] have this property and there are also two single tones [H\*, L\*] which do not” (31, our emphasis). Note that this treatment implicitly posited that relative heights of other tones in sequence (for example, L\* followed by H\*) were unconstrained in their relative heights by phonology, leaving a legacy of inconsistent treatment in ToBI\*'s notational conventions regarding which pairs of tones in a sequence code for syntagmatic relative tone heights, and which do not. As we already noted, the tone scaling rules did not actually restrict the syntagmatic relative heights of the two tones of bitonal pitch accents to surface with the intended F0 contours.

The piecemeal handling of syntagmatic restrictions further complicated the treatment of pitch accents through the adoption of descriptive devices termed “floating low” tones. For example, the H\*+L bitonal accent in Pierrehumbert was treated as exceptional, in that the +L tone was assumed to be “floating.” It was assumed to be never directly realized phonetically as a low F0 event, but instead was assumed to be the causal factor in an observed F0 peak (which is normally thought of as an index of an H tone) being relatively lower than another F0 peak in the same phrasal constituent.<sup>4</sup> This indirect floating-low device as a means of accounting for a syntagmatic relationship among observed high-pitched events was borrowed from mid-1970s African linguistics (and Goldsmith 1976), according to which lexical L tones were sometimes associated with and/or synchronically traceable to observed lowering of subsequent H-toned units, resulting in iterative phonetic lowering of the H-tone syllables in a claimed phenomenon termed “downstep.”

### Complications in When F0 Curves Correspond to Phonetic Interpolation versus Tones

The exceptional treatment of the L tone in H\*+L accents as a floating-low tone in Pierrehumbert (1980) in order to codify a syntagmatic relationship among high-toned events necessitated a further theory-internal complication regarding F0 interpolation contours. Building on a core assumption that phonological specification of tones is usually sparse in intonation languages, Pierrehumbert proposed that, in general, F0 interpolation functions that connect phonological tones are monotonic: increasing functions should only increase, not decrease, and decreasing functions should only

decrease, not increase. Because it was assumed that the L in an HLH sequence was a floating low that could never “surface” as an F0 valley, this precluded any description that treated the F0 valley as a low tone when the following peak was not lower than an earlier peak. The theoretic choice to prioritize the descriptive device of floating low from mid-1970s African linguistics over phonetic consistency meant that for an F0 peak-valley-peak sequence in which the two peaks were of equal height, the F0 valley could not be described as an L tone between the two H tones, based on these theory-internal assumptions. It was therefore necessary for Pierrehumbert (1980) to posit an exceptional nonmonotonic (“sagging”) interpolation function only in the case of two H tones, when the second H tone was not lower.

Evidence against this function was demonstrated in Ladd and Schepman (2003). Specifically, not only did the F0 valley in question show consistent alignment with respect to the phonological structures in utterances, but varying the temporal alignment of the low tone within a phrase changed listeners’ interpretation of that phrase. Both kinds of evidence were consistent with the F0 valley being a reflex of a low tone that was phonological in nature. To further complicate matters, Pierrehumbert (1980) assumed H tones sometimes were realized with a “late peak” on a nonprominent syllable following the accented syllable. This assumption constitutes a lesser-known exception to the monotonic interpolation rule, one not commented on by Pierrehumbert, and amounts to a second type of nonmonotonic function, termed a “bulging interpolation” by Dilley and Heffner (2013).

Moreover, examination of how phonological theories have handled cases of phonetically flat pitch reveals another case of how failing to codify syntagmatic relationships has complicated theories. Consider that *monotonic* can also mean “unchanging in pitch or tone”; a monotonic interpolation between two tones at the same level should yield a flat pitch, where a temporally later tone has an equal pitch relative to an earlier tone and to everything in between.<sup>5</sup> To account for regions of flat pitch, descriptive work on African languages in the 1970s (e.g., Goldsmith 1976; Hyman and Schuh 1974; Leben 1973; Williams 1971/1976) posited phonological rules that enacted tone copying or spread to account for regions of flat pitch, that is, cases where lexically specified H or L tone showed a sustained pitch at the same level over multiple syllables.<sup>6</sup> We note that tone spread is assumed to result in an F0 “elbow” that marks the right edge of the flat-pitched region before a subsequent rise or fall.<sup>7</sup>

An alternative to the tone-spreading rule proposed by Pierrehumbert and Beckman (1988) involved accounting for syntagmatically level stretches of F0 according to a phonological rule known as “secondary association,” in which a single tone could be anchored to two timing slots separately. This idea was productively used to account for level stretches in a variety of languages (Grice 1995; Grice, Ladd, and Arvaniti 2000; Prieto, D’Imperio, and Gili Fivela 2005). However, note that this proposal (and tone spreading) requires inconsistency in treatment of autosegmental association. That is, while the original idea of Goldsmith (1976) was that tones occupy a single timing slot (i.e., that they occur at a single moment in time), secondary association entails that tones can occupy multiple timing slots and “persist” over long stretches of time.

The tension among the tone spreading account, the secondary association account, tone copying, and/or a single tone per timing slot with monotonic interpolation has not to date been resolved. The core facts motivating these proposals, however, were strikingly similar. That is, cross-linguistically, there are many attested cases in which an equal height relationship exists among successive, adjacent tones, where change points may be separated by long distances.

### Summary

In conclusion, the assumption of strictly paradigmatic tonal phonological representations had a cascade of negative consequences for phonetic transparency and theoretic consistency. We point to the obfuscation of syntagmatic relationships as an underappreciated, but truly fundamental flaw in traditional AM theory and the ToBI\* enterprise. A span of forty years' time also reveals that a strictly paradigmatic phonological treatment is simply inadequate. In spite of the best efforts of Pierrehumbert (1980), Pierrehumbert and Beckman (1988), and others, a supplementary phonetic module has not been put forward that sufficiently constrains relative tone heights to generate the correct F0 curves from phonological tones. The legacy of this inadvertent obfuscation masquerading as theoretical simplification has indelibly imprinted in ToBI\*'s notational apparatus. Failure to clearly codify the relationship between syntagmatic aspects in the signal and abstract theoretical constructs means that ToBI's descriptive apparatus for these aspects of representations is highly inconsistent, unconstrained, and unprincipled.

Fortunately, ToBI\* systems have been used by communities of scholars as though syntagmatic tonal relationships are part of the phonology, even though they are not. For example, scholars have annotated L+H\* as a low valley plus a rising pitch, even though this choice is not supported by the underlying theories, as discussed. In the following section, we demonstrate that a simple theoretical change—to assume that syntagmatic features are directly part of the phonological representations of tones—allows building on the last forty years of insights in an “enhanced” AM framework.

### A Way Forward: AM<sup>+</sup> Theory and the RaP Transcription System

An “enhanced” AM theory (AM<sup>+</sup>) is proposed here. AM<sup>+</sup> integrates insights from more than forty years of empirical work in intonational phonology, as well as research in speech perception, music cognition, and cognitive neuroscience, building on the proposals of Dilley (2005). These proposals develop a notational device adapted for linguistic systems that is derived from insights about cognitive representations of nonlinguistic tonal information from auditory streaming studies, music cognition studies, and music theories for the world's musical systems (Bregman 1994; Burns 1999; Dowling and Fujitani 1971; Hannon and Trainor 2007; Jones, Fay and Popper 2010; Patel 2010).

A central part of the AM<sup>+</sup> theory is its assumption that, across languages, cognitive representations of tonal systems include both syntagmatic and paradigmatic aspects of tone. That is, they are both part of the phonology. Furthermore, paradigmatic and syntagmatic aspects of tone operate according to similar constraints such that they are specified lexically in some tonal systems and postlexically in others. It is proposed that each language draws on a combination of paradigmatic and syntagmatic tonal specifications, where there will be different densities of specification at the lexical or postlexical levels.<sup>8</sup>

Note that by including both paradigmatic and syntagmatic aspects of tone in phonology, AM<sup>+</sup> theory is not more complex than proposals of traditional AM theory, which assumed strictly paradigmatic tonal features (e.g., Goldsmith 1976). This is because, as we show for AM<sup>+</sup> theory, paradigmatic aspects of tone reduce to syntagmatic feature specifications. This is a core insight of AM<sup>+</sup> theory—namely, that paradigmatic tonal features can be formally reexpressed as syntagmatic ones. AM<sup>+</sup> thus preserves the economy of featural specification that was appealing in traditional AM theory. Furthermore, in nearly forty years, no theory based on paradigmatic level tones plus phonetic implementation rules has been put forward that successfully maps sequences of H and



L tones to their expected F0 outputs for intonation languages, as was conclusively shown in Dilley and Brown (2007).<sup>9</sup>

AM<sup>+</sup> conceives of tones in cognitive, abstract terms. In this theory, tones are abstract pitch targets that involve language-specific sensorimotor mappings. Conceiving of tones as abstract pitch targets that instantiate experience-dependent sensorimotor mappings is well grounded in empirical research from the past two decades in speech perception, music cognition, and cognitive neuroscience (Burnett et al. 1998; Chen et al. 2007; Guenther 2016; Guenther and Hickok 2015; Hutchins and Peretz 2011; Ning, Shih, and Loucks 2014; Patel et al. 2011; Pfordresher et al. 2015). To relate concepts of tones in traditional AM theory to AM<sup>+</sup>, note that an H tone, which in traditional AM theory was taken to correspond to an F0 peak (see Pierrehumbert 1980), can be fundamentally reexpressed as an abstract pitch target that is syntagmatically constrained to be higher in pitch than a tonal target to the left and to the right. Viewed in this way, a syllable that is autosegmentally associated with an H tone naturally maps in most speaking situations to an F0 peak. However, because tonal targets are intrinsically perceptual in nature, other F0 mappings are possible, such as F0 plateaus (Dilley and Brown 2007; Knight 2008) or variations in the F0 shape as given by, for example, the tonal center of gravity (Barnes et al. 2012; D'Imperio 2000; Niebuhr 2007a).

Likewise, an L tone that in traditional AM theory was taken to correspond to an F0 valley (e.g., Pierrehumbert 1980) can be reexpressed in AM<sup>+</sup> as an abstract pitch target that is constrained to be syntagmatically lower in pitch than a tonal target to the left and to the right. Finally, an L or H tone that in traditional AM theory was taken, for example, to correspond to the right edge of a stretch of level pitch, is reexpressed in AM<sup>+</sup> as an abstract pitch target that is constrained to be syntagmatically at the same pitch level as a tonal target to the left. The syntagmatic relationship between that L or H tone and the tone that follows will then dictate whether the contour subsequently rises or falls (i.e., a following tonal target that is higher or lower, respectively). Given that perceptual pitch lawfully relates to F0 in speech (d'Alessandro and Mertens 1995; House 1990; Mertens 2004), AM<sup>+</sup> provides a unifying explanation for observed correspondences between abstract tones and their typical F0 consequences, as with F0 peaks, valleys, and plateaus.<sup>10</sup>

An experiment from Dilley and Brown (2007) provides further support for the proposal that categorical differences in F0 turning point timing, for example, of F0 peaks and valleys, derive fundamentally from pitch targets whose cognitive representations involve syntagmatic specifications. Dilley and Brown created synthetic stimuli with flat, level-pitched F0 across critical syllables, without F0 peaks or valleys. Using an imitation task, the gold standard test of categories in intonation (Gussenhoven 2004; Pierrehumbert and Steele 1989), Dilley and Brown showed that speakers imitated the level-pitched syllables by producing categorical shifts in F0 peak and valley timing; furthermore, the categorical timing was predicted by the syntagmatic relationship of relative height borne by a level-pitched syllable to adjacent syllables, not by the syllable's relation to the pitch range. These findings further supported a view that F0 peaks and valleys derive from syntagmatic relationships among tones and provide experimental evidence for a tonal phonology that includes syntagmatic features.

The RaP prosodic transcription system (Breen et al. 2012; Dilley and Brown 2005) instantiates the proposals of AM<sup>+</sup> theory.<sup>11</sup> The phonological representations in AM<sup>+</sup> are based on two syntagmatic tone features: [+/- same], which distinguishes same and different, and [+/- higher], which distinguishes higher and lower. [+/- higher] is only specified in the case of [-same]. RaP includes the symbols **H**, **L**, and **E**, which capture the syntagmatic relationship borne by a tone,  $T_n$ , with respect to a previous tone,  $T_{n-1}$ ;

boldface type will be used for RaP symbols to distinguish them from ToBI\* notations (in this section, for MAE\_ToBI, in particular).<sup>12</sup> RaP's **H** designates a tone that has a feature specification [-same, +higher] and is phonetically higher than the previous tone. **L** designates a tone that has a feature specification [-same, -higher] and is phonetically lower than the previous tone. **E** designates a tone with a feature specification [+same], which is phonetically equal in pitch to the previous tone.<sup>13</sup>

In RaP, the features [+/- same] and [+/- higher] are specified for pairs of adjacent tones,  $T_{n-1}$  and  $T_n$ , on an AM\* grid tier. In other words, it is claimed that the cognitive representation for tone in languages entails phonological encoding of the relative heights of tones in a sequence. The most basic aspect of the representation is that a given tone,  $T_n$ , is specified to have a pitch value that is higher than, lower than, or equal to that of a prior tone in the sequence,  $T_{n-1}$ . An AM\* grid tier is a hybrid concept that generalizes across notions of a metrical grid row (e.g., Halle and Idsardi 1995) and an autosegmental tier à la Goldsmith (1976); it conceives of autosegmental association as expressly hierarchical, elaborating on the assumed, and eponymous, metrical representations of traditional AM theory. Furthermore, the notation  $T_n/T_{n-1}$  is adopted to represent a pair of adjacent tones on an AM\* grid tier that is constrained by a given syntagmatic feature; the entity on the right of the / is the referent entity. For example,  $T_n/T_{n-1}$  = [-same, +higher] means that  $T_n$  is higher than  $T_{n-1}$ . Phonetically, this corresponds to a rise. By extension, a reciprocal relationship exists between two tones captured through the relationality of this expression. A rise in forward time is just a fall in reverse time, which is captured by a sign change when the referent entity is in the future, for example,  $T_{n-1}/T_n$  = [-same, -higher]. In AM\* theory, this is termed the *reciprocal property*.<sup>14</sup>

*Paradigmatic features* have been traditionally characterized as “tone levels” according to which tones are defined relative to a speaker's pitch range. AM\* offers a formalization of this view according to which paradigmatic tone levels arise from a syntagmatic relationship between a tone, on the one hand, and an abstract (phonological) referent quantity, on the other, which is phonetically defined with respect to a speaker's own pitch range. Specifically, paradigmatic tonal representations are formally codified as a syntagmatic relationship between a lexically specified tone,  $T$ , and an abstract referent level,  $r$ ; the value  $r$  is phonetically interpreted as the speaker's mean pitch (or habitual pitch).<sup>15</sup> A high tone which is high in speakers' pitch ranges is represented as  $T/r$  = [-same, +higher], a low tone which is low in speakers' ranges is  $T/r$  = [-same, -higher], and a tone at speakers' mean or habitual pitch levels is  $T/r$  = [+same].<sup>16</sup> If a tone,  $T$ , is not specified in the lexicon to have a particular featural relationship with respect to  $r$ , then at the speech motor planning stage, we propose that the first tone in an utterance,  $T_1$ , receives postlexical assignment of features for  $T_1/r$ . Thereafter, lexically specified features for tones, together with postlexical expressive factors such as prominence and intended meaning, will determine the overall placement of tones in the speaker's pitch range and the syntagmatic pitch distances among tone pairs.

Importantly, paradigmatic representations specified according to a common referent have an interesting benefit: they allow obtaining syntagmatic relationships “for free” when tones are strung together by default in sequence.<sup>17</sup> For example, a language with two lexical tones,  $T_H$  for high tone and  $T_L$  for low tone, might specify that  $T_H/r$  = [-same, +higher] and  $T_L/r$  = [+same].<sup>18</sup> Because  $T_H$  is higher than  $r$  and  $T_L$  is at the same level as  $r$ , deductive reasoning ensures that by default,  $T_H$  will be higher than  $T_L$ . Language-specific rules might modify default syntagmatic relationships in ways that could be used to distinguish meanings (Odden 1995). This account appears to fit well in the case of Hausa,

where syntagmatic relative heights of H tones in HL sequences distinguish statements from questions (Inkelas and Leben 1990; Inkelas, Leben, and Cobler 1986).

These proposals entail that syntagmatic relationships can be derived from paradigmatic specifications, or syntagmatic relationships can even be separately specified in the lexicon or the morphosyntax, perhaps without a paradigmatic specification. These properties therefore endow this framework with the ability to more elegantly account for so-called floating-tone phenomena than was previously possible. Specifically, if a tone-bearing unit of a given tone is deleted, there is still another tone that retains, or holds onto, the syntagmatic relational featural specification. A new anchor point for this abstract relational structure can then be found, leading to realization on a different syllable the relative height differences taken to be phonetic hallmarks of floating tones, which are here viewed as reflexes of inherently syntagmatic tonal specifications.

Elaborating on Dillely (2005), five tonal levels can be captured in AM<sup>+</sup> by proposing the feature [+/- small]. This feature codifies tonal distance: [+small] represents a small tonal distance, while [-small] indicates a large tonal distance (Patel 2010; Vos and Troost 1989). We propose that [+/-small], like [+/-high], is specified only for [-same]. We further propose that in most cases, tones are unmarked for [+/-small], so pitch range can vary expressively. A language with five level tones—extra high, high, mid, low, extra low (EH, H, M, L, EL, respectively)—could thus be described as in table 4c.1.<sup>19</sup>

RaP and AM<sup>+</sup> theory elaborate productively on the relationship between hierarchical metrical structure and tonal associations. AM<sup>+</sup> and RaP adopt the starred \* tone notation used previously to describe tones that autosegmentally associate with a metrically prominent syllable. Metrically prominent syllables are marked in RaP with x (moderate prominence) or X (strong prominence), where the latter would occupy a higher grid tier position than the former. Importantly, AM<sup>+</sup> theory proposes that starred tones that associate with prominent metrical positions propagate upward to be represented in positions of adjacency on higher grid tiers. Following the idea of traditional metrical grid formalisms (e.g., Halle and Idsardi 1995; Hayes 1995), higher levels of AM<sup>+</sup> grid tiers entail adjacency of elements that occupy them. The significance of this is that on higher grid tiers, nonadjacent tones may be specified for syntagmatic featural relationships lexically or postlexically. This allows an account of tone register phenomena, for example, downstep, downdrift, and upstep (Clements and Goldsmith 1984; Hyman 1993; Inkelas and Leben 1990; Inkelas et al. 1986; Ladd 1988; Snider 1999; Truckenbrodt 2002).<sup>20</sup> A metrical account is consistent with a growing body of

**Table 4c.1**

Five level paradigmatic tone specifications derived from syntagmatic features [+/- same], [+/- higher], and [+/- small]

Lexical specification in phonology	Phonetic interpretation
$T_{EH}/r = [-\text{same}, +\text{higher}, -\text{small}]$	Substantially higher than the mean pitch; high in the pitch range
$T_H/r = [-\text{same}, +\text{higher}, +\text{small}]$	Slightly higher than the mean pitch
$T_M/r = [+same]$	Equal to the mean pitch
$T_L/r = [-\text{same}, -\text{higher}, +\text{small}]$	Slightly lower than the mean pitch
$T_{EL}/r = [-\text{same}, -\text{higher}, -\text{small}]$	Substantially lower than the mean pitch; low in the pitch range

evidence of metrical interactions in a variety of languages with very different tonal systems (de Lacy 2002; Hayes 1995; Manfredi 1993; Rice 1987; Zec 1999).

Phonetically, RaP requires a local phonetic pitch change (F0 turning point or F0 slope change) for a starred tone to be indicated on a metrically prominent syllable. This is consistent with prior assumptions of sparse representations, e.g., that only a subset of metrically prominent syllables are pitch accented (i.e., associated with a starred tone). As a consequence, a stretch of flat pitch can never have pitch accents, only metrical prominences, in contrast to Pierrehumbert's (1980) proposal that strings of low pitch accents can be present in regions of flat pitch. Based on these ideas, RaP distinguishes three categories of syllable prominence: metrically nonprominent, metrically prominent without a pitch accent (i.e., without a pitch change, such as for flat pitch), and metrically prominent with a pitch accent (i.e., with a pitch change). By contrast, the ToBI\* system allows for only two levels of prominence, pitch accent or no pitch accent, in contrast to multiple studies demonstrating that speakers produce, and listeners perceive, at least three levels of prominence (Fitzroy and Breen 2019; Greenberg, Carvey, and Hitchcock 2002).

There are several other notational conventions and standardizations that are instantiated in AM<sup>+</sup> and codified in RaP's conventions which enhance phonetic transparency and explanatory power relative to ToBI\*:

- *Strictly monotonic interpolation functions:* AM<sup>+</sup> retains the assumption that tones may be sparsely distributed in intonation languages and do not often occur on every syllable or word, such that adjacent tones separated by multiple syllables or words may be connected via interpolations. In this theory, interpolation functions are strictly monotonic, ensuring that all turning points are coded as tones. Multiple studies have demonstrated evidence against Pierrehumbert's (1980) proposal that certain F0 turning points are not tones but rather reflexes of exceptional nonmonotonic interpolation functions. (See Dilley and Heffner 2013; Dilley, Ladd, and Schepman 2005; Ladd and Schepman 2003.)
- *Tones and timing slots:* An aspect of AM<sup>+</sup> theory that notably increases phonetic transparency is that every tone must be associated with a timing slot. This effectively disallows floating tones and multiple associations between a single tone and more than one timing slot (tone spread or secondary association). Like ToBI\*, RaP allows multiple tones to be associated with a syllable.

The assumption that every tone is associated with a timing slot allows for the consistent treatment of unstarred tones in bitonal pitch accents. Pierrehumbert (1980) predicted a constant timing relationship between the two tones of bitonal pitch accents, but this prediction has not been borne out in production studies (Arvaniti, Ladd, and Mennen 1998, 2000; Dilley, Ladd, and Schepman 2005; Ladd 2008). Following AM<sup>+</sup>, RaP treats pitch accents as prominence-lending pitch movements, that is, locations of a local change in pitch. RaP assumes that unstarred tones can participate in pitch accents by associating with a nonprominent slot adjacent to a timing slot with a starred tone, or they can associate with constituent edges, following Pierrehumbert and Beckman (1988). The + symbol is used in RaP for the unstarred tones of pitch accents. Unstarred tones with + are assumed to associate directly with respect to a metrically nonprominent position; however, their eligibility to so associate is limited to the set of nonprominent positions that are adjacent to a prominent position associated with a starred tone. The + is put on the right side of the unstarred tone when that tone is to the left of a

starred prominent syllable (i.e., a metrically prominent syllable that is autosegmentally associated with a starred tone). For example, RaP's **L+ H\*** is annotated as a sequence of two tones with a space between them and maps to ToBI\*'s **L+ H\***. Otherwise, the + is put on the left side of the tone, as in RaP's **L\* +H**, which can be compared to ToBI\*'s **L\* +H**. Note that this formulation predicts that unstarred tones can be associated with metrically nonprominent positions both before and after a starred tone.

- *Meaningful pitch range differences:* The theory outlined here, which places primacy on syntagmatic tone features, readily accounts for examples of meaningful differences in pitch range. For example, RaP accounts for ToBI's much-studied distinction of **H\*** versus **L+ H\***; phonetically, there is a small rise to a peak for ToBI's **H\*** versus a large rise to a peak for ToBI's **L+ H\***.<sup>21</sup> These phonetic differences are important for capturing distinctions of focus (Breen et al. 2010; Féry and Krifka 2008; Katz and Selkirk 2011; Xu and Xu 2005). Noting that both contours rise to a peak, RaP captures the contours as **L+ !H\*** (for ToBI's **H\***) versus **L+ H\*** (for ToBI's **L+ H\***). This treatment has the further effect of filling a theoretical gap in Pierrehumbert's (1980) theory having to do with the status of **F0** values on phrase-initial unstressed syllables preceding an **H\***. When **H\*** was on a noninitial syllable in the phrase and there was no phrase-initial boundary tone (such as Pierrehumbert's %H), it was theoretically unclear how phrase-initial unstressed syllables could obtain an **F0** value under the theory, because there was no phrase-initial tone prior to the **H\*** with respect to which to carry out **F0** interpolation (Dilley 2010). RaP assumes that every phrase starts and ends with a tone, thus overcoming this theoretical gap.
- *Phrase edges:* Regarding phrase-initial tones, a further point is warranted about RaP notation. Recall that in RaP, the codes **H**, **L**, and **E** describe the syntagmatic relationship between the later-occurring tone and the earlier-occurring one. By definition, a phrase-initial tone has no earlier-occurring tone in the same phrase; rather, its phonological status as high or low is fully determined by the following tone (if there is no paradigmatic lexical specification, that is). The phrase-initial tone thus redundantly corresponds to the reciprocal (via the reciprocal property) of the tone in second position in the phrase. This status of the phrase-initial tone is designated by prepending the : symbol. That is to say, the three ways of beginning a phrase are a rise, symbolized :**L H** (omitting + and \*); a fall, symbolized :**H L**; or a level pitch, symbolized :**E E**.

Regarding phrase-final tones, RaP and AM<sup>+</sup> assume that right edges of constituents license a variable number of unstarred tones, depending on postlexical, language-specific rules. Thus, for example, final rises are treated as two unstarred tones (**H H %**) when a slope change is observed; otherwise, in the case of monotonic rise, only one tone (**H %**) is warranted. As a further illustration of how RaP characterizes meaningful pitch range differences, take the *calling contour*, which entails a stepping down by a small pitch interval from one level-pitched stretch to another, as in *An-na-belle!* RaP repurposes ToBI's ! symbol to indicate a small pitch interval ([+small]). RaP's transcription for the calling contour is therefore :**E\* E+ !L\* +E %**.

- *Slope changes:* RaP codifies vertices corresponding to a slope change as tones. This allows a principled means of accounting for phenomena like ToBI's **H- H%** in cases when a shallow rise transitions on the last syllable to a steep rise. RaP allows the notation > for the upper edge of the pitch range or < for the lower edge of the pitch

range. Pierrehumbert (1980) assumed uniformly that all intonation phrases end with two tones, a phrase accent and a boundary tone, that reduced phonetic transparency. RaP assumes that the number of edge tones may vary. This convention increases phonetic transparency because it is not necessary to assume that tones are present when there is no change in F0 slope. Note that this convention provides a principled account for the well-known problem of Greek prenuclear accents, which were problematic for ToBI\*; this is because these Greek prenuclear accents show evidence of a slope change as their phonetic correlate, with a characteristic F0 valley and peak on the syllables preceding and following the accented syllable, respectively. RaP characterizes Greek prenuclear accents as L+ !H\* +H, capturing the observed slope change (Arvaniti, Ladd, and Mennen 1998, 2000).

- *Sparse tonal representation*: Consistent with a sparse tonal representation, adjacent syntagmatic features are required to have different featural specifications. As a result, for example, when two rising intervals—[–same, +higher]—are adjacent to one another, one of them must be [–small] and the other [+small]. Phonologically, adjacent syntagmatic features of [+same] are thus banned for T<sub>1</sub> T<sub>2</sub> T<sub>3</sub>. Phonetically, this corresponds to a slope change, with a tone—starred or unstarred—indicated at the locus of the slope change. As a consequence of these assumptions, there are no sequences like E\* +E, E\* E+ or E\* E, meaning that Pierrehumbert’s (1980) assertion that strings of L\* accents may give rise to a low, flat pitch is not supported in the present theory.

## Conclusion

AM<sup>+</sup> offers a simplified theory that accounts for a range of cross-linguistic tonal phenomena. Its approach is implemented with the RaP annotation system, which offers a phonetically transparent alternative to ToBI\*. This phonetic transparency makes RaP a useful starting point for developing a phonologically principled International Prosodic Alphabet (IPrA) (Hualde and Prieto 2016).<sup>22</sup> RaP has been implemented as a full annotation system, with a publicly available set of interactive training materials<sup>23</sup> and a corpus of RaP-labeled speech (Breen et al. 2018). A large-scale study comparing annotation agreement between labelers trained in both the RaP and ToBI systems demonstrated RaP agreement levels that were equal to, and in some cases exceeded, agreement levels for ToBI (Breen et al. 2012). Finally, recent studies have successfully used the RaP system to accurately assess prosodic structure (Sharpe, Fogerty, and van Ouden 2017).

AM<sup>+</sup> and RaP have already yielded new insights into metrical interactions between segmental and suprasegmental structures. Pierrehumbert (2000) noted that ToBI\* failed to capture observed restrictions on the sequencing of tones—for example, the observation that tones tend to be repeated. Using the approach outlined in AM<sup>+</sup> theory, Dilley and colleagues have experimentally demonstrated powerful perceptual constraints on metrical structure that can perceptually “garden-path” listeners into hearing different organizations of words (Breen et al. 2014; Dilley, Mattys, and Vinke 2010; Dilley and McAuley 2008; Morrill, Dilley, and McAuley 2014; Morrill et al. 2014).

In sum, in this chapter, Jun presents a useful outline for prosody researchers who are unfamiliar with ToBI\*, but it only skimmed the surface with respect to problems behind ToBI\*. Here, we have summarized several problematic aspects of the traditional AM theory that are irreconcilably part of ToBI\*’s notations. We feel this is a critical juncture in time that will determine the usefulness of transcription choices to fields that stand to gain the most from consistent prosodic annotation. AM<sup>+</sup> is a theory that retains the insights of traditional AM approaches that have stood the test of time, while

affording new insights and considerable improvement in phonetic transparency. RaP and AM\* are informed by more than forty years of research in phonetics, phonology, music cognition, and cognitive neuroscience. We hope that researchers will embrace paradigm change by moving toward AM\* and a phonetically transparent system like RaP, in the interests of fostering further discovery in prosody research.

## Notes

1. However, see Yi Xu and colleagues (Xu and Wang 2001; Xu and Xu 2005) for an opposing point of view.
2. Goldsmith's (1976) specific proposal was that H and L tones were based on features of [+/- high] and [+/- low]; H was [+high, -low], L was [-high, +low], and M was [+high, +low].
3. Following an a priori premise through to absurd illogical conclusions is an important part of philosophical scholarly enterprise. Philosophers routinely engage in thought experiments that involve alternative conceptualizations of reality that can lead to new insights (e.g., the Chinese room argument; Searle 1980). There is value in the philosophical traditions in the humanities by their permitting deeper understanding of what *is* by entertaining what *is not*. However, if applications of a priori reasoning result in unrealistic ideas about speech perception and production, the result will be to disconnect scholarly linguistic enterprises from science. The fact that unscientific approaches are common in linguistics (de Lacy 2014; Gibson and Fedorenko 2010) reflects well-known tensions between science- and humanities-oriented scholars who bump elbows in many linguistics departments.
4. The H+L\* accent was later quietly “rescinded” from the English inventory by grouping it together with H\* in MAE-ToBI and explicitly marking the lowering of a H\* that follows another H\* as a downstepped !H\*. The idea of floating low tones being responsible for lowering of H—rather than some direct syntagmatic relationship—has never been retracted.
5. Mathematically, if  $f: X \rightarrow Y$  is a set function from a collection of sets  $X$  to an ordered set  $Y$ , then  $f$  is said to be monotone if whenever  $A \subseteq B$  as elements of  $X$ ,  $f(A) \leq f(B)$ .
6. The obligatory contour principle (OCP), originally proposed by Leben (1973) to account for syntagmatic level pitch observed for sequences of paradigmatically specific lexical high tones, is described in AM\* as the default assignment of the syntagmatic feature [+same] to sequences of lexically specified tones sharing a common paradigmatic tonal referent,  $r$ . Given that the OCP descriptively captures widespread phonological perceptual phenomena across languages (Berent, Shimron, and Vaknin 2001; Coetzee 2005; McCarthy 1986), we speculate that the OCP and the feature [+/- same] both reflect cognitive processes of attention and memory consolidation for processing sensory information that is the *same* versus *different* (Jones 1976; Large and Jones 1999).
7. Pitch targets associated with F0 elbows at right edges of flat stretches of pitch marking transition points to a rise or a fall are simply annotated E in RaP.
8. For example, in Mandarin, nearly every syllable is lexically specified for tone (Xu and Wang 2001), except for neutral tone syllables (Chen and Xu 2006; Lai and Dillely 2016). By contrast, in Chichewa, a Bantu language spoken mainly in Malawi, paradigmatic tonal specification appears to be much sparser (Myers 1998).
9. Dillely and Brown (2007) pointed out that the precise fix to the phonetic module proposed in Pierrehumbert and Beckman (1988) is to specify a rule that H tones cannot fall below adjacent L tones. Dillely and Brown argue that this rule-based systematicity is better captured by revisiting the idea of syntagmatic specifications as part of phonological representations.

10. We propose that pitch targets are associated with timing slots that specify locations of change in the velocity of pitch change over time. Viewed in this way, pitch targets encode the points of pitch *acceleration* as temporally coordinated with the metrical structure of speech utterances. The acoustic consequences of these pitch targets are predicted to roughly correspond to the second derivative of a function  $y = f(x)$ —that is  $f''(x)$ —where  $x$  is time and  $y$  is the F0 value. Then cases when  $f'(x) = 0$  for a smoothed or stylized F0 contour correspond to F0 maxima and minima, whereas cases where  $f''(x) = 0$  correspond to other slope changes, including elbows and changes in the steepness of a rise or fall, as approximated via the juncture point of two piecewise-linear F0 functions. An example of such a slope change is the +!H\* tonal target in ToBI\*'s H + !H\* or a H- tonal target characterized as a slope change in Pierrehumbert's (1980) and ToBI\*'s H-H%.

11. See Breen et al. (2012) for a conversion from MAE\_ToBI to RaP. Converting from RaP to MAE\_ToBI entails information loss.

12. RaP's E option appears to provide a solution to accounting for an interesting accentual distinction reported recently by Niebuhr and Hoekstra (2015) for North Frisian involving pointed versus plateau-shaped pitch accents.

13. The AM\* theory can be viewed as validating or mirroring the phonetically transparent INTSINT (International Transcription System for Intonation) approach by Daniel Hirst (Hirst and Di Cristo 1998), providing a theoretical framework that supports its insights.

14. Research within the AM framework in intonational phonology has sometimes oversimplified the debates regarding the nature of intonational representations and prematurely rejected syntagmatic representations (see, e.g., Arvaniti, Ladd, and Mennen 1998, 23). Although rise/fall approaches (e.g., 't Hart, Collie, and Cohen 1990) did not correct predict timing aspects of F0 curves (Arvaniti, Ladd, and Mennen 1998; Ladd 2008), these approaches did not suffer from the serious problems of underdetermined F0 contour shape from phonological primitives, which exists with the theories of Pierrehumbert (1980) and Pierrehumbert and Beckman (1988), as pointed out in Dilley and Brown (2007).

15. In Dilley (2005), the tonal referent was symbolized  $\mu$ , the mathematical symbol for mean. We rename it  $r$  here to avoid confusion with moras.

16. In musical scale systems, each scale tone is defined by a relationship to a common paradigmatic referent pitch—termed the *tonic* in Western tonal music. This allows the notes to be “scaled” up or down—or instruments to be “tuned” up or down—creating completely different sets of absolute frequencies in hertz, while still allowing the melody (i.e., sequence of syntagmatically related pitches) to be recognized.

17. Indeed, musical scale systems have this property because they involve a common referent note. In Western music, the first scale note is called the tonic, and this note also names the key. For example, in the key of C, C is the tonic. When keys are changed in scale systems, the presence of a common referent note allows listeners to perceive the melody as constant even though the frequencies are shifted up or down (Dowling and Harwood 1986; Jones, Fay, and Popper 2010). In cases when there is no common referent note, as in atonal music, the pattern of ups and downs is the basis of the cognitive representation of the pitch sequence (Dowling and Fujitani 1971; Dowling and Harwood 1986), consistent with the primacy of syntagmatic features in the present theory. Note that musical intervals are perceived categorically (Siegel and Siegel 1977), as are lexical tones (Hallé, Chang, and Best 2004); thus, frequency ratios do not need to be exact to instantiate a tonal category. (See also Pfordresher et al. 2015.)

18. For example, Myers (1998) states that in Chichewa, a string of morphemes with low tone “is always realized unchanged with all low tones.” He describes low tone as “phonologically inert, because it is simply the absence of tone” (367). This leads him to propose



that low tone is underspecified in the surface representation. This description is consistent with a lexical paradigmatic specification that low tone in Chichewa is at the same level as  $r$ ,  $T_L / r = [+same]$ , which is the speaker's habitual (or mean) pitch phonetically.

19. There is considerable evidence that world musical systems are based on frequency ratios (Burns 1999; Perlman and Krumhansl 1996; Wright 2009). Dilley (2005) outlines an elaboration of the ideas presented here to account for how specific frequency ratios, or ranges of ratios, could become lexicalized. Fundamentally, we assume that F0 is a low-bandwidth channel in an information-theoretic sense (Shannon 1948), limiting the number of paradigmatic tonal contrasts that can be transmitted through it. In the speech-to-song transformation, a phrase that is repeated several times shifts perceptually to being heard as sung (Falk, Rathcke, and Dalla Bella 2014; Tierney, Patel, and Breen 2018). This phenomenon supports the contention that the pitches in speech are subject to similar cognitive organizing principles as in music.

20. Syntagmatic featural specifications at higher  $AM^+$  grid tiers constrain the possible steps among tones on lower grid tiers. Tones that propagate to higher grid tiers are the more "important" ones; though they are nonadjacent in time, they are heard to form a cohesive syntagmatic structure. For example, in J. S. Bach's Toccata and Fugue in D Minor, the solo unaccompanied allegro passage involves an alternation between low and high notes; the low notes in this passage are metrically prominent and are heard to form a coherent melody, even though they are nonadjacent in the note sequence. Dilley (2005) proposes the multiplicative property to capture how the syntagmatic relationships among tones on different  $AM^+$  grid tiers are constrained relative to one another. Generalizing Dilley's formulation, this property says that for two tones,  $T_n$  and  $T_N$  for  $n, n+1, \dots, N$  which are adjacent on a higher grid tier,  $M+1$ , the syntagmatic features of tones subtended by  $T_n$  and  $T_N$  constrains the sequence of steps at the next lower grid tier,  $M$ , such that  $T_N / T_n$  must equal  $(T_{n+1} / T_n) \cdot (T_{n+2} / T_{n+1}) \cdot \dots \cdot (T_N / T_{N-1})$ . The notation conveys abstract relationships, but it also has a direct mathematical interpretation, because frequency ratios in music are multiplicative (Wright 2009). Note that frequency ratios do not need to be exact to be heard as instances of a musical category (Siegel and Siegel 1977).

21. Other phonetic differences have sometimes been found between accents realizing focus differences, such as a later peak for  $L+H^*$ . RaP would capture an audibly late peak that occurred within the accented syllable as an extra unstarred tone, leading to a variant contour:  $L+H^*!H(+)$  . . .

22. See also Hirst (chapter 3, this volume) and Hirst and Di Cristo (1998).

23. Available at [http://tedlab.mit.edu/tedlab\\_website/RaPHome.html](http://tedlab.mit.edu/tedlab_website/RaPHome.html).

## References

- Arvaniti, A., D. R. Ladd, and I. Mennen. 1998. "Stability of Tonal Alignment: The Case of Greek Prenuclear Accents." *Journal of Phonetics* 26:3–25.
- Arvaniti, A., D. R. Ladd, and I. Mennen. 2000. "What Is a Starred Tone? Evidence from Greek." In *Papers in Laboratory Phonology V*, edited by M. B. Broe and J. B. Pierrehumbert, 119–130. Cambridge: Cambridge University Press.
- Barnes, J., N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel. 2012. "Tonal Center of Gravity: A Global Approach to Tonal Implementation in a Level-Based Intonational Phonology." *Laboratory Phonology* 3:337–383.
- Beckman, M., and J. Hirschberg. 1994. "The ToBI Annotation Conventions." Ohio State University, Columbus, Ohio. <https://web.archive.org/web/20130218081131/http://www.ling.ohio-state.edu/~tobi/>.

- Beckman, M., J. Hirschberg, and S. Shattuck-Hufnagel. 2005. "The Original ToBI System and the Evolution of the ToBI Framework." In *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun, 9–54. New York: Oxford University Press.
- Berent, I., J. Shimron, and V. Vaknin. 2001. "Phonological Constraints on Reading: Evidence from the Obligatory Contour Principle." *Journal of Memory and Language* 44:644–665.
- Bishop, J., and P. A. Keating. 2012. "Perception of Pitch Location within a Speaker's Range: Fundamental Frequency, Voice Quality, and Speaker Sex." *Journal of the Acoustical Society of America* 132:1100–1112.
- Breen, M., L. C. Dilley, M. Brown, and E. Gibson. 2018. Rhythm and Pitch (RaP) Corpus. LDC2018S04. Philadelphia: Linguistic Data Consortium.
- Breen, M., L. C. Dilley, J. Kraemer, and E. Gibson. 2012. "Inter-Transcriber Reliability for Two Systems of Prosodic Annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch)." *Corpus Linguistics and Linguistic Theory* 8:277–312. <https://doi.org/10.1515/cllt-2012-0011>.
- Breen, M., L. C. Dilley, J. D. McAuley, and L. D. Sanders. 2014. "Auditory Evoked Potentials Reveal Early Perceptual Effects of Distal Prosody on Speech Segmentation." *Language, Cognition and Neuroscience* 29:1132–1146. <https://doi.org/10.1080/23273798.2014.894642>.
- Breen, M., E. Fedorenko, M. Wagner, and E. Gibson. 2010. "Acoustic Correlates of Information Structure." *Language and Cognitive Processes* 25:1044–1098.
- Bregman, A. S. 1994. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.
- Burnett, T. A., M. B. Freedland, C. R. Larson, and T. C. Hain. 1998. "Voice F0 Responses to Manipulations in Pitch Feedback." *Journal of the Acoustical Society of America* 103:3153–3161.
- Burns, E. M. 1999. "Intervals, Scales, and Tuning." In *The Psychology of Music*, 2nd ed., edited by D. Deutsch, 215–264. San Diego: Academic Press.
- Chen, S., H. Liu, Y. Xu, and C. R. Larson. 2007. "Voice F0 Responses to Pitch-Shifted Voice Feedback during English Speech." *Journal of the Acoustical Society of America* 121:1157–1163.
- Chen, Y., and Y. Xu. 2006. "Production of Weak Elements in Speech: Evidence from F0 Patterns of Neutral Tone in Standard Chinese." *Phonetica* 63:47–75.
- Clements, G. N., and J. Goldsmith, eds. 1984. *Autosegmental Studies in Bantu Tone*. Dordrecht, the Netherlands: Foris.
- Coetzee, A. 2005. "The Obligatory Contour Principle in the Perception of English." In *Prosodies: With Special Reference to Iberian Languages*, edited by S. Frota, M. Vigário, and M. J. Freitas, 223–246. Berlin: Walter de Gruyter.
- Cole, J. 2015. "Prosody in Context: A Review." *Language, Cognition and Neuroscience* 30:1–31.
- Cutler, A., D. Dahan, and W. van Donselaar. 1997. "Prosody in the Comprehension of Spoken Language: A Literature Review." *Language and Speech* 40:141–201.
- d'Alessandro, C., and P. Mertens. 1995. "Automatic Pitch Contour Stylization Using a Model of Tonal Perception." *Computer Speech and Language* 9:257–288.
- D'Imperio, M. 2000. "The Role of Perception in Defining Tonal Targets and Their Alignment." PhD diss., Ohio State University.
- D'Imperio, M., B. Gili Fivela, and O. Niebuhr. 2010. "Alignment Perception of High Intonational Plateaux in Italian and German." Paper presented at the International Conference on Speech Prosody, Chicago, IL, USA, May.

- de Lacy, P. 2002. "The Interaction of Tone and Stress in Optimality Theory." *Phonology* 19:1–32.
- de Lacy, P. 2014. Evaluating Evidence for Stress Systems." In *Word Stress: Theoretical and Typological Issues*, edited by H. v. d. Hulst, 149–193. Cambridge: Cambridge University Press.
- del Giudice, A., R. Shosted, K. Davidson, M. Salihie, and A. Arvaniti. 2007. "Comparing Methods for Locating Pitch `Elbows.'" In *Proceedings of the Sixteenth International Congress of Phonetic Sciences*, <http://www.icphs2007.de/conference/Papers/1283/index.html>.
- Dilley, L. C. 2005. "The Phonetics and Phonology of Tonal Systems." PhD diss., MIT.
- Dilley, L. C. 2010. "Pitch Range Variation in English Tonal Contrasts Is Continuous, Not Categorical." *Phonetica* 67:63–81.
- Dilley, L. C., and M. Brown. 2005. *The RaP (Rhythm and Pitch) Labeling System, Version 1.0*. East Lansing: Michigan State University. <http://speechlab.cas.msu.edu/rap-system.htm>.
- Dilley, L. C., and M. Brown. 2007. "Effects of Pitch Range Variation on F0 Extrema in an Imitation Task." *Journal of Phonetics* 35:523–551.
- Dilley, L. C., and C. Heffner. 2013. "The Role of f0 Alignment in Distinguishing Intonation Categories: Evidence from American English." *Journal of Speech Sciences* 3:3–67.
- Dilley, L. C., and J. D. McAuley. 2008. Distal Prosodic Context Affects Word Segmentation and Lexical Processing." *Journal of Memory and Language* 59:294–311. <https://doi.org/10.1016/j.jml.2008.06.006>.
- Dilley, L. C., D. R. Ladd, and A. Schepman. 2005. "Alignment of L and H in Bitonal Pitch Accents: Testing Two Hypotheses." *Journal of Phonetics* 33:115–119.
- Dilley, L. C., S. Mattys, and L. Vinke. 2010. "Potent Prosody: Comparing the Effects of Distal Prosody, Proximal Prosody, and Semantic Context on Word Segmentation." *Journal of Memory and Language* 63:274–294. <https://doi.org/10.1016/j.jml.2010.06.003>.
- Dowling, W. J., and D. S. Fujitani. 1971. "Contour, Interval, and Pitch Recognition in Memory for Melodies." *Journal of the Acoustical Society of America* 49:524–531.
- Dowling, W. J., and D. L. Harwood. 1986. *Music Cognition*. Orlando: Academic Press.
- Falk, S., T. Rathcke, and S. Dalla Bella. 2014. "When Speech Sounds like Music." *Journal of Experimental Psychology: Human Perception and Performance* 40:1491–1506.
- Féry, C., and M. Krifka. 2008. "Information Structure: Notional Distinctions, Ways of Expression." In *Unity and Diversity of Languages*, edited by P. v. Sterkenburg, 123–136. Amsterdam: John Benjamin.
- Fitzroy, A. B., and M. Breen. 2019 "Metric Structure and Rhyme Predictability Modulate Speech Intensity During Child-Directed and Read-Along Productions of Children's Literature." *Language and Speech* 63:292–305. <https://doi.org/10.1177/0023830919843158>.
- Gibson, E., and E. Fedorenko. 2010. "Weak Quantitative Standards in Linguistics Research." *Trends in Cognitive Science* 14:233–234.
- Goldsmith, J. 1976. "Autosegmental Phonology." PhD diss., MIT.
- Greenberg, S., H. Carvey, and L. Hitchcock. 2002. "The Relationship between Stress Accent and Pronunciation Variation in Spontaneous American English Discourse." In *Proceedings of the ISCA Workshop on Prosody and Speech Processing*. [https://www.isca-speech.org/archive\\_open/prosody\\_2001/prsr\\_009.html](https://www.isca-speech.org/archive_open/prosody_2001/prsr_009.html).
- Grice, M. 1995. "Leading Tones and Downstep in English." *Phonology* 12:183–233.

- Grice, M., D. R. Ladd, and A. Arvaniti. 2000. "On the Place of Phrase Accents in Intonational Phonology." *Phonology* 17:143–186.
- Guenther, F. H. 2016. *Neural Control of Speech*. Cambridge, MA: MIT Press.
- Guenther, F. H., and G. Hickok. 2015. "Role of the Auditory System in Speech Production." In *Handbook of Clinical Neurology*, edited by M. J. Aminoff, F. Boller, and D. Swaab, 161–175. Amsterdam: Elsevier.
- Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Halle, M., and W. Idsardi. 1995. "General Properties of Stress and Metrical Structure." In *The Handbook of Phonological Theory*, edited by J. A. Goldsmith, 403–441. Oxford: Blackwell.
- Hallé, P. A., Y.-C. Chang, and C. T. Best. 2004. "Identification and Discrimination of Mandarin Chinese Tones by Mandarin Chinese vs. French Listeners." *Journal of Phonetics* 32 (3): 395–421.
- Hannon, E. E., and L. J. Trainor. 2007. "Music Acquisition: Effects of Enculturation and Formal Training on Development." *Trends in Cognitive Science* 11:466–472.
- Hayes, B. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
- Hirst, D., and A. Di Cristo. 1998. *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.
- Honorof, D. N., and D. H. Whalen. 2005. "Perception of Pitch Location within a Speaker's F0 range." *Journal of the Acoustical Society of America* 117:2193–2200.
- House, D. 1990. *Tonal Perception in Speech*. Lund, Sweden: Lund University Press.
- Hualde, J. H., and P. Prieto. 2016. "Towards an International Prosodic Alphabet (IPrA)." *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7:1–25.
- Hutchins, S., and I. Peretz. 2011. "Perception and Action in Singing." In *Progress in Brain Research*, edited by A. M. Green, C. E. Chapman, J. F. Kalaska, and F. Lepore, 103–118. Amsterdam: Elsevier.
- Hyman, L. H., and R. G. Schuh. 1974. "Universals of Tone Rules: Evidence from West Africa." *Linguistic Inquiry* 5:81–115.
- Hyman, L. M. 1993. "Register Tones and Tonal Geometry." In *The Phonology of Tone: The Representation of Tonal Register*, edited by H. van der Hulst and K. Snider, 75–108. Berlin: Mouton de Gruyter.
- Inkelas, S., and W. R. Leben. 1990. "Where Phonology and Phonetics Intersect: The Case of Hausa Intonation." In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, edited by J. Kingston and M. Beckman, 17–34. Cambridge: Cambridge University Press.
- Inkelas, S., W. R. Leben, and M. Cobler. 1986. "The Phonology of Intonation in Hausa." Paper presented at the Sixteenth Annual Meeting of the North East Linguistic Society, Amherst, MA, October.
- Jakobson, R., G. Fant, and M. Halle. 1952. *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.
- Jones, M. R. 1976. "Time, Our Lost Dimension: Toward a New Theory of Perception, Attention, and Memory." *Psychological Review* 83:323–355. <https://doi.org/10.1037/0033-295X.83.5.323>.

- Jones, M. R., R. Fay, and A. Popper, eds. 2010. *Music Perception*. New York: Springer.
- Katz, J., and E. Selkirk. 2011. "Contrastive Focus vs. Discourse—New: Evidence from Phonetic Prominence in English." *Language: Journal of the Linguistic Society of America* 87:771–816.
- 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* 51:223–244.
- Knight, R.-A., and F. Nolan. 2006. "The Effect of Pitch Span on Intonational Plateaux." *Journal of the International Phonetic Association* 36:21–38.
- Ladd, D. R. 1986. "Intonational Phrasing: The Case for Recursive Prosodic Structure." *Phonology Yearbook* 3:311–340.
- 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. 2008. *Intonational Phonology*. 2nd ed. Cambridge: Cambridge University Press.
- Ladd, D. R., and A. Schepman. 2003. "'Sagging Transitions' between High Accent Peaks in English: Experimental Evidence." *Journal of Phonetics* 31:81–112.
- Lai, W., and L. C. Dilley. 2016. "Cross-Linguistic Generalization of the Distal Rate Effect: Speech Rate in Context Affects Whether Listeners Hear a Function Word in Chinese Mandarin." *Proceedings of the International Conference on Speech Prosody* 8:1124–1128.
- Large, E. W., and M. R. Jones. 1999. "The Dynamics of Attending: How People Track Time-Varying Events." *Psychological Review* 106:19–159. <https://doi.org/10.1037/0033-295X.106.1.119>.
- Leben, W. R. 1973. "Suprasegmental Phonology." PhD diss., MIT.
- Lee, C.-Y. 2009. "Identifying Isolated, Multispeaker Mandarin Tones from Brief Acoustic Input: A Perceptual and Acoustic Study." *Journal of the Acoustical Society of America* 125:1125–1137.
- Lieberman, M. 1975. *The Intonation System of English*. PhD diss., MIT.
- Lieberman, M., and J. Pierrehumbert. 1984. "Intonational Invariance under Changes in Pitch Range and Length." In *Language Sound Structure*, edited by M. Aronoff and R. Oerhle, 157–233. Cambridge, MA: MIT Press.
- Manfredi, V. 1993. "Spreading and Downstep: Prosodic Government in Tone Languages." In *The Phonology of Tone: The Representation of Tonal Register*, edited by H. van der Hulst and K. Snider, 133–184. Berlin: Mouton de Gruyter.
- McCarthy, J. 1986. "OCP Effects: Gemination and Antigemination." *Linguistic Inquiry* 17:207–263.
- Mertens, P. 2004. "The Prosogram: Semi-Automatic Transcription of Prosody Based on a Tonal Perception Model." Paper presented at the International Conference on Speech Prosody, Nara, Japan, March.
- Monelle, R. 2014. *Linguistics and Semiotics in Music*. London: Routledge.
- Morrill, T., L. C. Dilley, and J. D. McAuley. 2014. "Prosodic Patterning in Distal Speech Context: Effects of List Intonation and f0 Downtrend on Perception of Proximal Prosodic Structure." *Journal of Phonetics* 46:68–85.

- Morrill, T., L. C. Dilley, J. D. McAuley, and M. A. Pitt. 2014. "Distal Rhythm Influences Whether or Not Listeners Hear a Word in Continuous Speech: Support for a Perceptual Grouping Hypothesis." *Cognition* 131:69–74. <https://doi.org/10.1016/j.cognition.2013.12.006>.
- Myers, S. 1998. "Surface Underspecification of Tone in Chichewa." *Phonology* 15:367–392.
- Niebuhr, O. 2007a. *Perzeption un kognitive Verarbeitung der Sprechmelodie: Theoretische Grundlagen und empirische Untersuchungen*. Berlin: Walter de Gruyter.
- Niebuhr, O. 2007b. "The Signalling of German Rising-Falling Intonation Categories: The Interplay of Synchronization, Shape, and Height." *Phonetica* 64:174–193.
- Niebuhr, O., and J. Hoekstra, J. 2015. "Pointed and Plateau-Shaped Pitch Accents in North Frisian." *Laboratory Phonology* 6:433–468.
- Ning, L. H., C. Shih, and T. M. Loucks. 2014. "Mandarin Tone Learning in L2 Adults: A Test of Perceptual and Sensorimotor Contributions." *Speech Communication* 63:55–69.
- O'Connor, R. J., and G. F. Arnold. 1973. *Intonation of Colloquial English*. Bristol, UK: Longman.
- Odden, D. 1995. "Tone: African Languages." In *The Handbook of Phonological Theory*, edited by J. Goldsmith, 444–475. Oxford: Blackwell.
- Patel, A. D. 2010. *Music, Language, and the Brain*. New York: Oxford University Press.
- Patel, R., C. Niziolek, K. Reilly, and F. H. Guenther. 2011. "Prosodic Adaptations to Pitch Perturbation in Running Speech." *Journal of Speech, Language, and Hearing Research* 54:1051–1059.
- Peng, G., C. Zhang, H.-Y. Zheng, J. Minnett, and, W. S.-Y. Wang. 2012. "The Effect of Intertalker Variations on Acoustic-Perceptual Mapping in Cantonese and Mandarin Tone Systems." *Journal of Speech, Language, and Hearing Research* 55:579–595.
- Perlman, M., and C. L. Krumhansl. 1996. "An Experimental Study of Internal Standards in Javanese and Western Musicians." *Music Perception* 14:95–116.
- Pfordresher, P. Q., S. M. Demorest, S. Dalla Bella, S. Hutchins, P. Loui, J. Rutkowski, and G. F. Welch. 2015. "Theoretical Perspectives on Singing Accuracy: An Introduction to the Special Issue on Singing Accuracy (Part 1)." *Music Perception: An Interdisciplinary Journal* 32:227–231.
- Pierrehumbert, J. 1980. "The Phonology and Phonetics of English Intonation." PhD diss., MIT.
- Pierrehumbert, J. 2000. "Tonal Elements and Their Alignment." In *Prosody: Theory and Experiment*, edited by M. Horne, 11–36. Dordrecht, the Netherlands: Kluwer.
- Pierrehumbert, J., and M. Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.
- Pierrehumbert, J., and S. A. Steele. 1989. "Categories of Tonal Alignment in English." *Phonetica* 46:181–196.
- Prieto, P., M. D'Imperio, and B. Gili Fivela. 2005. "Pitch Accent Alignment in Romance: Primary and Secondary Associations with Metrical Structure." *Language: Journal of the Linguistic Society of America* 48:359–396.
- Rice, K. D. 1987. "Metrical Structure in a Tone Language: The Foot in Slave (Athapaskan)." Paper presented at the Twenty-Third Annual Regional Meeting of the Chicago Linguistics Society.
- Searle, J. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3:417–457.
- Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27:379–423, 623–656.

- Sharpe, V., D. Fogerty, and D.-B. van Ouden. 2017. "The Role of Fundamental Frequency and Temporal Envelope in Processing Sentences with Temporary Syntactic Ambiguities." *Language and Speech* 60:399–426.
- Siegel, J. A., and W. Siegel. 1977. "Categorical Perception of Tonal Intervals: Musicians Can't Tell *Sharp* from *Flat*." *Perception and Psychophysics* 21:399–407.
- Snider, K. 1999. *The Geometry and Features of Tone*. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington Publications in Linguistics.
- 't Hart, J., R. Collier, and A. Cohen. 1990. *A Perceptual Study of Intonation*. Cambridge: Cambridge University Press.
- Tierney, A., A. D. Patel, and M. Breen. 2018. "Repetition Enhances the Musicality of Speech and Tone Stimuli to Similar Degrees." *Music Perception: An Interdisciplinary Journal*. 35 (5): 573–578.
- Truckenbrodt, H. 2002. "Upstep and Embedded Register Levels." *Phonology* 19:77–120.
- Vos, P., and J. Troost. 1989. "Ascending and Descending Melodic Intervals: Statistical Findings and Their Perceptual Relevance." *Music Perception* 6:383–396.
- Welby, P. 2006. "French Intonational Structure: Evidence from Tonal Alignment." *Journal of Phonetics* 34:343–371.
- Williams, E. S. (1971) 1976. "Underlying Tone in Margi and Igbo." *Linguistic Inquiry* 7:463–484. Manuscript was written in 1971.
- Wong, P. C. M., and R. L. Diehl. 2003. "Perceptual Normalization for Inter- and Intra-Talker Variation in Cantonese Level Tones." *Journal of Speech, Language, and Hearing Research* 46:413–421.
- Wright, D. 2009. *Mathematics and Music*. Providence, RI: American Mathematical Society.
- Xu, Y., and C. Xu. 2005. "Phonetic Realization of Focus in English Declarative Intonation." *Journal of Phonetics* 33:159–197.
- Xu, Y., and W. E. Wang. 2001. "Pitch Targets and Their Realization: Evidence from Mandarin Chinese." *Speech Communication* 33:319–337.
- Zec, D. 1999. "Footed Tones and Tonal Feet: Rhythmic Constituency in a Pitch-Accent Language." *Phonology* 16:225–271.

© 2022 The Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data is available.

Names: Barnes, Jonathan, 1970– editor. | Shattuck-Hufnagel, Stefanie, editor.

Title: Prosodic theory and practice / edited by Jonathan Barnes and Stefanie Shattuck-Hufnagel.

Description: Cambridge, Massachusetts : The MIT Press, 2022. | Includes bibliographical references and index.

Identifiers: LCCN 2021000764 | ISBN 9780262543170 (paperback)

Subjects: LCSH: Prosodic analysis (Linguistics)

Classification: LCC P224 .P739 2022 | DDC 414/.6—dc23

LC record available at <https://lcn.loc.gov/2021000764>