An *enhanced* autosegmental-metrical theory (AM^{+}) facilitates phonetically transparent prosodic annotation: A reply to Jun

Laura C. Dilley[1] and Mara Breen[2]

[1]Michigan State University [2]Mount Holyoke College

To appear in *Prosodic Theory and Practice*, J. Barnes and S. Shattuck-Hufnagel (eds.), Cambridge, MA: MIT Press.

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)

Research under a paradigm must be a particularly effective way of inducing paradigm change…Almost always the *men* who achieve these fundamental inventions of a new paradigm have been either very young or very new to the field whose paradigm they change.
– Thomas S. Kuhn (1962), *The Structure of Scientific Revolutions*, supplied emphasis

If I have seen further it is by standing on the shoulders of giants.
– Isaac Newton (1675), letter to Robert Hooke

We welcome this opportunity to respond to the well-organized, thoughtful chapter by Jun and to share our perspective on the ToBI* enterprise – where by ToBI* we mean all ToBI-like annotation systems, including MAE-ToBI, GToBI, etc. – and how this enterprise fits in with the scientific study of tone and intonation in language. Much time has now passed – some 40 years – since the core theoretic ideas behind ToBI* were put forward in groundbreaking, well-cited Ph.D. dissertations at Massachusetts Institute of Technology (MIT) by Goldsmith (1976), Pierrehumbert (1980), and Liberman (1975) – hereafter referred to as G76, P80, and L75, respectively – which formed the core ideas in what has come to be known as autosegmental-metrical (AM) theory (Ladd, 2008). Further, more than twenty-five years have passed since the original MAE-ToBI was developed (Beckman & Hirschberg, 1994; Beckman, Hirschberg, & Shattuck-Hufnagel, 2005); this development included the third (and most recent) ToBI workshop in Columbus, Ohio in 1993, which the first author of this chapter attended following her first undergraduate year at MIT. This long time span provides perspective on strengths and weaknesses of ToBI*, as well as the theory that underlies it.

Our chapter aims to contextualize Jun's chapter by highlighting theoretical insights from 40 years ago that led to the broad adoption of AM theory, an approach which has facilitated 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 phonology. Drawing on empirical evidence about the attested cognitive representations for pitch in the world's non-linguistic communicative tonal systems (i.e., music), we propose a theoretical clarification of syntagmatic elements in phonology, leading to a proposal for an *enhanced AM theory*, or *AM^{+}*. 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, & Mennen, 1998), (ii) influences of contour shape and slope on perception of phonological contrasts (Barnes, Veilleux, Brugos, & Shattuck-Hufnagel, 2012; D'Imperio, 2000; Niebuhr, 2007b) (iii) evidence against a non-monotonic interpolation function account of F0 turning points on metrically nonprominent syllables (Dilley & Heffner, 2013; Ladd & Schepman, 2003), (iv) the lack of invariant timing in bitonal pitch accents (Dilley, Ladd, & Schepman, 2005), (v) characterization of pointed vs. plateau-shaped pitch accents (Niebuhr & Hoekstra, 2015), and several others. We present the Rhythm and Pitch (RaP) prosodic transcription system as an AM^{+}-based, empirical tool which can be extended toward the goal of developing an International Prosodic Alphabet (IPrA) (Hualde & Prieto, 2016).

Enduring insights from 40+ years of traditional AM theory

Elaborating on Jun, we highlight below some key ideas and findings from the last 40+ years which 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 in Jun's chapter. This idea provided the basis for ToBI*'s descriptive notations, in which entities like H and L are viewed as discrete tonal events which have abstract associations with segments (e.g., Ladd, 2008).
- *Surface intonation contours reflect sparse tonal representations.* Another core idea highlighted in Jun's chapter was the idea that tones are sparse, e.g., they do not occur on every syllable and are connected via F0 interpolations.[1]
- *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 in Jun's chapter, 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 brief chapter 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, & Niebuhr, 2010; del Giudice, Shosted, Davidson, Salihie, & Arvaniti, 2007; Knight & 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). Further, 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's chapter hints at some problems with the ToBI* framework's handling of accounting for F0 turning points and facts about importance of pitch for phonology, a topic we explore below.
- *Tones have paradigmatic phonological status*. A core proposal of G76 and P80 was that tones have paradigmatic phonological status, meaning that they are defined relative to the speaker's pitch range. A core observation about how lexical tone languages work is that a single-syllable word can be spoken in isolation with a level tone, and perceivers can recognize the tone (Lee, 2009; Peng, Zhang, Zheng, Minnett, & Wang, 2012). 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 & Keating, 2012; Honorof & 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", i.e. tones which participated 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 L75 was not explored in P80. However, the explanatory value of viewing stress as hierarchical and metrical survive to the present day.

We agree with Jun that the above theoretical points capture important generalizations about tonal systems made possible by the invention of ToBI*. However, in the next section we argue that G76's, and later, P80's, 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 leads to descriptive inadequacy and inconsistency in traditional AM theory and ToBI*

Jun's chapter vaguely alludes to theoretical problems with ToBI*. She states: "some of [the] challenges [of ToBI*'s handling of intonational phenomena] stem from properties of the AM theory that ToBI adopts" (p. 33). Jun elsewhere cites the lack of phonetic transparency in ToBI* to be one of its key problems, without elaboration. 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, & van Donselaar, 1997; Ladd, 2008; Lee, 2009). Syntagmatic properties have long been thought to be central to tonal representations across languages (Cole, 2015; Jakobson, Fant, & Halle, 1952; Ladd, 2008; O'Connor & Arnold, 1973; Odden, 1995). There is considerable evidence that cognitive representations of tonal information include syntagmatic relationships in lexical tone languages (Odden, 1995; Wong & Diehl, 2003), intonation languages (Dilley & Brown, 2007), and non-linguistic tonal systems (cf. world musical traditions) (Burns, 1999; Dowling & Fujitani, 1971; Monelle, 2014; Patel, 2010).

Both G76 and P80 acknowledged the importance of syntagmatic relationships for tonal representations, but 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 G76 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.[2] Indeed, no linguistic theoretic notational device had yet been developed which could yield conceptual insight into how tones *themselves* can be dually paradigmatic *and* syntagmatic. (AM$^+$ develops such a device, as discussed later.)

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 anything but 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), hereafter PB88, who state (p. 4): "…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 detailed below, 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); (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 & 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 than or lower than another tone – are meaningful, and until P80, 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 e.g., biomechanical forces during speech production – was endowed with the power to make meaning-based distinctions.[3]

To supplement this "weak" phonology, it was necessary to invent a "strong" phonetics which consisted, in Jun's words, of "rules that map the phonological representation (abstract level tone target sequences) to the phonetic representation (the f0 contour)" (p. 6). These rules, comprising a complex set of equations laid out in an entire chapter of P80, 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 which was equated with F0), along with abstruse parameters *n* and *k*, which lacked a phonetic interpretation. PB88 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 F0. A variety of other proposals were put forward which 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 restrictions on relative tone heights, as in Liberman and Pierrehumbert (1984). These accounts steadfastly ignored the issue of how listeners could perceptually recover phonological representations from F0, or else obfuscated the issue by assigning meaning to the "phonetic" component, rather than to phonology.

There was, furthermore, a serious problem with the phonetic rules in P80: they did not actually restrict syntagmatic relative F0 heights of tones. As demonstrated in Dilley and Brown (2007, pp. 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 show that even for bitonal accents like L+H* and L*+H – uniformly assumed to correspond to rising contours – the rules permitted H tones to fall below adjacent L tones, allowing L+H* and L*+H to map onto *falling* contours. The revised theory of PB88 also suffered from the same serious problem, as further shown in Dilley and Brown (p. 548-9), so that again, rising L+H* and L*+H contours were permitted to map onto falling contours. Dilley and Brown show 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 P80. The theoretical distinction between bitonal accents like L+H* and single-tone accents like H* was itself motivated in part as a means of capturing syntagmatic relations. Specifically, P80 (p. 4, supplied emphasis) 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". 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 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 adoption of descriptive devices termed "floating low" tones. That is, the H*+L bitonal accent in P80 was treated as exceptional, in that the +L tone was assumed to be "floating". It was never directly realized as a low F0 event, but instead was attributed to be the causal factor in an observed F0 *peak* (which is normally thought of as an index of a H tone) being *relatively lower than another F0 peak* in the same phrasal constituent.[4] This indirect "floating low" device as a means of accounting for a syntagmatic relationship among observed high-pitched events was borrowed from mid-1970's African linguistics (and G76), 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 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 P80 in order to codify a syntagmatic relationship among high-toned events necessitated a further, somewhat bizarre, theory-internal complication regarding interpolation contours. Building on a core assumption that phonological specification of tones is usually sparse in intonation languages, P80 proposed that, in general, F0 interpolation functions that connect up phonological tones are monotonic: increasing functions should only increase, not decrease, and decreasing functions should only decrease, not increase. Since it was assumed that the L in a 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-1970's 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 a L tone between the two H tones, based on these theory-internal assumptions. It was therefore necessary to posit an exceptional non-monotonic 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). To further complicate matters, P80 assumed H tones sometimes were realized with a "late peak" on a nonprominent syllable following the accented syllable. This assumption constitutes a further lesser-known exception to the monotonic interpolation rule, one not commented on in P80, and amounts to a second type of non-monotonic function, termed "bulging interpolation" by Dilley and Heffner (2013).

As if this weren't enough, 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.[5] To account for regions of flat pitch, descriptive work on African languages in the 1970's (e.g., G76, Hyman & Schuh, 1974; Leben, 1973) posited phonological rules that enacted tone copying or spread to account for regions of flat pitch, i.e., cases where lexically-specified H or L tone showed a sustained pitch at the same level over multiple syllables.[6] 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.[7]

An alternative to the tone spreading rule that was proposed by PB88 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, & Arvaniti, 2000; Prieto, D'Imperio, & Gili Fivela, 2005). Note that this proposal (and tone spreading) requires inconsistency in treatment of autosegmental association. That is, while the original idea of G76 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 have 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 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 40 years' time also reveals that a strictly paradigmatic phonological treatment is simply inadequate – in spite of the best efforts of P80, PB88 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 if syntagmatic tonal relationships are part of the phonology, even though they are not. That is, scholars have annotated e.g. L+H* as a low valley plus a rising pitch, even though this choice is not supported by the underlying theories. In the following section, we demonstrate that a simple theoretical change – to assume that syntagmatic features are directly part of the representations of tones – allows building on the last 40 years of insights in an "enhanced" autosegmental-metrical framework.

A way forward: An enhanced autosegmental-metrical (AM$^+$) theory and the Rhythm and Pitch (RaP) transcription system

An "enhanced" autosegmental-metrical theory is proposed here, termed *AM$^+$ ("AM plus")*. AM$^+$ integrates insights from 40+ years of empirical work in intonational phonology, as well as research in speech perception, music cognition, and cognitive neuroscience. These proposals develop a notational device adapted for linguistic systems that is derived from insights about cognitive representations of non-linguistic tonal information – from auditory streaming studies, music cognition studies, and music theories for world's musical systems (Bregman, 1994; Burns, 1999; Dowling & Fujitani, 1971; Hannon & Trainor, 2007; Jones, Fay, & Popper, 2010; Patel, 2010).

A central part of AM$^+$ theory is its assumption that syntagmatic aspects of tone are part of cognitive representations for tonal systems cross-linguistically. AM$^+$ assumes syntagmatic features are part of *phonology*. AM$^+$ assumes that paradigmatic aspects of tones are *also* part of cross-linguistic tonal systems, and to be specified lexically in some tonal systems, and post-lexically in others. Syntagmatic aspects of tones, which specify the relationships of tones with one another in sequence, are likewise assumed to be specified in the lexicon in some cases, and to be assigned post-lexically 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 post-lexical levels.[8]

Note that by including both paradigmatic and syntagmatic aspects of tone in phonology, AM$^+$ theory is *not*, in fact, more complex than proposals of traditional AM theory, which assumed strictly paradigmatic tonal features (e.g., G76). This is because, as we show for AM$^+$ theory, *paradigmatic aspects of tone reduce to syntagmatic feature specifications*. This is a core insight of AM$^+$ theory – namely, that apparently paradigmatic tonal features can be formally re-expressed as syntagmatic ones. AM$^+$ thus preserves the economy of featural specification that was appealing in traditional AM theory. Further, as discussed above, in nearly 40 years, no theory based on paradigmatic level tones plus phonetic implementation rules has been put forward which successfully maps sequences of H and L tones to their expected F0 outputs for intonation languages, as conclusively shown in Dilley and Brown (2007).[9]

AM$^+$ 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 which 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, Freedland, Larson, & Hain, 1998; Chen, Liu, Xu, & Larson, 2007; Guenther, 2016; Guenther & Hickok, 2015; Hutchins & Peretz, 2011; Ning, Shih, & Loucks, 2014; Patel, Niziolek, Reilly, & Guenther, 2011; Pfordresher et al., 2015). To relate concepts of tones in traditional AM theory to AM$^+$, note that a "H" tone which in traditional AM theory was taken to correspond to an F0 peak (cf. P80) can be fundamentally re-expressed 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 which is autosegmentally associated with a H tone naturally maps in most speaking situations to an F0 peak. However, since tonal targets are intrinsically perceptual in nature, other F0 mappings are possible, such as F0 plateaux (Dilley & Brown, 2007; Knight, 2008) or variations in the F0 shape as given by e.g., 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., P80) can be re-expressed in AM$^+$ 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, a “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 re-expressed in AM$^+$ 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 which follows will then dictate whether the contour subsequently rises or falls (cf. a following tonal target that is higher or lower, respectively). Given that perceptual pitch lawfully relates to F0 in speech (d'Alessandro & Mertens, 1995; House, 1990; Mertens, 2004), AM$^+$ provides a unifying explanation for observed correspondences between abstract tones and their typical F0 consequences, cf. F0 peaks, valleys and plateaux.[10]

An experiment from Dilley and Brown (2007) provides further support for the proposal that categorical differences in F0 turning point timing, e.g. of F0 peaks and valleys, derive fundamentally from pitch targets that 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 & Steele, 1989) – Dilley and Brown showed that speakers imitated the level-pitched syllables by producing *categorical shifts* in F0 peak and valley timing; further, the categorical timing was predicted by *the syntagmatic relationship of relative height* borne by a level-pitched syllable to adjacent syllables, and 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 *Rhythm and Pitch (RaP)* Prosodic Transcription System (Breen, Dilley, Kraemer, & Gibson, 2012; Dilley & Brown, 2005), instantiates the proposals of AM$^+$ theory. The phonological representations in AM$^+$ 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). RaP’s **H** designates a tone that has feature specification [-same, +higher] and is phonetically *higher* than the previous tone. **L** designates a tone that has feature specification [-same, -higher] and is phonetically *lower* than the previous tone. E designates a tone with feature specification [+same] which is phonetically *equal* in pitch compared with the previous tone.[11,12,13]

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*. An AM$^+$ grid tier is a hybrid concept which generalizes across notions of a metrical grid row (e.g., Halle & Idsardi, 1995) and an autosegmental tier *a la* G76; it conceives of autosegmental association as expressly hierarchical, elaborating on the assumed, and eponymous, metrical representations of traditional AM theory. Further, 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 past, e.g., T_{n-1} / T_n = [-same, -higher]. In AM$^+$ theory, this is termed the *Reciprocal Property*.[14]

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. 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).[15] 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 which is at speakers' mean or habitual pitch levels is T / *r* = [+same].[16] 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 post-lexical assignment of features for T_1 / *r*. Thereafter, lexically-specified features for tones, together with post-lexical expressive factors like 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.[17] 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].[18] 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 the case of Hausa, for which syntagmatic relative heights of H tones in HL sequences distinguishes statements from questions (Inkelas & Leben, 1990; Inkelas, Leben, & Cobler, 1986).

Elaborating on Dilley (2005), five tonal levels can be captured in AM^+ 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 & Troost, 1989). We propose that [+/-small], like [+/-high], is only specified for [-same]. We further propose that tones are, in most cases, "unmarked" for [+/-small], in which case pitch range can vary expressively. A language with five level tones – Extra High, High, Mid, Low, Extra Low (EH, H, M, L, EL) – could thus be described as in Table 1.[19,20]

Table 1. Five level "paradigmatic" tone specifications derived from syntagmatic features [+/- same], [+/-higher], and [+/-small].

Lexical specification in phonology	*Phonetic interpretation*
T_{EH} / r = [-same, +higher, -small]	*substantially higher* than the mean pitch; high in the pitch range
T_H / r = [-same, +higher, +small]	*slightly higher* than the mean pitch
T_M / r = [+same]	*equal to* the mean pitch
T_L / r = [-same, -higher, +small]	*slightly lower* than the mean pitch
T_{EL} / r = [-same, -higher, -small]	*substantially lower* than the mean pitch; low in the pitch range

RaP and AM^+ theory elaborate productively on the relationship between hierarchical metrical structure and tonal associations. AM^+ and RaP adopt the starred tone notation "*" used previously to describe tones which 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^+ theory proposes that starred tones which 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 & Idsardi, 1995; Hayes, 1995), higher levels of AM+ 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 post-lexically. This allows an account of tone register phenomena, e.g., downstep, downdrift, and upstep (Clements & Goldsmith, 1984; Hyman, 1993; Inkelas & Leben, 1990; Inkelas et al., 1986; Ladd, 1988; Snider, 1999; Truckenbrodt, 2002).[21] A metrical account is consistent with a growing body of 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) in order for a starred tone to be indicated on a metrically prominent syllable. This is consistent with the restriction 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 P80's 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 non-prominent, metrically prominent without a pitch accent (i.e., without a pitch change, e.g., for flat pitch), and metrically prominent with 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 & Breen, in prep; Greenberg, Carvey, & Hitchcock, 2002).

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

- *Strictly monotonic interpolation functions.* Interpolation functions are strictly monotonic, ensuring that all turning points are coded as tones. Multiple studies have demonstrated evidence against P80's proposal that certain F0 turning points are not tones but rather reflexes of exceptional non-monotonic interpolation functions (Dilley & Heffner, 2013; Dilley et al., 2005; Ladd & Schepman, 2003).
- *Tones and timing slots*. An aspect of AM$^+$ theory which 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 (cf. 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. P80 predicted a constant timing relationship between the two tones of bitonal pitch accents, but this prediction has not been bourne out in production studies (Arvaniti et al., 1998; Arvaniti, Ladd, & Mennen, 2000; Dilley et al., 2005; Ladd, 2008). Following AM$^+$, RaP treats pitch accents as prominence-lending pitch *movements*, that is, locations of a local *change* in pitch. Following PB88, 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. 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 which is autosegmentally associated with a starred tone) – e.g., RaP's **L+ H***, which annotated as a sequence of two tones with a space between them, in contrast to ToBI*'s L+H*. Otherwise, the "+" is put on the right side of the tone, cf. 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 between H* vs. 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*.[22] These phonetic differences are important for capturing distinctions of *focus* (Breen, Fedorenko, Wagner, & Gibson, 2010; Féry & Krifka, 2008; Katz & Selkirk, 2011; Xu & Xu, 2005). Noting that both contours rise to a peak, RaP captures the contours as **L+ !H*** (for ToBI's H*) vs. **L+ H*** (for ToBI's L+H*). This treatment has the further effect of filling a theoretical gap in P80's theory having to do with the status of F0 values on phrase-initial unstressed syllables preceding a H*. When H* was on a non-initial syllable in the phrase and there was no phrase-initial boundary tone (such as P80's %H), it was theoretically unclear how phrase-initial unstressed syllables could obtain a F0 value under the theory, since 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 the RaP codes **H**, **L**, and **E** describe the syntagmatic relationship between the later-occurring tone and the earlier-occurring tone. 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^{+} assume that right edges of constituents license a variable number of unstarred tones, depending on post-lexical, language-specific rules. Thus, for example, final rises are treated as two unstarred tones (**H H** %) when a slope change is observed; otherwise, 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 upper edge of the pitch range or "<" for lower edge of the pitch range. P80 assumed uniformly that all intonation phrases end with two tones, a phrase accent and a boundary tone, which 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 provides a principled account for the well-known problem of Greek prenuclear accents, which were problematic for ToBI*. RaP characterizes Greek prenuclear accents as **L+ !H* +H**, predicting the observed slope change (Arvaniti et al., 1998; Arvaniti et al., 2000).
- *Sparse tonal representation*. Consistent with a sparse tonal representation, adjacent syntagmatic features are required to have different featural specifications. As a result, 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_1 T_2 T_3. 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 P80's assertion that strings of L* accents may give rise to a low, flat pitch is not supported in the present theory.

Conclusions. As we have outlined here, AM$^+$ offers a simplified theory which 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 the International Prosodic Alphabet (IPrA) (Hualde & Prieto, 2016).[23] RaP has been implemented as a full annotation system, with a publically available set of interactive training materials[24], and a corpus of RaP-labeled speech (Breen, Dilley, Brown, & Gibson, 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 RaP system to accurately assess prosodic structure (Sharpe, Fogerty, & van Ouden, 2017).

AM$^+$ 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$^+$ theory, Dilley and colleagues have experimentally demonstrated powerful perceptual constraints on metrical structure which can perceptually "garden path" listeners into hearing different organizations of words (Breen, Dilley, McAuley, & Sanders, 2014; Dilley, Mattys, & Vinke, 2010; Dilley & McAuley, 2008; Morrill, Dilley, McAuley, & Pitt, 2014; Morrill, Dilley, & McAuley, 2014).

In sum, Jun's chapter presents a useful outline for beginners unfamiliar with ToBI*, but it only skimmed the surface with respect to problems behind ToBI*. We have laid out some of the most damning aspects of the traditional AM theory behind ToBI*, which 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$^+$ is a theory which retains the insights of traditional AM approaches which have stood the test of time, while affording new insights and considerable improvement in phonetic transparency. RaP and AM$^+$ are informed by 40+ 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, rather than hindering it.

References

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation*. Cambridge: Cambridge University Press.

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 *Papers in Laboratory Phonology V* (pp. 119-130): Cambridge University Press.

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., & Hirschberg, J. (1994). *The ToBI annotation conventions. Technical report, The Ohio State University and AT&T Bell Laboratories, unpublished manuscript.* Retrieved from ftp://ftp.ling.ohio-state.edu/pub/phonetics/TOBI/ToBI/ToBI.6.html

Beckman, M., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 9-54): Oxford University Press.

Berent, I., Shimron, J., & Vaknin, V. (2001). Phonological constraints on reading: Evidence from the obligatory contour principle. *Journal of Memory and Language, 44*(4), 644-665.

Bishop, J., & Keating, P. A. (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*(2), 1100-1112.

Breen, M., Dilley, L. C., Brown, M., & Gibson, E. (2018). Rhythm and Pitch (RaP) Corpus. In. Philadelphia: Linguistic Data Consortium.

Breen, M., Dilley, L. C., Kraemer, J., & Gibson, E. (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*(2), 277-312. doi:10.1515/cllt-2012-0011

Breen, M., Dilley, L. C., McAuley, J. D., & Sanders, L. D. (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Language, cognition and neuroscience, 29*(9), 1132-1146. doi:10.1080/23273798.2014.894642

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*(7-9), 1044-1098.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feeback. *Journal of the Acoustical Society of America, 103*(6), 3153-3161.

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 215-264). San Diego: Academic Press.

Chen, S., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *Journal of the Acoustical Society of America, 121*(2), 1157-1163.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech - Evidence from F_o patterns of neutral tone in Standard Chinese. *Phonetica, 63*, 47-75.

Clements, G. N., & Goldsmith, J. (1984). Autosegmental Studies in Bantu Tone.

Coetzee, A. (2005). The Obligatory Contour Principle in the perception of English. In S. Frota, M. Vigário, & M. J. Freitas (Eds.), *Prosodies: With Special Reference to Iberian Languages* (pp. 223-246): Walter de Gruyter.

Cole, J. (2015). Prosody in context: A review. *Language, cognition and neuroscience, 30*(1-2), 1-31.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*, 141-201.

d'Alessandro, C., & Mertens, P. (1995). Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language, 9*(3), 257-288.

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

D'Imperio, M., Gili Fivela, B., & Niebuhr, O. (2010). *Alignment perception of high intonational plateaux in Italian and German.* Paper presented at the Proceedings of the International Conference on Speech Prosody, Chicago, US.

de Lacy, P. (2002). The interaction of tone and stress in Optimality Theory. *Phonology, 19*(1), 1-32.

de Lacy, P. (2014). Evaluating evidence for stress systems. In H. v. d. Hulst (Ed.), *Word Stress: Theoretical and Typological Issues* Cambridge: Cambridge University Press.

del Giudice, A., Shosted, R., Davidson, K., Salihie, M., & Arvaniti, A. (2007). Comparing methods for locating pitch "elbows". In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 1117-1120).

Dilley, L. C. (2010). Pitch range variation in English tonal contrasts is continuous, not categorical. *Phonetica, 67*, 63-81.

Dilley, L. C., & Brown, M. (2005). *The RaP (Rhythm and Pitch) Labeling System, Version 1.0*. Michigan State University. Retrieved from http://speechlab.cas.msu.edu/rap-system.htm

Dilley, L. C., & Brown, M. (2007). Effects of pitch range variation on F0 extrema in an imitation task. *Journal of Phonetics, 35*, 523-551.

Dilley, L. C., & Heffner, C. (2013). The role of f0 alignment in distinguishing intonation categories: Evidence from American English. *Journal of Speech Sciences, 3*(1), 3-67.

Dilley, L. C., Ladd, D. R., & Schepman, A. (2005). Alignment of L and H in bitonal pitch accents: Testing two hypotheses. *Journal of Phonetics, 33*(1), 115-119.

Dilley, L. C., Mattys, S., & 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*, 274-294. doi:10.1016/j.jml.2010.06.003

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*(3), 294-311. doi:10.1016/j.jml.2008.06.006

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49*(2), 524-531.

Dowling, W. J., & Harwood, D. L. (1986). *Music Cognition*. Orlando, FL: Academic Press, Harcourt Brace Jovanovich Publishers.

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1491-1506.

Féry, C., & Krifka, M. (2008). Information structure: Notional distinctions, ways of expression. In P. v. Sterkenburg (Ed.), *Unity and diversity of languages* (pp. 123-136). Amsterdam: John Benjamis.

Fitzroy, A. F., & Breen, M. (in prep). Metric structure and rhyme predictability modulate speech intensity during child-directed reading.

Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends Cogn Sci, 14*(6), 233-234.

Goldsmith, J. (1976). *Autosegmental phonology.* (Ph.D. dissertation), MIT, Cambridge, MA.

Greenberg, S., Carvey, H., & Hitchcock, L. (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* (pp. 56-61).

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

Grice, M., Ladd, D. R., & Arvaniti, A. (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., & Hickok, G. (2015). Role of the auditory system in speech production. In M. J. Aminoff, F. Boller, & D. Swaab (Eds.), *Handbook of Clinical Neurology. Vol. 129: The Human Auditory System: Fundamental Organization and Clinical Disorders* (pp. 161-175): Elsevier.

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

Halle, M., & Idsardi, W. (1995). General properties of stress and metrical structure. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 403-441).

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, forthcoming.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends Cogn Sci, 11*(11), 466-472.

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

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

Honorof, D. N., & Whalen, D. H. (2005). Perception of pitch location within a speaker's F0 range. *Journal of the Acoustical Society of America, 117*(4), 2193-2200.

House, D. (1990). *Tonal perception in speech*. Lund: Lund University Press.

Hualde, J. H., & Prieto, P. (2016). Towards an International Prosodic Alphabet (IPrA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 7*(1), 1-25.

Hutchins, S., & Peretz, I. (2011). Perception and action in singing. In A. M. Green, C. E. Chapman, J. F. Kalaska, & F. Lepore (Eds.), *Progress in Brain Research* (Vol. 191, pp. 103-118).

Hyman, L. H., & Schuh, R. G. (1974). Universals of tone rules: Evidence from West Africa. *Linguistic Inquiry, 5*(1), 81-115.

Hyman, L. M. (1993). Register tones and tonal geometry. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 75-108). Berlin, New York: Mouton de Gruyter.

Inkelas, S., & Leben, W. R. (1990). Where phonology and phonetics intersect: the case of Hausa intonation. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 17-34). New York: Cambridge University Press.

Inkelas, S., Leben, W. R., & Cobler, M. (1986). *The phonology of intonation in Hausa.* Paper presented at the Proceedings of the 16th Annual Meeting of NELS, Amherst.

Jakobson, R., Fant, G., & Halle, M. (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*(5), 323-355. doi:10.1037/0033-295X.83.5.323

Jones, M. R., Fay, R., & Popper, A. (Eds.). (2010). *Music perception*. New York: Springer.

Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language: Journal of the Linguistic Society of America, 87*(4), 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*(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.

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*(2), 530-544.

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

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

Lai, W., & Dilley, L. C. (2016). Cross-linguistic generalization of the distal rate effect: Speech rate in context affects whether listeners hear a function word in Chinese Mandarin. In *Proceedings of the International Conference on Speech Prosody* (Vol. 8, pp. 1124-1128).

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychological Review, 106*(1), 119-159. doi:10.1037/0033-295X.106.1.119

Leben, W. R. (1973). *Suprasegmental phonology.* (Ph.D. dissertation), MIT, Cambridge, MA.

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.

Liberman, M. (1975). *The intonation system of English.* (Ph.D. Ph.D. dissertation), MIT, Cambridge, MA.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oerhle (Eds.), *Language Sound Structure* (pp. 157-233). Cambridge, MA: MIT Press.

Manfredi, V. (1993). Spreading and downstep: prosodic government in tone languages. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 133-184). Berlin, New York: 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 Proceedings of the International Conference on Speech Prosody, Nara, Japan.

Monelle, R. (2014). Linguistics and semiotics in music. In: Routledge.

Morrill, T., Dilley, L., McAuley, J. D., & Pitt, M. A. (2014). Distal rhythm influences whether or not listeners hear a word in continuous speech: support for a perceptual grouping hypothesis. *Cognition, 131*(1), 69-74. doi:10.1016/j.cognition.2013.12.006

Morrill, T., Dilley, L. C., & McAuley, J. D. (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.

Myers, S. (1998). Surface underspecification of tone in Chichewa.

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*(2-3), 174-193.

Niebuhr, O., & Hoekstra, J. (2015). Pointed and plateau-shaped pitch accents in North Frisian. *Laboratory Phonology, 6*(3-4), 433-468.

Ning, L. H., Shih, C., & Loucks, T. M. (2014). Mandarin tone learning in L2 adults: A test of perceptual and sensorimotor contributions. *Speech Communication, 63*, 55-69.

O'Connor, R. J., & Arnold, G. F. (1973). *Intonation of colloquial English*. Bristol, U.K.: Longman Group Ltd.

Odden, D. (1995). Tone: African languages. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 444-475): Blackwell.

Patel, A. D. (2010). *Music, Language, and the Brain*: Oxford University Press.

Patel, R., Niziolek, C., Reilly, K., & Guenther, F. H. (2011). Prosodic adaptations to pitch perturbation in running speech. *Journal of Speech, Language, and Hearing Research, 54*(4), 1051-1059.

Peng, G., Zhang, C., Zheng, H.-Y., Minnett, J., & Wang, W. S.-Y. (**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., & Krumhansl, C. L. (1996). An experimental study of internal standards in Javanese and Western Musicians. *Music Perception, 14*(2), 95-116.

Pfordresher, P. Q., Demorest, S. M., Dalla Bella, S., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Theoretical perspectives on singing accuracy: an introduction to the special issue on singing accuracy (Part 1). *Music Perception: An Interdisciplinary Journal, 32*(3), 227-231.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation.* (Ph.D. dissertation), MIT, Cambridge, MA.

Pierrehumbert, J. (2000). Tonal Elements and Their Alignment. In M. Horne (Ed.), *Prosody: Theory and Experiment* (pp. 11-36). Dordrecht: Kluwer Academic Publishers.

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

Pierrehumbert, J., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica, 46*, 181-196.

Prieto, P., D'Imperio, M., & Gili Fivela, B. (2005). Pitch accent alignment in romance: primary and secondary associations with metrical structure. *Language: Journal of the Linguistic Society of America, 48*(4), 359-396.

Rice, K. D. (1987). *Metrical Structure in a Tone Language: The Foot in Slave (Athapaskan).* Paper presented at the Proceedings of the 23rd Annual Regional Meeting of Chicago Linguistics Society.

Searle, J. (1980). Minds, brains, and programs. *The 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., Fogerty, D., & van Ouden, D.-B. (2017). The role of fundamental frequency and temporal envelope in processing sentences with temporary syntactic ambiguities. *Language and Speech, 60*(3), 399-426.

Siegel, J. A., & Siegel, W. (1977). Categorical perception of tonal intervals: Musicians can't tell *sharp* from *flat*. *Perception & Psychophysics, 21*(5), 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 133.

Tierney, A., Patel, A. D., & Breen, M. (in press). Repetition enhances the musicality of speech and tone stimuli to similar degrees. *Music Perception*.

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

Vos, P., & Troost, J. (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception, 6*(4), 383-396.

Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics, 34*, 343-371.

Wong, P. C. M., & Diehl, R. L. (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*: American Mathematical Society.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication, 33*, 319-337.

Xu, Y., & Xu, C. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*, 159-197.

Zec, D. (1999). Footed tones and tonal feet: rhythmic constituency in a pitch-accent language. *Phonology, 16*(2), 225-271.

[1] However, see Yi Xu and colleagues (Xu & Wang, 2001; Xu & Xu, 2005) for an opposing point of view.

[2] G76’s 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 which can lead to new insights (e.g., the Chinese roomargument; 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 & Fedorenko, 2010) reflects well-known tensions between science- and humanities-oriented scholars who bump elbows in most 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 $f : X \to Y$ 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 paradigmatically-specified high tones sharing a common syntagmatic tonal referent, *r*. Given that the OCP descriptively captures widespread phonological perceptual phenomena across languages (Berent, Shimron, & Vaknin, 2001; Coetzee, 2005; McCarthy, 1986), we speculate that the OCP and the feature [+/- same] in fact both reflect cognitive processes of attention and memory consolidation for processing sensory information that is the *same* vs. *different* (Jones, 1976; Large & 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 & Wang, 2001), except for neutral tone syllables (Chen & Xu, 2006; Lai & Dilley, 2016). By contrast, in Chichewa, a Bantu language spoken mainly in Malawi, paradigmatic tonal specification appears to be much sparser (Myers, 1998).

[9] Dilley and Brown (2007) pointed out that the precise fix to the phonetic module proposed in PB88 is to specify a rule that H tones cannot fall below adjacent L tones. Dilley and Brown argue that this rule-based systematicity is better captured by revisiting the idea of syntagmatic specifications as part of phonological representations.

[10] In fact, 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 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 P80 and ToBI*’s H-H%).

[11] See Breen et al., 2012, for a conversion from MAE-ToBI to RaP. Converting RaP to MAE-ToBI entails information loss.

[12] The AM$^+$ theory can be viewed as validating the phonetically transparent INTSINT approach by Daniel Hirst (Hirst & Di Cristo, 1998 , this volume) and developing a theory to support its insights.

[13] 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 vs. plateau-shaped pitch accents.

[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 (cf. Arvaniti et al., 1998, p. 23). Although rise/fall approaches (e.g., 't Hart, Collier, & Cohen, 1990) did not correct predict timing aspects of F0 curves (cf. Arvaniti et al., 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 P80 and PB88, as pointed out in Dilley and Brown (2007).

[15] In Dilley (2005), the tonal referent was symbolized μ, the mathematical symbol for mean. We rename it *r* here, in order to avoid confusion with moras.

[16] In musical scale systems, each scale tone is defined by a relationship 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 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 & Harwood, 1986; Jones et al., 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 & Fujitani, 1971; Dowling & Harwood, 1986), consistent with the primacy of syntagmatic features in the present theory. Note that musical intervals are perceived categorically (Siegel & Siegel, 1977) as are lexical tones (Hallé, Chang, & Best, 2004); thus, frequency ratios do not need to be exact in order 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" (p. 367). He described low tone as "phonologically inert, because it is simply the absence of tone" (p. 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 & 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.

[20]In the Speech to Song Transformation, a phrase which is repeated several times shifts perceptually to being heard as sung (Falk, Rathcke, & Dalla Bella, 2014; Tierney, Patel, & Breen, in press). This phenomenon supports the contention that the pitches in speech are subject to similar cognitive organizing principles as in music.

[21] 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*, … *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 $(T_{n+1} / T_n) \cdot (T_{n+2} / T_{n+1}) \cdot \ldots (T_N / T_{N-1})$. The notation conveys abstract relationships, but also has a direct mathematical interpretation, since 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 & Siegel, 1977).

[22] 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(+)** …

[23] See also Hirst (this volume) and Hirst and Di Cristo (1998).

[24] Available at http://tedlab.mit.edu/tedlab_website/RaPHome.html