4 The ToBI Transcription System: Conventions, Strengths, and Challenges

Sun-Ah Jun

4.1 Introduction

Decisions regarding what information a prosodic transcription should include vary depending on the goal of transcription. For example, the goal may be to (i) represent the distinctive features of the prosodic system of a language, (ii) learn how prosodic categories are phonetically realized and how variable the categories are, (iii) train machines to achieve better synthesis and recognition of human speech, (iv) compare prosody across dialects or across languages, or (v) develop a corpus to explore various linguistic phenomena related to prosody. Thus, prosodic transcriptions may include annotations of prosodic properties that are meaningful at the level of phonology, phonetics, or even lower (i.e., the implementation of phonetic rules for a specific language), or they can be nondistinctive but still categorical in nature, capturing prosodic properties not specific to a certain language.

ToBI, which stands for tones and break indices, is the most well-known annotation system for prosody at the level of phonology. It aims to transcribe the phonological properties of intonation (the "tones" part) and the perceived degree of juncture between words (the "break indices" part), which together represent the prominence patterns and prosodic structure of an utterance. Thus, the tone labels used in the ToBI system of a specific language are meant to represent underlying tonal targets that are both meaningful to native speakers of the language and systematic and consistent across native speakers of the language variety.

The ToBI system was originally designed for transcribing the intonation and prosodic structure of English utterances (Silverman et al. 1992; Beckman and Hirschberg 1994; Beckman and Ayers Elam 1997; Beckman, Hirschberg, and Shattuck-Hufnagel 2005), and so it began as a transcription system for the prosody of English, more specifically for mainstream American English. However, after the development of the ToBI systems for other dialects of English, such as GlaToBI for Glasgow English (Mayo, Aylett, and Ladd 1997), as well as a few typologically different languages—for example, GToBI for German (Grice and Benzmüller 1995), K-ToBI for Korean (Beckman and Jun 1996; Jun 2000), and J_ToBI for Japanese (Venditti 1997), *ToBI* has come to refer to a general framework for prosodic transcription systems based on phonological properties.

The original ToBI was therefore renamed MAE_ToBI (Mainstream American English ToBI, although sometimes it is still simply called "English ToBI") when several ToBI systems of typologically various languages (e.g., Bininj Gun-wok, Cantonese, Chickasaw, German, Greek, Japanese, Korean, Mandarin, Serbo-Croatian) were presented at a satellite meeting of the Fourteenth International Congress of Phonetic Sciences in San Francisco in 1999.[1] Since then, many ToBI systems have been developed, covering languages whose

prosodic systems had been rarely studied or studied in different frameworks, such as Spanish (Beckman et al. 2002; Face and Prieto 2007), Bangladesh Bengali (Khan 2008, 2014), Mongolian (Indjieva 2009), Catalan (Prieto et al. 2009; Prieto 2014), European Portuguese (Frota 2014; Frota et al. 2015), and French (Delais-Roussarie et al. 2015), among others.

This chapter introduces what the ToBI conventions are and how the ToBI transcription system works and discusses the strengths and weaknesses of the ToBI system. It also describes some of the recent developments in the prosodic transcription system that attempt to address the ToBI system's known limitations. Before introducing how the ToBI system works, section 4.2 offers a brief description of the theoretical background and framework on which ToBI is based. Section 4.3 presents the ToBI conventions and how the ToBI system has been applied to typologically various languages, showing the workings of the phonological theory it has adopted. Section 4.4 presents the strengths of the ToBI system; section 4.5 discusses the problems and challenges ToBI users face, as well as recent developments that have been made in response to such challenges; and section 4.6 concludes the chapter.

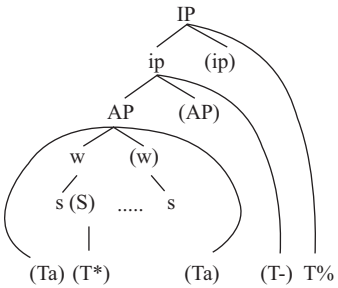
4.2 The Theoretical Background of the ToBI System

The ToBI system includes two main types of annotation: tonal labels and break index labels. The tonal labels are based on the autosegmental-metrical (AM) model of intonational phonology. (See Bruce 1977; Liberman 1975; Pierrehumbert 1980; Liberman and Pierrehumbert 1984; Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988; Ladd 1983, 2008a; and Pierrehumbert, commentary to chapter 11, this volume. See also Arvanti, chapter 1, this volume, for a detailed description of the AM model of intonational phonology.) In this model, an intonational contour is analyzed as a linear sequence of high (H) and low (L) tones. As an autosegment, the intonational tones can be associated with a specific syllable or mora in a word, that is, head, with the edge of a specific prosodic unit, or both, reflecting the metrical and/or prosodic structure of the utterance (Beckman 1996). When the tone is associated with an intonationally prominent syllable (either because it is metrically or rhythmically strong or because it is lexically marked), it is called a *pitch accent* (marked with a star [*], e.g., H*, following Goldsmith 1976 and Pierrehumbert 1980). When the tone is associated with the edge of a prosodic unit, it is called a *boundary tone* (e.g., H%, following Liberman 1975 and Pierrehumbert 1980). In this way, these intonational tones achieve two major functions: (i) marking the prominence relationship among the syllables (or moras) within a word and among the words within a phrase and (ii) marking prosodic grouping and hierarchical prosodic structure of the utterance.[2]

The AM model of intonational phonology, illustrated well in the models proposed by Beckman and Pierrehumbert (1986) and Pierrehumbert and Beckman (1988), assumes three prosodic units defined by intonation: an *intonational phrase* (IP), which is the largest prosodic unit; an *intermediate phrase* (ip), a prosodic unit smaller than an IP; and an *accentual phrase* (AP), which is slightly larger than a word (w) and smaller than an ip. These prosodic units form a strictly layered hierarchical prosodic structure (Selkirk 1986; Nespor and Vogel 1986; Pierrehumbert and Beckman 1988). A higher-level prosodic unit is assumed to be exhaustively parsed into one or more lower-level prosodic units. Therefore, any word or syllable should be part of each prosodic unit, and because an IP dominates an ip, which dominates an AP, an IP-final syllable is assumed to also be an ip-final syllable, as well as an AP-final syllable. A tree diagram of intonationally defined prosodic structure is shown in (1).[3] Each prosodic unit can be marked by a boundary tone (T% for an IP boundary tone, T- for an ip boundary tone, Ta for an

AP boundary tone; where T = H or L or its combinations), which is typically realized on the last syllable of a prosodic unit, regardless of whether the syllable is stressed. Among the boundary tones, only an IP boundary tone is obligatory. Lowercase *s* refers to a syllable, and *S* is a metrically strong/lexically marked syllable (the latter of which can be associated with a pitch accent, T*).

(1) Intonationally defined prosodic structure.



The AM model of intonational phonology allows only two tonal levels, H and L, and as in the analysis of African tone languages (Goldsmith 1976; Leben 1976), a rising or falling contour tone is represented by a combination of these level tones. This differs from the British intonation models (e.g., Palmer 1922; Halliday 1967; Crystal 1969) and from Bolinger's (1951, 1958) model where a contour tone (configuration) was represented as a single tonal unit. That is, a pitch accent in the AM model can be monotonal or bitonal, depending on the shape of the fundamental frequency (F0) contour around an accented syllable. A monotonal pitch accent represents a level tone, such as H* and L*, and a bitonal pitch accent represents a rising tone such as L + H* or a falling tone such as H* + L. By decomposing a contour tone into an L and an H level tone, the AM model can also distinguish contour tones where the F0 peak or valley is aligned differently relative to an accented syllable. That is, when a bitonal pitch accent is associated with an accented syllable of a prominent word, the starred tone of the pitch accent is realized on the accented syllable of the word. The unstarred tone, which either precedes (called a *leading tone*) or follows (called a *trailing tone*) the starred tone, is typically realized on the syllable immediately preceding or following the star-toned syllable, respectively. For example, L + H* is a rising tone with an F0 peak during the accented syllable and an F0 valley on the immediately preceding syllable, while L* + H is a rising tone with an F0 valley during the accented syllable and an F0 peak on the immediately following syllable.

As a phonological model of intonation adapted mostly from Pierrehumbert (1980), the AM model of intonational phonology posits the simplest possible abstract underlying representation. It accounts for the surface variation of underlying tones by rules that map the phonological representation (abstract level tone target sequences) to the phonetic representation of intonation (the F0 contour). Thus, even though the AM model assumes only two tonal levels in the underlying representation, various tonal levels realized in the surface F0 contour are explained by the rules that affect pitch range, such as upstep or downstep relative to the immediately preceding tone, the strength of prominence (e.g., a high nuclear pitch accent is often higher than a high prenuclear pitch accent in English), the lexical status of a tone (e.g., a lexical H tone is higher than a phrasal H tone in Japanese), or the function of tone (e.g., an L% boundary tone is typically lower than a L* pitch accent).

In addition to the rules adjusting the pitch range for a tonal category's realization, the model also adopts the target-interpolation rule to explain the surface F0 contour over syllables that have no underlying tonal targets. As a phonological model, only distinctive pitch targets are associated with a certain syllable, and syllables without a tonal target receive their surface F0 values through direct interpolation between two adjacent tonal targets, that is, the tones before and after the toneless syllable(s). In general, interpolation is done locally between two tonal targets, but it can be sensitive to the tone's function. For example, in English, interpolation does not occur between a prominence-marking pitch accent and a boundary tone in the same way that interpolation occurs between two pitch accents.

Distinctive intonational tone categories, such as pitch accents and boundary tones, proposed for a specific variety of a language will form the tonal inventory for that variety's ToBI system. But a complete ToBI annotation system should also include labels for the break index (BI) of the language. The BI in the ToBI system is in general marked by Arabic numerals, representing the perceived degree of juncture between any two adjacent words. Typically, the higher the numeral, the larger the perceived juncture between the words. Therefore, the distribution of BIs reflects prosodic grouping of words and thus the utterance's hierarchical prosodic structure.

The convention of using numbers to represent the degree of juncture is adopted from work by Price and her colleagues (Price et al. 1991; Wightman et al. 1992), who examined the mapping between syntactic and prosodic structure, and the role of prosody (especially boundary marking and prominence) in the resolution of various types of syntactic ambiguity. As Beckman et al. (2005) described in detail, the ideas that Price and her colleagues investigated have a long tradition, going back to the work known as *instrumental phonetics* (see Ladd [1996] 2008a and Price et al. 1991 for a review), where researchers examined the phonetic correlates of syntactic structure (e.g., Lieberman 1967; Lehiste 1973; Klatt 1975; Cooper and Paccia-Cooper 1980; Scott 1982), and work on *prosodic phonology* (e.g., Gee and Grosjean 1983; Selkirk 1986; Nespor and Vogel 1986), where researchers tried to predict prosodic structure from syntactic structure and explained the domain of postlexical phonological rules based on a sentence's syntactic structure.

In sum, the ToBI system is a transcription system for annotating the phonological properties of prosody, especially the prominence and prosodic structure of an utterance, analyzed in the AM model of intonational phonology for a specific language variety. The following section describes the conventions of the ToBI system based on data from English and other typologically various languages, illustrating how the phonological approach of AM theory is manifested in various ToBI systems.

4.3 ToBI Conventions and Various ToBI Systems

The goal of developing the original ToBI system was to provide a common vocabulary for tagging phonological properties of prosody and intonation to a community of users working at different sites. This common vocabulary would then facilitate the ability to share and interpret each other's data and to build a large prosodically transcribed database. In their chapter describing the original ToBI system, Beckman et al. (2005, 12–14) state that a viable ToBI system must conform to the following four principles. First, the conventions should be "accurate," conforming to an established body of research in the intonational phonology of the language variety, and, when possible, the conventions should also be informed by work on dialectology, pragmatics, and discourse analysis

for the language. Second, the conventions should be "efficient," by not transcribing information that can be extracted from the signal or derived from online resources automatically (e.g., location of lexical stress). Also, prosodic phenomena should not be labeled if they cannot be labeled consistently due to a lack of understanding of the phenomena by the research community. Third, the conventions should be "easy" enough to learn by consulting a manual (which includes sound files illustrating ToBI transcriptions), and "reliable" enough to be applied consistently among transcribers. For this, intertranscriber consistency should be evaluated and the conventions should be reviewed regularly and upgraded, if needed, by agreed-on groups. Finally, a ToBI transcription should never replace a permanent record of the speech signal with a symbolic record. Instead, it should integrate symbolic labels with continuous measures derived directly from the signal. This is because a ToBI system is not simply a transcription system, but also "a tool for observing the signal and communicating one's observations to the larger community in a common language" (Beckman et al. 2005, 14).

A ToBI transcription therefore requires an audio recording of an utterance and a record of the F0 contour, aligned with symbolic labels written on four parallel quasi-independent tiers: words, tones, BIs, and miscellaneous (misc). In the original ToBI system, UNIX-based xwaves software was used to display the waveform, pitch track, and four tiers (placed above the pitch track), and a label on each tier was placed right-justified relative to a certain time point. That is, the end of each label is aligned with the end of the acoustic signal corresponding to each word (for the words and BIs tiers) or with the time point of a tonal event (for the tones tier) or a nonspeech event or interval (for the misc tier). Figure 4.1 provides an example, showing the English utterance "Okay, they have a couple flights," transcribed in MAE_ToBI, using xwaves. Later ToBI systems have used different software, such as PitchWorks (Scicon R&D; http://www.sciconrd.com/index.aspx), WaveSurfer (Department of Speech, Music, and Hearing, KTH; http://www.speech.kth.se/wavesurfer/), or Praat (http://www.fon.hum.uva.nl/praat/). Praat has been used widely in recent years by ToBI users, and unlike the format used in xwaves, has two different types of labeling tiers: it places a label either as an event that has some duration, in an *interval tier*, or as an event occurring at a specific time point, in a *point tier*. In this format, an interval tier is appropriate for the words tier, which marks the beginning and the end of each word determined by examining the waveform and spectrogram, and a point tier is appropriate for the other three tiers. Figure 4.2 shows an example of ToBI transcription of the same utterance as in Figure 4.1, but displayed in Praat. The subsections that follow describe the conventions on each tier based on data from English and other languages.

4.3.1 The Words Tier

The words tier is where an orthographic transcription of each word in the utterance appears, aligned with the acoustic representation of the word (in the waveform or the spectrogram, or both). What counts as a word is language-specific, and how to transcribe it can be decided by the research community for a specific ToBI system. For example, even though the morphosyntactic function of a case marker or a postposition is similar between Japanese and Korean, these morphemes are treated differently prosodically in each language. They are entered as an independent word in J_ToBI (aligned with BI 1), but as part of a word in K-ToBI. This is because case markers and postpositions are sometimes produced separately from a stem in Japanese, but not in Korean.

The orthography of a word can be written in language-specific characters if the font is available, or in a romanized text of the word, as in J_ToBI or K-ToBI. In Pan-Mandarin

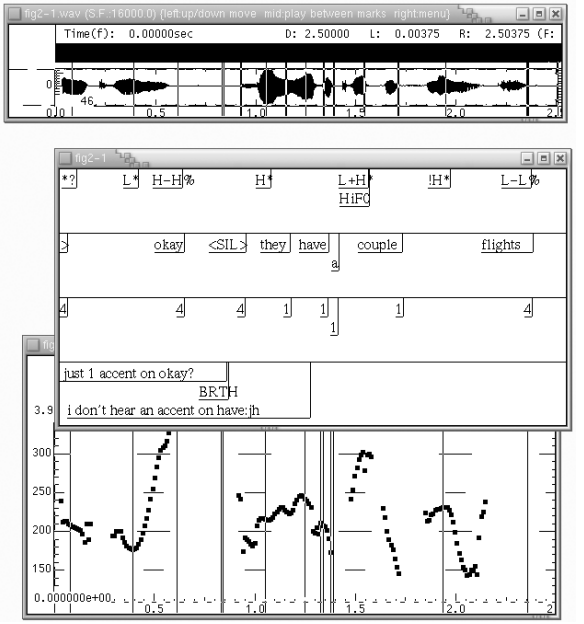


Figure 4.1
An example English utterance, "Okay. They have a couple flights," transcribed in MAE_ToBI in xwaves. The tiers are shown in the order of words, tones, BIs, and miscellaneous. *Source*: Courtesy of Beckman, Hirschberg, and Shattuck-Hufnagel (2005, figure 2.1); reproduced by permission of Oxford University Press. https://global.oup.com/academic/product/prosodic-typology-9780199208746?cc=us&lang=en&. BRTH, breath.

ToBI (Peng et al. 2005), syllable-by-syllable transcriptions in Chinese characters are given on the words tier, but transcription of syllables in modified Pinyin romanization is given on the romanization tier, called the *Romazi tier*. Similarly, the words tier in Cantonese C-ToBI (Wong, Chan, and Beckman 2005) provides a syllable-by-syllable transcription in Chinese characters, but because not every labeling platform has Chinese characters, the C-ToBI developers suggest making the words tier optional and a syllables tier obligatory, where an alphabetic transliteration is provided for every syllable in the words tier.

Finally, when the boundary of a word is not easy to locate (for example, due to cliticization of adjacent function words), the orthography of a prosodic word rather than a morphosyntactic word (e.g., *gonna* rather than *going to*) can be written on the words tier.

4.3.2 The Tones Tier

The tones tier is for transcription of distinctive tonal events and other tone-related tags. Distinctive tonal events can include (i) phrasal tones, marking the edge of a prosodic unit (e.g., Ha for an accentual phrase boundary tone in Bengali [Khan 2008, 2014] and Korean [Jun 2000, 2005a] or L% for a low boundary tone of an IP in many languages); (ii) postlexical pitch accents, marking the prominence of a word (e.g., L*, L+H* in English, German [Grice, Baumann, and Benzmueller 2005], Greek [Arvaniti and Baltazani 2005], or Bininj Gun-wok [Bishop and Fletcher 2005]); or (iii) lexically specified tonal events (e.g., lexical pitch accent H*+L in Japanese [Venditti 1997, 2005] and

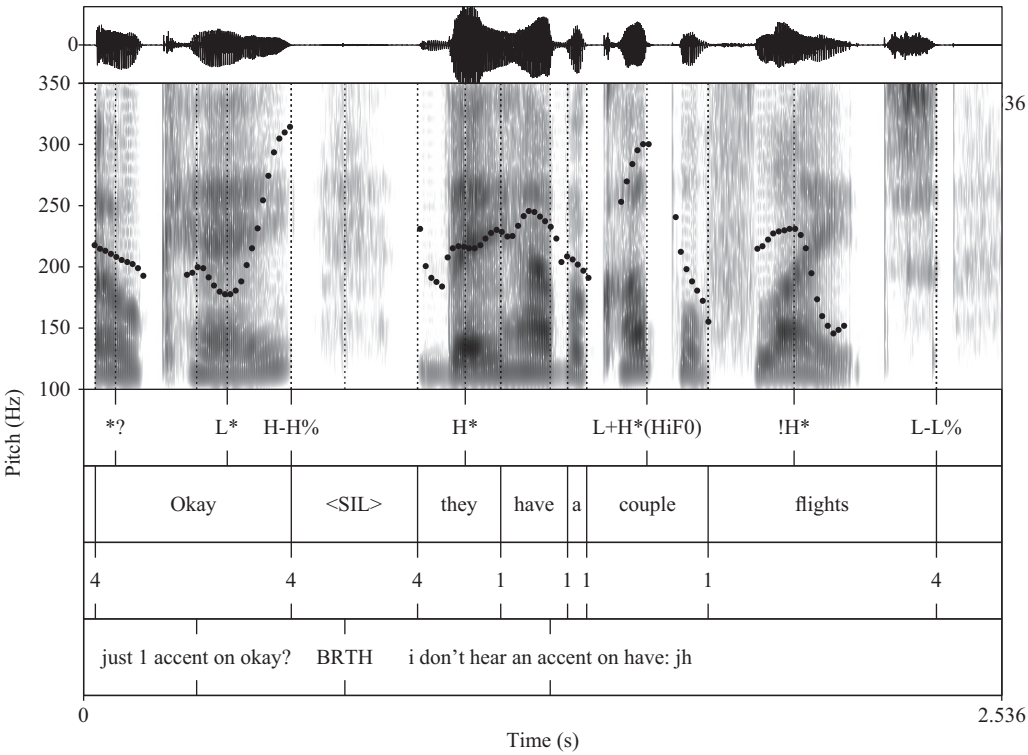


Figure 4.2
An example of MAE_ToBI transcription of the same utterance shown in figure 4.1 but transcribed in Praat.

L* + H in Serbo-Croatian [Godjevac 2005]), lexical tones in Cantonese [Wong, Chan, and Beckman 2005], or H^{λ} morphological pitch accent in Chickasaw [Gordon 2005]).

Because distinctive tonal categories and their realizations are language- or dialect-specific, the same tonal label is not necessarily realized in the same way across languages, and the same F0 contour can be transcribed with different tonal labels across languages. For example, as mentioned earlier for English intonation, a rising pitch accent L* + H is used to label an F0 contour with an F0 valley on an accented syllable and an F0 peak on the immediately following syllable. However, in Greek, the same tonal label is used for an F0 contour with an F0 valley at the very beginning or slightly before the onset of the accented syllable and an F0 peak at the beginning of the postaccentual vowel (Arvaniti, Ladd, and Mennen 1998, 2000; Arvaniti and Baltazani 2005). This is possible because each of these languages has another, contrasting rising pitch accent, such as L + H*, whose F0 peak is aligned later than that of L* + H. Thus, what justifies a tonal label in a language is its status as contrastive or noncontrastive, not the exact details of its phonetic realization.

Furthermore, the same F0 contour with the same tone-text alignment can be transcribed differently across languages depending on the lexical prosody of the target language. For example, a phrase-medial rising F0 contour over a two syllable word can be categorized as either L* + H or L + H* if the first or the second syllable, respectively, is the metrical head of the word (as in English, German, and Spanish ToBIs). If both tones are

not associated with a head syllable, the contour can be categorized as a LH boundary tone or a LH tonal melody (as in Mongolian and Korean ToBIs) or an L boundary tone followed by an H phrasal one (as in J_ToBI). Or, if the first syllable is the head syllable of a word while the following syllable is the last syllable of a word, the sequence can be categorized as an L* pitch accent followed by an H boundary tone (as in Bangladesh Bengali ToBI).

In addition to the difference in tone-text alignment, tone labels reflect differences in the tonal scale or pitch range. While the AM model assumes only two tonal levels, H and L, it is also known that an H tone target is sometimes lower compared to the preceding H target, more so than would be expected given simple declination over the utterance. Furthermore, such reduction in pitch range may apply iteratively to the subsequent H tone targets. H tones realized in such a reduced pitch range are called "downstepped H tones." Because the context of downstep is similar to the downdrift phenomenon in African tone languages (Leben 1973), Beckman and Pierrehumbert (1986) proposed downstep for their AM model of English intonation to be triggered by the preceding bitonal pitch accent within the same intermediate phrase. However, the original ToBI conventions for English, influenced by Ladd (1983), adopted a theory-neutral downstep diacritic, !, before H to annotate a downstepped H tone. With this change, the MAE_ToBI revised the tonal label H+L* in Beckman and Pierrehumbert (1986) to H+!H* to reflect an F0 fall from a high to mid target rather than from a high to low target, and revised the label H*+L to H* followed by !H*. This therefore increased the tone inventory for English by adding a downstepped version of all pitch accents that have an H tone component (i.e., !H*, L+!H*, L*+!H, !H+!H*) and also a downstepped version of the H phrase accent (i.e., !H-).[4] This introduction of a downstep diacritic made the tonal category in the ToBI system more faithful to the surface F0 contour than that proposed in the AM model.[5]

The tones tier also includes tones marking the prosodic structure of an utterance. As mentioned in section 4.1, a language can have at least three prosodic units marked by intonation above the word level: IP, ip, and AP.[6] The tones marking the edge of a prosodic unit, called *boundary tones*, are accompanied by a diacritic, indicating that the prosodic unit that the tone is affiliated with. For example, *a* is used to mark an AP boundary tone (e.g., Ha, La), - for an ip boundary tone (e.g., H-, L-), and % for an IP boundary tone (e.g., H%, L%). Unlike the pitch accent, which is located within the accented syllable, specifically aligned with the F0 minimum for L*-type tones (e.g., L*, L*+H) and with the F0 maximum for H*-type tones (e.g., H*, L+H*), the boundary tone is typically located on the tones tier, aligned with the end of the relevant prosodic unit. When an ip is IP-final, though, both ip-boundary and IP-boundary tones can be realized sequentially at the edge of an IP, as in English (e.g., L-H%). In some languages, however, the boundary tone marking a lower-level unit (e.g., an ip) can be overridden by that marking a higher-level unit (e.g., an IP), as in Bangladesh Bengali or Korean. Thus, how multiple boundary tones are realized when they are hosted by the same syllable is language-specific.

The pitch accents and boundary tones mentioned so far are all postlexical tones, but the tones forming an intonation contour can also originate from a lexical tone, such as a lexical pitch accent (in Japanese and Swedish), a morphological pitch accent (in Chickasaw), or a lexical tone (in Mandarin and Cantonese). In the ToBI system of most languages, both lexical and phrasal tones are labeled on the same tier, a tones tier, because the lexical status of the tone is recognizable from the diacritics on the tone symbol. For example, in J_ToBI (1997, 2005), there is only one lexical pitch accent and

it is marked with a star diacritic (i.e., H* + L). All other tones are phrasal tones: an AP phrasal tone marked with a diacritic - (i.e., H-) or a boundary tone marked with a diacritic % (i.e., an L% AP-boundary tone and a few IP-boundary tones [e.g., H%, LH%]). In Chickasaw ToBI (Gordon 2005), there is one morphological pitch accent, and it is marked with a lambda (i.e., H^{λ}). All other tones are phrasal tones: a postlexical pitch accent marked with a star (e.g., H*), AP phrasal tones without any diacritics (e.g., L, H), and an IP boundary tone marked with % (e.g., H%). Similarly, both lexical tones and IP-boundary tones are labeled on the tones tier in Cantonese ToBI (Wong, Chan, and Beckman 2005), but lexical tones are represented using Chao's system of tone numbers (e.g., 35, 53) and IP tones are marked with % (e.g., L%, HL%). However, lexical tones and postlexical tones can also be labeled on two separate tiers. In the Pan-Mandarin ToBI (Peng et al. 2005), the lexical tone is labeled on the Romazi tier while postlexical tonal events such as IP-boundary tones (i.e., L%, H%) and pitch-range–related labels (e.g., %reset, %e-prom) are labeled on the tones tier.

Finally, as shown in the original ToBI system, the tones tier can also include tone-related tags that are not intended to mark distinctive tonal events. Such tags are instead used to mark surface F0 patterns related to the phonetic realization of a tonal category or to discourse information. The most common example is to use a diacritic to mark atypical or allotonic realizations of an underlying tonal category, such as the use of < to mark a delayed F0 peak and > for an early F0 peak when the F0 peak is realized later or earlier than expected, respectively, from the typical tone-text alignment of the underlying tone. Another example is the use of *w* to mark a weak realization of an underlying tonal target, as in the undershoot of an L target in Greek prenuclear pitch accents (i.e., wL* + H) or in Japanese AP boundary tones (wL% or %wL).[7] Tone-related tags can also include a label that can be used to retrieve F0 values to study discourse structure (e.g., HiF0 in English on the highest F0 point among pitch accents within an ip), or a label to capture how pitch range is influenced by disfluency (e.g., %r at the onset of pitch-range reset after disfluency). Because a ToBI system is a tool for observing the signal and creating a communal corpus, adding these tags on surface F0 information related to the distinctive tonal categories allows researchers to learn how an underlying tone is phonetically realized in what context and how intonation interacts with other components of grammar and reflects various communicative pressures.

This idea is further extended in K-ToBI, version 3 (Jun 2000, 2005a), where the tones tier is split into two tiers, a phonological tone tier and a phonetic tone tier, to capture the variability and distribution of the surface tonal categories of the AP. Unlike English intonation, where all tonal categories are distinctive and each type of pitch accent contributes compositionally to the pragmatic interpretation of an utterance (Pierrehumbert and Hirschberg 1990), most of the tonal targets in Korean intonation are not distinctive. Distinctive tones are those marking the end of prosodic units larger than a word, and tones that are not AP-final are not distinctive. For example, an AP-initial tone can be L or H depending on the laryngeal feature of the AP-initial segment,[8] that is, predictable and thus not distinctive. AP-internal tones can be H on the second syllable or L on the penultimate syllable, but the presence of these tones varies by factors not fully known, though some of the tendencies can be explained by the AP length and by speech style. Even the AP-final tone, though usually H, can be L for factors that are also not well understood. Because the tones in ToBI systems are intended to be distinctive, nondistinctive yet categorical F0 targets such as these should not be included in the tones tier. For this reason, Jun (2000) chose to place these on a separate tier called the phonetic tone tier and put only distinctive tones on the phonological tone tier for

K-ToBI. It was hoped that labeling these surface tonal categories in a large corpus will help researchers to determine what factors and conditions trigger such variation, and thus improve the current model of Korean intonational phonology.

In sum, the tones tier is used to transcribe distinctive tonal categories of a specific language variety, regardless of their function (marking prominence or prosodic structure) and lexical status (lexical or phrasal), as well as to transcribe surface tonal events that are meaningful and interesting to the research community of the target language.

4.3.3 The Break Indices Tier

The BIs tier is for transcribing the perceived degree of juncture or the strength of association between each pair of words (and between the final word and the silence at the end of the utterance), capturing the hierarchical nature of an utterance's prosodic grouping. The original ToBI system (for English) proposed five indices in numerical numbers: four numbers for four different degrees of disjuncture and one number for a mismatch between the degree of juncture and the tonal cues that mark prosodic grouping. Among the four degrees of juncture, two corresponded to prosodic boundaries higher than the word level, that is, IP and ip, and two lower than the ip-level juncture, that is, ip-medial word boundary and a boundary weaker than a typical word boundary. (2) shows a brief description of five BIs in MAE_ToBI, represented in numeric numbers from 0 to 4. On the BIs tier, each index is labeled and aligned with the end of each word in the words tier.

(2) The BIs in MAE_ToBI

- 0 A juncture smaller than a typical phrase-medial word boundary. It corresponds to the juncture between a content word and a function word when the function word is cliticized. It can also be used for a weakened word juncture due to coalescence of adjacent segments across a word boundary.
- 1 A typical phrase-medial word boundary.
- 3 A juncture corresponding to an ip boundary.
- 4 A juncture corresponding to an IP boundary.
- 2 A mismatch between a tonal cue and the degree of juncture that mark a prosodic boundary. It includes two types of mismatch: (i) a juncture corresponding to ip or IP but with no tonal event defining ip or IP, and (ii) a juncture weaker than an ip level but with a clear tonal event for an ip or IP.

Because the BI is tightly linked to the prosodic structure of a language, the number of BI categories will differ across languages, reflecting the number of prosodic units a language has. However, the number of prosodic units defined by a tone does not necessarily match the number of BIs, because a prosodic unit is not necessarily marked by intonation. For example, Cantonese has only one prosodic unit defined by intonation, IP, but its ToBI system has three BI numbers: 2 for an IP boundary, 1 for a foot boundary, and 0 for a foot-internal syllable boundary. Furthermore, the Pan-Mandarin ToBI has six BI numbers, but only one BI (BI 4) corresponds to a prosodic unit defined by pitch reset.

As the list of BIs shows, an increasing value of BI refers to an increasing degree of juncture: $0 < 1 < 3 < 4$. For this reason, BI 2, which refers to a mismatch in the original ToBI, has sometimes been misinterpreted as a juncture larger than BI 1 and smaller than BI 3. Furthermore, the BI 2 in MAE_ToBI included two types of mismatch cases, as described in the BIs list in (2). To improve this situation, some of the later ToBI systems (e.g., Japanese, Korean, Greek, and Chickasaw) used BI 2 as the juncture larger than BI 1

and smaller than BI 3, and added a diacritic *m* after a BI number for mismatch cases. For example, 2m is used to label a BI 2–like juncture but without having tonal cues for the prosodic unit corresponding to BI 2 (which is an AP in Japanese, Korean, and Chickasaw, but an ip in Greek). Similarly, 3m is used to label a BI 3–like juncture but without having tonal cues for the prosodic unit corresponding to BI 3 (which is an IP in all four languages mentioned).

Because some of the BIs do correspond to a tonally defined prosodic unit, labeling both a BI on the BIs tier and a boundary tone on the tones tier would mark the same prosodic unit. For example, in MAE_ToBI, BI 3 and BI 4 are labeled on the BIs tier aligned with a phrase accent and an IP boundary tone, respectively, on the tones tier. This seems to violate the "efficiency" principle of the ToBI system, mentioned at the beginning of section 4.3. Namely, the system has to be efficient and should not transcribe predictable information. So, the researchers of a specific ToBI system could decide not to label BIs that are predictable from other tiers but label only those not predictable such as mismatch cases. But this would work if both tones and BI tiers are fully specified. In the cases where a research group wants to examine only the tonal patterns of a prosodic structure or only the degree of juncture or grouping between words, the tier of interest should be fully labeled regardless of whether some label is predictable from the other tier. That is, depending on the goal of the researchers, what to label on what tier can be different.[9]

Finally, a BI can also transcribe a disfluent juncture, a juncture produced disfluently or unnaturally for various reasons (e.g., a memory problem, a sudden change in a word choice, interruption by other speakers, stuttering), and the original ToBI proposed three types of labels for a disfluent juncture, as in (3).

(3) ToBI labels for disfluent juncture
1p for an abrupt cutoff of a word
2p for prolongation, but without a phrase accent
3p for prolongation, with an accompanying phrase accent

These two types of disfluency, cutoff and prolongation, have been adopted by most of the later ToBI systems, but these are not enough to transcribe all types of disfluencies found in speech.[10] To transcribe stuttered speech, Arbisi-Kelm (2006) increased the types of disfluency from two to five by adding "restart," "pause," and "filler." However, BIs describing disfluent junctures have not been used extensively across ToBI systems, probably because most ToBI systems have been proposed based on reading and scripted speech or semi-spontaneous speech and not much from free conversation speech.

4.3.4 The Miscellaneous Tier

The misc tier is for transcription of any comments or remarks about the utterance. It can be a comment about the utterance such as silence, audible breaths (an example is "BRTH" in figure 4.1), laughter, false start, hesitation, disfluencies, and other spontaneous speech effects. It can also be used to include labelers' remarks for themselves about a transcription, or questions to other labelers (e.g., "just 1 accent on okay?" in figure 4.1). If a label on the misc tier is for a localized event, it should be located at the time point of the event (e.g., a label "disfl" is written at the approximate time point where some disfluency is perceived), but if the event has an identifiable interval such as laughter does, the convention is to use paired labels to mark the starting point of the event with < and the ending point of the event with > (e.g., laughter< . . . laughter>).

4.3.5 ToBI Labels for Uncertainty and an Alt Tier

The original ToBI system provided a convention to transcribe a labeler's uncertainty about a BI or a tonal category. When a labeler is not sure about the degree of juncture between two adjacent BIs, the convention is to add a dash (-) after a BI of a larger juncture (e.g., 4- is used for a BI ambiguous between 3 and 4).[11] For a tonal category, *? is used when a labeler is not sure whether a syllable is pitch accented or not, and X*? is used when a labeler is sure about the presence of pitch accent on a syllable but not sure about the pitch accent type. The same rule is applied to other tonal categories. So, uncertainty about the presence and the type of a phrase accent is labeled as -?and X-?, respectively, and that of an IP boundary tone is labeled as %?and X%?, respectively.

However, this way of labeling uncertainty is often found to be unsatisfactory because it does not record relevant details of the ambiguity. That is, while it tags that the labeler is uncertain about the transcription, it does not specify what possible tonal labels were considered by the labeler, or whether one potential label should be preferred over another. To fix this problem, a fifth tier, Alt (alternative), was proposed at the Workshop on ToBI for Spontaneous Speech, held in Boston in 2004, and the use of an Alt tier has been tested by labeling a large speech sample in Brugos et al. (2008). Brugos and her colleagues proposed adding a question mark next to an ambiguous or uncertain tonal label (e.g., H*?) or a BI label (e.g., 3?) in the respective tier and providing an alternative tonal or BI label on the Alt tier (e.g., !H* or 2), with the alternative label aligned with the corresponding labels on the tones or BIs tiers. If a sequence of labels is ambiguous, instead of one label, they proposed adding a square bracket before and after the affected labels on the Alt tier (e.g., [2 L + H*]).

4.3.6 Other Tiers

As mentioned, the original ToBI system provided four core tiers (tones, words, BIs, misc), but later ToBI systems have added new tiers either to capture language-specific prosodic properties (e.g., Romazi, syllable, stress, and sandhi tiers in Pan_Mandarin ToBI; foot and syllable tiers in C_ToBI), to study the interface between prosody and other areas of linguistics (e.g., the phonetic transcription tier in GRToBI and Chickasaw ToBI, the prosodic word tier in GRToBI, the finality tier in J_ToBI), or to help labelers transcribe non-native or unfamiliar languages (e.g., the gloss tier in Chickasaw ToBI and Bininj Gun-wok ToBI; see Jun 2005b for examples of each tier used in different ToBI systems). Creation of these noncore tiers in various languages exemplifies the flexibility and extendibility of ToBI system as a tool for transcribing language-specific prosodic features as well as other linguistic properties that are useful to the research community or system users.

4.4 Strengths of the ToBI System

The strengths of the ToBI system come from the transcription of tones being phonological, especially based on the AM model of intonational phonology, and from the architecture and mechanism of the transcription system.

First, as a phonological transcription system based on the AM framework of intonational phonology, the ToBI system transcribes distinctive tonal categories of intonation that mark prominence and a hierarchical prosodic structure in a specific language variety. That is, the tones that are transcribed are not just a sequence of F0 turning points on the surface F0 contour, but are contrastive in the language by performing linguistic functions, either marking prominence or information structure, or marking prosodic structure, or both.

Though changes in pitch heavily contribute to the perception of prominence and the degree of juncture, perception of prominence and juncture is influenced by multiple sources of information such as acoustic differences in F0, duration and intensity, voice quality, and even linguistic structures (Cole, Mo, and Baek 2010; Cole, Mo, and Hasegawa-Johnson 2010; Bishop 2012). Because perception can be subjectively influenced by an individual's bias and linguistic background, ToBI requires observation of visual representations of acoustics, especially the F0 track and the waveform/spectrogram of the utterance. This holistic approach to transcription is necessary because ToBI transcription is phonological, capturing the information of metrical and prosodic structure as well as the tonal category. In this way, it is different from other transcription systems focusing on only F0 representations, such as INTSINT (International Transcription System for Intonation; Hirst 1998, and chapter 3, this volume; Hirst and Di Cristo 1998). In INTSINT, an F0 contour is labeled as a sequence of symbols referring to either an absolute tone level (e.g., *t* or ⇑ for the top of the speaker's pitch range) or a relative tone (e.g., *h* or ↑ for a higher tone than the preceding tonal segment), without considering the function of the tone or the prosodic structure of an utterance. Thus, the ToBI tone labels can be equated to broad phonemic transcription of the International Phonetic Alphabet (IPA), while INTSINT symbols correspond to a narrow phonetic transcription of the IPA. That is, the goal of ToBI is not simply to regenerate the surface F0 contours of any sentence in any language, but to capture tonal properties that are linguistically meaningful, especially for the function of marking prominence and prosodic structure.

Because ToBI is language-specific and the tonal categories of each ToBI system are contrastive in each language, defined based on the same theoretical assumptions and principles and using a common vocabulary, comparing the tonal types and the structure of tonal sequences across various ToBI systems allows us to identify intonational universals and study prosodic typology.

For example, as mentioned in sections 4.2 and 4.3.2, the function of ToBI tones, whether they mark prominence or a prosodic structure, can be identified by tonal symbols, especially diacritics. In the ToBI system of languages that have stress, a prominence marking tone is labeled with a diacritic * (e.g., H*, L*+H), called a (postlexical) pitch accent, while tones marking the boundary of a prosodic unit are labeled with an edge-marking diacritic such as %, -, or *a* (to mark a boundary tone of an IP, ip, and AP, respectively). But the * diacritic is also added to a lexical pitch accent, as shown in the Japanese and Serbo-Croatian ToBI systems. Unlike the postlexical pitch accent found in English, the lexical pitch accent in Japanese is not metrically strong, so an accented mora is not realized with higher amplitude or longer duration than an unaccented one (Beckman 1986). Thus, a starred tone marks a syllable or mora's special status as the "head" of a word, but that head is not necessarily associated with stress. If a language has no lexical prosody and has no word head, as in Korean, Mongolian, and West Greenlandic, no starred tone is used to categorize the F0 contour. In these headless languages, distinctive tonal categories are all phrasal tones or boundary tones marking a prosodic unit (e.g., an HLH tonal rhythm in West Greenlandic aligned with the right edge of a phonological word [Arnhold 2014] or a rising boundary tone on the left edge of an intermediate phrase [i.e., -LH] in Mongolian [Karlsson 2014]). In these headless languages, a word becomes prominent by being located at the edge of a larger prosodic unit. In Jun's (2014a) model of prosodic typology, these languages are categorized as edge-prominence languages, while languages like English, where prominence is marked by the head of a prosodic unit, are categorized as head-prominence languages (cf. Beckman 1996).[12] So, comparing diacritics used in the tonal labels of

a ToBI system can provide data for typological analyses of prominence marking and prosodic structure.

Along the same line, by comparing tonal categories across ToBI systems, we can identify the most common versus rare tonal categories, or the most common prosodic units across languages. We can also learn which languages have a contrast on the rising slopes (e.g., H* versus L + H*) or on differences in the alignment of rises (e.g., L + H* versus L* + H) or falls (e.g., H + L*, H* + L), and so on. Languages can also be classified based on the number and type of pitch accents and boundary tones they have, and on how regular their tonal pattern is (for such comparisons, see Jun 2005b, 2014a).

Labeling distinctive categories of prosody by referring to their phonetic realizations in ToBI can also provide excellent data for studying the acquisition of prosody as a second or foreign language or the prosody of languages in contact. This is because one needs to know the distinctive tonal categories and prosodic structure, as well as their phonetic realizations, in both languages to study prosodic transfer or interference between them (e.g., McGory 1997; Ueyama and Jun 1998; Jun and Oh 2000; Dupoux et al. 2008; Atterer and Ladd 2004; Mennen 2004, 2007, 2015; Schepman, Lickley, and Ladd 2006; Trofimovich and Baker 2006; O'Brien and Gut 2010; Lee and Jun 2016). For example, both Japanese and Korean L2 (second language) speakers of English produced the nuclear pitch accent on the focused word in English declarative sentences (i.e., L + H*) better than that in English interrogative sentences (L*). This is because a focused word in Japanese and Korean is always produced with an H tone regardless of the sentence type (McGory 1997; Ueyama and Jun 1998). This suggests that difficulty in producing L2 prosody is due to the differences in the prosodic category between L2 and the speaker's native language (L1). Difficulty in producing L2 prosody can also come from the differences in phonetic realizations of the same prosodic category between L1 and L2. An example can be seen when bilingual speakers of German tend to produce a delayed onset of F0 rise when producing English prenuclear rising pitch accent, because the onset of F0 rise is aligned later in German than in English (Atterer and Ladd 2004). Similarly, both Greek and Dutch have a prenuclear rising pitch accent, but Dutch speakers of Greek often produce an earlier F0 peak when producing Greek prenuclear rising pitch accent, because the F0 peak is aligned later than the accented syllable in Greek than in Dutch (Mennen 1998, 2004). Transcribing the prosody of utterances produced by L2 speakers or bilinguals by referring to the ToBI systems of the two languages involved reveals how the prosodic properties of two languages, either categorical or phonetic, interact in the speech of L2 learners or bilinguals (Lee and Jun 2016).

An additional strength of the ToBI system can come from the flexibility of adding new tiers. Each tier in ToBI can be used to transcribe different prosodic and linguistic information, and the misc tier can even carry nonlinguistic information. As noted in section 4.3.6, ToBI systems for various languages have added new tiers to tag language-specific prosodic information as well as other linguistic information tailored to the particular interests of the research community of the ToBI variety. For example, C-ToBI added a foot tier to transcribe the fusion forms, that is, to mark the degree of lenition between two adjacent syllables, as well as emphasized syllables and phrases. The goal of transcribing syllable fusion was to explore the factors affecting syllable fusion, and the goal of transcribing emphasized syllables was to explore whether the domain of prominence has any relationship with the prosodic hierarchy.[13] The addition of a phonetic tone tier in K-ToBI also allowed researchers to explore the factors affecting the various tonal patterns observed in the AP. By transcribing the allotonic surface tonal categories in the phonetic tone tier for various utterances produced in different speech styles, Yoo

and Jun (2016) recently found that, between the two AP-medial tones, the H tone is realized more frequently than the L tone and is more common in spontaneous speech than in formal speech. A new tier can therefore be added to a ToBI system of a specific language for an exploratory purpose, and what to label in the tier can be decided by the particular group of researchers working on the target language variety.

In addition to testing and improving existing models of intonation by transcribing language-specific information in a new tier, ToBI also represents a useful tool for studying the interfaces between prosody and various subareas of linguistics. Utilizing parallel quasi-independent multitiers aligned with each other, researchers can create and examine ToBI-labeled corpora (e.g, Ostendorf et al. 2001; Grice and Savino 2003) or add ToBI labels to an existing database such as the English Broadcast News Speech (Fiscus et al. 1998) and the Buckeye Corpus (Pitt et al. 2005). For example, the MAE_ToBI-labeled corpus of the Boston University Radio Speech Corpus (Ostendorf, Price, and Shattuck-Hufnagel 1995) was examined to study the realization of glottal stops and glottalization of word-initial vowels in different prosodic positions (e.g., Dilley, Shattuck-Hufnagel, and Ostendorf 1996; Garellek 2013). News speech corpus (Fiscus et al. 1998) and National Public Radio news excerpts have also been labeled in MAE_ToBI to examine the prosodic realization of reflexive pronouns (e.g., *himself, themselves*) relative to their syntactic and semantic structures (Ahn 2015). Similarly, Grice and Savino (2003) analyzed dialogues from the map tasks in Bari Italian by adopting the coding scheme used in the English Human Communication Research Center (HCRC) Map Task (Anderson et al. 1991) and transcribed the intonation of the utterances in the corpus following the ToBI notation in Grice et al. (2005). They discovered that in Bari Italian, polar questions asking about new information employ a rising pitch accent (L + H*), while questions about given information (i.e., confirmation-seeking questions) are expressed with a falling pitch accent (H* + L or H + L* depending on a contrast setting). Furthermore, Stirling et al. (2001) transcribed part of the speech corpora from the Australian National Database of Spoken Language (ANDOSL) map task corpus (Miller et al. 1994) in Australian English ToBI (AuE_ToBI, Fletcher and Harrington 1996) and investigated correlations between discourse structure and various prosodic properties.

Finally, researchers can also use tonal categories and BIs from ToBI transcriptions to test linguistic theories and hypotheses in experimental contexts. ToBI is a particularly useful tool when manipulating prosody for experimental stimuli and for evaluating and categorizing prosodic structures produced by experimental subjects. Perhaps the best illustration of this type of use is research on the role of prosody in sentence processing (e.g., Speer, Kjelgaard, and Dobroth 1996; Schafer et al. 1996; Kjelgaard and Speer 1999; Schafer et al. 2000; Schafer, Speer, and Warren 2005; Ito and Speer 2008; Snedeker and Casserly 2010; Jun 2010; Lee and Watson 2011; Speer and Foltz 2015; Jun and Bishop 2015). These studies explored how manipulations of prosodic structure influence listeners' interpretation and processing of syntactic structure, or, conversely, how their production of prosody changes as a result of manipulating syntactic structure or pragmatic context. Prosodic manipulation and evaluation in these studies was facilitated by the adoption of the tonal categories and/or BIs from the ToBI conventions of the target language. For example, to test Fodor's implicit prosody hypothesis (Fodor 1998, 2002),[14] the prosodic phrasing of sentences produced by subjects (e.g., "Someone shot the servant of the actress who was on the balcony") was assessed by labeling a BI and a boundary tone, if present, after each head noun ("the servant" and "the actress"; see Bergmann, Armstrong, and Maday 2008 and Jun 2010 for English data using MAE_ToBI; Jun and Koike 2008 for Japanese data using J_ToBI; and Jun and

Kim 2004 for Korean data using K-ToBI). Speer and Foltz (2015) also tested the implicit prosody hypothesis by manipulating the presence and type of pitch accent on target words in visual-to-auditory cross-modal priming experiments. Based on ToBI-labeling speech samples taken from paragraphs read aloud, they found that only speakers who reliably produced L+H* pitch accents on contrastively focused words in the produced speech sample (i.e., with explicit prosody) responded faster to an auditory probe word in L+H* following the silent reading of a correctively focused prime word. That is, priming of contrastive intonation from implicit to explicit prosody depended on the speech styles of individual speakers, which were established by way of a ToBI annotation of individual speakers' prosody.

In sum, the ToBI system, based on the AM model of intonational phonology, is an important tool for advancing knowledge on the phonetics and phonology of prosody. It is particularly useful for illuminating the relation between prosody and other linguistic phenomena in a way not possible before the system's creation. As mentioned in Beckman, Hirschberg, and Shattuck-Hufnagel (2005, 46), ToBI is "an ongoing research program rather than a set of rules cast in stone for all time." In addition to facilitating linguistic research, it is a valuable tool for creating communal corpora that can be used to address a variety of research questions. It can also be used by engineers and speech scientists to enhance the performance of systems for recognizing and synthesizing intonation. With ToBI-labeled data, we can investigate how prosodic categories are phonetically realized, and how prosody delivers phonological, syntactic, semantic, and discourse-related information. We can also explore how languages differ in their marking of such complex information prosodically.

4.5 Challenges and Recent Developments

This section describes challenges that ToBI users face as the result of some of the system's weaknesses. While some of these challenges stem from properties of the AM theory that ToBI adopts, others arise from the nature of transcribing real speech using a discrete phonological transcription system. This section also describes some of the recent developments the ToBI community has devised to address ToBI's weaknesses.

One of the challenges ToBI users encounter is that the prominence patterns and prosodic structure captured by ToBI transcriptions are limited to the size of an utterance corresponding to an IP. This is a consequence of the fact that, in the AM model of intonational phonology, the IP is the highest prosodic unit defined by a tone.[15] Therefore, prosodic events that occur across an IP boundary or over a sequence of IPs cannot be captured by the current ToBI system. This makes it difficult to use ToBI to capture how prosody is used to mark complex discourse structure, or possible syntagmatic relations between tones across IPs (e.g., Ladd 1990, [1996] 2008a; Calhoun 2006, 2012; Bishop 2013).[16] One of the main prosodic cues known to mark discourse structure is pitch range (Brown, Currie, and Kenworthy 1980; Hirschberg and Pierrehumbert 1986). While ToBI includes pitch reset among the acoustic cues defining a prosodic unit, no labels exist to cover pitch range changes across ips and IPs. Mandarin ToBI (Peng et al. 2005) includes a limited number of labels for pitch range, but they refer to effects within one sentence or phrase (i.e., one breath group): that is, %reset to mark pitch reset over a declarative phrase, %q-raise to mark raised pitch range (as in echo questions), %e-prom for the local expansion of pitch range within a phrase due to emphatic prominence, and %compressed for the local compression of pitch range after %e-prom. The MAE_ToBI system does include the label HiF0, intended to mark the

highest F0 point for a pitch accent within an ip. But this was simply intended as a tag for accessing the F0 value for use later as a measure of pitch range to study the relationship between pitch range and discourse structure. Instead of using a tones tier, a new tier could be added for pitch-range information. J_ToBI, for example, added an extra tier to mark finality. Though the finality tier for J_ToBI only marks a juncture larger than an IP (delivering "the sense that the speaker is 'done' in her turn in discourse planning"; Venditti 1997, 2005), this type of tier could be used to label acoustic cues to the relationship between IPs and larger discourse segments.

Because most of the existing ToBI systems were developed on the basis of sentence-length laboratory speech with relatively simple discourse structure (e.g., reading of a story, news reports, interviews, short dialogues, speech from map tasks), ToBI transcriptions have not been widely used by researchers working on the discourse structure of free conversational speech. However, as noted in Beckman, Hirschberg, and Shattuck-Hufnagel (2005, 12), transcription conventions for ToBI are ideally "based on a large and long-established body of research on dialectology, pragmatics and discourse analysis for the target language," in addition to study of its intonational phonology. The inclusion of a tier that allows for annotating relationships between ips and IPs (or between larger discourse segments) would thus be a straightforward way to address ToBI's current weakness in this area.

In a related vein, the current ToBI system is also not optimized for labeling global, and gradual, prosodic events. This stems from basic properties of the AM model of intonation, which represents the intonation contour as a linear sequence of low and high tonal targets. If a researcher is more interested in the domain of global prosodic events (e.g., pitch range reset at the beginning of prosodic unit X or Y) or relative differences between domains (e.g., whether the pitch range of unit X is larger or smaller than that of unit Y), this could be addressed by creating a label to mark the onset or magnitude of the prosodic feature at the edges of units. However, modifications that require a researcher to address more gradual changes over domains present a much greater challenge to the ToBI system (e.g., annotating changes in the slope of declination, which marks sentence types in Danish intonation; Thorsen 1980, 1983).

Given the variability and gradience inherent in acoustic signals, the next challenge ToBI users face is not due to the system's reliance on the AM model specifically but to its basic status as phonological. First, prosodic categories are not always realized in the same way. As Cole and Shattuck-Hufnagel (2016), Cangemi and Grice (2016), and Arvaniti (2016) have emphasized, acoustic cues to a prosodic category are not always realized consistently. Instead, factors such as time pressures and segmental/prosodic context often render cues reduced or ambiguous, and prosodic categories can also be realized differently depending on the individual speaker's strategy to encoding prosodic contrasts (e.g., Mo, Cole, and Lee 2008; Cole and Shattuck-Hufnagel 2016). Moreover, criteria for defining prosodic categories, such as phrase-final lengthening for an IP or ip boundary, are gradient rather than categorical. Categorizing a prosodic event based on gradient and variable acoustic cues is not straightforward in the absence of other information supporting the category, and the final decision can be further influenced by the individual labeler's weighting of the acoustic cues.

Second, and related to this, recall from section 4.4 how ToBI labels are decided based on the labeler's perception of prominence and degree of juncture from an audio file, while at the same time observing the visual display of a pitch track together with a spectrogram or waveform. This means that the labeling is not fully based on objective criteria and can be influenced by the labeler's perception of acoustic cues, interpretation

of linguistic context, and level of knowledge and experience in ToBI labeling. It is therefore recommended that ToBI labeling be done by multiple trained labelers to reduce the subjectivity of the annotations. However, because ToBI labeling is labor intensive and requires training in both acoustics and AM theory, it is not always easy to find multiple ToBI labelers to participate in the annotation of experimental data or databases.

The fact that ToBI systems are tools for phonological transcription means they are well suited to studying prosodic typology. One of the strengths of a phonological transcription system is that it enables comparison of how different languages (and language varieties) tonally mark prominence and prosodic structure. However, this also comes with a challenge. Because ToBI's tone labels are language-specific, reflecting contrasts in each language, the same surface F0 contour is not necessarily assigned the same labels across languages (just as the same Voice Onset Time [VOT] can be categorized as /p/ in Spanish but as /b/ in English). Additionally, as pointed out by Ladd (2008b) and Hualde and Prieto (2016), it is also possible that the same surface F0 contours, representing the same contrast in each language, can be labeled differently depending on what analysis and tonal categories are chosen by the developers of the ToBI system.

A good example illustrating this point can be found in the comparison of MAE_ToBI and GToBI (Grice et al. 2005). Both American English and Standard German have four levels of sentence-final pitch, representing the same basic meaning/function: low F0 at the end of a neutral statement, mid F0 at the end of a calling contour, high flat F0 for incompleteness, and super-high F0 at the end of a yes-no question. However, these four types of contours are transcribed differently: as L-L%, !H-L%, H-L%, and H-H% in MAE_ToBI, but as L-%, !H-%, H-%, and H-^H% in GToBI. This is because English ToBI developers adopted an upstep rule (i.e., an IP boundary tone is realized higher after a (!)H- phrase accent), but GToBI developers did not. Instead, GToBI's creators added two symbols intended to make their system for German more phonetically transparent. One symbol, an upstep diacritic ^, is used to directly mark a super-high tone, ^H; another, a toneless IP boundary symbol %, is used when the IP-final pitch level is the same as the ip-final pitch (i.e., H-% means a high plateau and L-% a low plateau). Thus the difference between these two ToBI systems was a matter of choosing how abstract versus how phonetically transparent the tonal representation should be. Because the tones in the original ToBI system, by introducing a downstep diacritic, were claimed to be more faithful to the surface F0 patterns than those proposed in the original AM model (see section 4.3.2), various ToBI systems have added tones and symbols to better capture surface F0 contours, but there are still no widely agreed-on rules or conventions regarding what symbols and labels to use and on what occasions to use them. (See the "portability" problem of ToBI notation argued in Hualde and Prieto 2016.) As more ToBI systems are developed in the future, there will be new symbols and labels created, making comparison across ToBI systems more difficult and the study of prosodic typology thus more challenging. Students and researchers will also likely find it more difficult to determine what some labels and symbols in different ToBI systems mean[17] and what level of abstraction they represent.[18]

In addition to the difficulty in interpreting symbols and levels of abstraction in different ToBI systems generally, there is also the challenge of deciding what level of abstraction is appropriate when a surface tone does not match an underlying tone. Should it be transcribed with an allophonic tone label or with an underlying tone label? As a system designed to transcribe phonological categories, clearly the underlying tone should be used. However, if the goal is simply to tag one's observations about the speech signal, one might want to use an allophonic tone label. Cases like this

have been a known source of confusion and disagreement among labelers (e.g., Escudero et al. 2012; Armstrong 2017; Hualde and Prieto 2016), and the chosen solutions vary across ToBI systems. For example, when an underlying L% boundary tone is truncated after an L+H* on a phrase-final syllable (and thus realized as an approximately mid-level tone), should the boundary tone be labeled as L%, !H%, or something else? Prieto and Ortega-Liebaria (2009) and Hualde and Prieto (2016) used !H% in Catalan and Peninsular Spanish ToBIs, but Grice et al. (2005) used (L%)—L% in parentheses—for southern Italian varieties, as a notation for a partially realized underlying tone.

As a tagging tool, the ToBI system can include labels that represent surface tonal patterns, and this makes its level of tonal representation clearly less abstract than that proposed in the AM model of intonational phonology. However, the common practice of using ToBI-style labeling (employing primarily the words and tones tiers) when trying to build a model of intonational phonology of a language seems to have blurred the distinction between the ToBI system and the AM model of intonational phonology. This trend can be observed in publications such as Prieto and Roseano (2010), Jun (2014b), and Frota and Prieto (2015), where most models of intonational phonology and most ToBI transcription systems look quite similar to each other. Consequently, the BI part of ToBI has also become less emphasized in the ToBI system. In fact, it is not uncommon to hear researchers refer to the "ToBI model" of a particular language, which would seem to equate the transcription system with the intonational phonology model of the language.[19]

Many of the issues and challenges described herein motivated a recent proposal to develop an international prosodic alphabet (IPrA) within the AM framework (Hualde and Prieto 2016 and Jun and Fletcher 2014; see also presentations by Jun, Prieto, and Hualde at the satellite workshop of the International Congress of Phonetic Sciences 2015 in Glasgow[20]). The current proposal calls for ToBI systems to include two levels of tonal transcription—one distinctive, thus phonological, and the other nondistinctive but categorical, thus phonetic—and to create a set of discrete IPrA tonal labels and diacritics (the details of which are to be agreed on by the international community of ToBI users). The goal is that the symbols in IPrA can be used for both phonological and categorical phonetic transcriptions, like the symbols in the IPA, so that they can be used in both phonemic and phonetic representations. The labels for *phonological* transcription would only mark phonological contrast in the language, and the *categorical phonetic* labels would be for tones that are categorical in nature but not distinctive, or whose distinctiveness is not yet known. These categorical phonetic labels can (i) represent allophonic realizations of an underlying tonal category, (ii) be used as temporary labels before their phonological status is established (e.g., at the beginning stage of building a model of intonational phonology), or (iii) represent hybrid or exceptional tonal categories that are not part of the intonational model of any specific language (e.g., L2 speech, bilinguals' speech, or speech produced by speakers of multiple languages in contact). The phonological tone labels can be used in the tones tier, and categorical phonetic tones can be labeled in a new, separate tier (e.g., an IPrA tier) or can be used in the tones tier together with the phonological tones, but written in square brackets (e.g., [L*] for an allophonic L*). It is hoped that separating categorical phonetic tones from phonological tones in ToBI will facilitate the development of more abstract phonological analyses of intonation and better represent the correspondence between underlying tonal categories and surface patterns in a systematic way. This will help us to clarify the levels of transcription that different ToBI systems are using and make more transparent comparisons across languages possible. Clarifying the levels of transcription will also

increase interlabeler agreement rates and facilitate labeling large corpora and further improve the strengths of the ToBI system mentioned in section 4.4.

In addition to the IPrA proposal, some of the weaknesses of the ToBI system motivated the recent development of a few other transcription systems such as rapid prosody transcription (RPT) and rhythm and pitch (RaP) transcription. The RPT system was developed by Cole and her colleagues (Mo, Cole, and Lee 2008; Cole, Mo, and Baek 2010; and Cole, Mahrt, and Hualde 2014). Unlike ToBI transcription, which is labor-intensive and requires training and knowledge of the theoretical background, the RPT transcription is, as the name implies, a coarse-grained transcription done by untrained naive native speakers of the target language.[21] Because this type of annotation can be done as a group, it is possible to generate transcriptions from many labelers in a very short time. In the RPT task, untrained native speakers (such as college students) are given instructions defining "prominent" words and "prosodic juncture" in running speech. They are then asked to listen to a sample of running speech and identify prominent words and prosodic boundaries on a printed or electronically presented transcript of the speech sample. Identification of these prosodic events is done in real time (i.e., the "rapid" part of RPT), with the subjects typically listening to the entire speech sample four times, twice each for prominence and boundary transcription (they are not allowed to replay any specific parts of the sample). The length of speech samples used can vary from ten to sixty seconds, which generally includes one or more syntactic clauses and one or more prosodic phrases. In Cole and colleagues' studies, the task was used to generate a prominence score (p-score) and a boundary score (b-score) for each word (i.e., the proportion of labelers who marked each word as prominent or as preceding a boundary, respectively). The magnitude of p-scores and b-scores for individual words and their distribution thus provides a measure of agreement among labelers and an estimation of the variability in the perception of prosodic events in the speech material (Mo, Cole, and Lee 2008; Cole and Shattuck-Hufnagel 2016).

Bishop and his colleagues (Bishop and Kuo 2016; Bishop, Kuo, and Kim 2020) compared the responses from an RPT task with the labels provided by expert ToBI annotators. Their results showed that RPT subjects identified just under 50% of the nuclear pitch accents (NPAs) and just under 50% of the IP boundaries that ToBI labelers assigned to the speech materials. Thus a little more than half of the time, the naive labelers failed to identify NPA-level prominence and IP-level boundaries. This is not surprising given that the RPT labelers were not trained to attend to prominence and phrasing at a detailed level, and the transcription was done in real time without examining pitch tracks or being able to hear parts of the sentence as many times as desired. However, from the words identified as prominent or having a prosodic boundary by the RPT labelers, pitch-accented words were identified as prominent more often (i.e., had higher p-scores) than unaccented words, and words at an IP boundary were identified as a boundary more often (i.e., had higher b-scores) than those at an ip boundary. This suggests that the p- and b-scores from a large set of speech materials annotated by a large group of RPT labelers can provide a good measure of the perceptual salience of prosodic categories. RPT may also provide useful information about the acoustic correlates of various prosodic categories, as well as about how the perception of prominence and boundaries is influenced by grammatical information. As pointed out in Cole and Shattuck-Hufnagel (2016), the patterning of the p- and b-scores from RPT may shed light on the mechanisms involved in prosodic processing in speech production and perception.

Next, the RaP system was developed by Dilley and her colleagues (Dilley 2005; Dilley and Brown 2005; Dilley et al. 2006; Breen et al. 2012) to address transcription difficulties associated with ToBI's vulnerability to the variability and gradience of tonal categories and to reflect the importance of rhythmic or metrical prominence distinct from pitch accent. In ToBI, tonal labels such as H* or L-H% include both the identity of the tonal target (e.g., H or L or !H) and information regarding its function in terms of both metrical prominence (i.e., pitch accents are explicitly marked by *) and prosodic structure (i.e., boundary tones are marked by % and -). The RaP system, in contrast, emphasizes the independence of pitch information from rhythmic/metrical and phrasal structure. Pitch information is transcribed in terms of three tonal targets (H, L, E) whose value is relative to a preceding tone (higher, lower, or equal to the preceding tonal labels, respectively). Tones in RaP, labeled in the pitch tier, are therefore represented phonetically. Metrical prominence (in three levels: strong, weak, none)[22] and prosodic structure (in two levels: IP and ip) are transcribed in the rhythm tier. Although RaP was proposed as "a method of labeling the rhythm and relative pitch of spoken English" (Dilley and Brown 2005, 2), the concept and principles of the system can be applied to other languages, and the system can be used by researchers interested in examining structural information separately from tonal information.

4.6 Conclusion

This chapter introduced the ToBI system for transcribing phonological aspects of prosody. It discussed both the strengths and challenges associated with the system, as well as some of the recent developments the research community has made in response to such challenges. As a phonological annotation system based on the AM model of intonational phonology, each language's ToBI system is unique, reflecting the prosodic properties of the language in question. The ToBI framework allows a common set of conventions to be used across languages, despite the fact that how prominence and phrasal structure are phonologically marked and phonetically realized is language-specific. In the ToBI system, tones and BIs are annotated based on the labeler's perceived prominence of words and degree of juncture between words, supplemented by the visual display of the utterance's acoustics, especially F0 track and waveforms (or spectrogram). Furthermore, as a tool to annotate and communicate observations made about the speech signal with other users of ToBI, the ToBI system allows communal corpora to be created for testing hypotheses about various linguistic and prosodic phenomena. The output of such research also serves as feedback to the ToBI transcription system itself and informs models of intonational phonology.

It was shown that most of the strengths as well as most of the weaknesses associated with ToBI stem from its being a phonological transcription system based on the AM model. The ToBI system's emphasis on transcribing distinctive phonological categories of tones and metrical/prosodic structure makes it the most valuable tool currently available for studying the grammar of prosody and intonation. ToBI users can pursue questions about prosody's relation to a number of other linguistic subfields, and it is especially well-suited to the study of prosodic typology and universals. However, as a phonological system, the labels proposed in various ToBI systems do not necessarily have the same phonetic values or represent the same level of abstraction from system to system, and so an effort should be made when comparing multiple ToBI systems to study prosodic typology. Efforts to address this particular issue are currently underway,

such as the recent proposal to develop an IPrA, allowing for the categorical phonetic representation of tones in each ToBI system.

ToBI is a valuable tool for conducting research on prosody, and it is a powerful and flexible transcription system. With approximately fifty ToBI systems now established from typologically different languages, tremendous progress has been made since the original ToBI system was published in 1994, in terms of both prosody's relation to other linguistic subfields and our understanding of prosodic typology. Although ToBI transcription is labor-intensive, and becoming a confident labeler requires significant training, the benefits seem clear and justified. Ongoing and future tasks for the ToBI research community include improving on the system's current limitations, as well as continuing to expand the number (and diversity) of languages described within the ToBI framework. Taken together, the ToBI framework represents a powerful and flexible tool, advancing our understanding of prosodic systems and prosodic typology.

Notes

1. Most of the ToBI systems presented at the meeting, including the MAE_ToBI, were later published in Jun (2005c).

2. As will be shown in sections 4.3.2 and 4.4, not all languages mark intonational prominence by way of pitch accent. If a language has no stress or lexically marked syllables—and thus has no designated terminal element for words (e.g., Korean, Mongolian, or West Greenlandic), prominence is not marked by a pitch accent but by positioning the prominent word at the edge of a prosodic unit. Thus how prosodic prominence is marked varies across languages. See Jun (2014a) for a typology of prominence marking.

3. This prosodic structure is similar to, but slightly different from, the prosodic hierarchy proposed by prosodic phonologists in the 1980s, which was defined indirectly based on the syntactic structure of a sentence (e.g., Nespor and Vogel 1986; Selkirk 1986; Hayes 1989). In that approach, prosodic units commonly assumed to be larger than a word include, from the largest to the smallest, an utterance (Utt) > an intonational phrase (IPh) > a phonological phrase (PPh) > clitic group. Both an Utt and an IPh correspond to an IP in intonational phonology, but a PPh corresponds to either an ip or an AP in intonational phonology (see Jun 1998 for comparisons between these two approaches).

4. The H*+L pitch accent proposed in Pierrehumbert (1980) and Beckman and Pierrehumbert (1986) is not a falling tone as implied by the trailing L tone. As mentioned in Beckman, Hirschberg, and Shattuck-Hufnagel (2005), it is for a simple H* pitch accent, followed by a downstepped H*. The trailing L was simply included to provide the necessary bitonality required, as shown in the research on African tone languages (e.g., Leben 1973; Goldsmith 1976), to trigger downstep of a following H tone. That is, a sequence of H*+L pitch accents was used to refer to a series of stepping-down pitch accents, examined in Liberman and Pierrehumbert (1984).

5. Allowing tonal labels to be more faithful to the surface F0 contour became the most prominent difference between the tonal inventory of the AM model of intonational phonology and that of the ToBI system. As will be discussed in section 4.5, this difference was a source of confusion to many researchers and significantly affected subsequent models of the intonational phonology of other languages.

6. A prosodic word can also be marked by intonation as in Serbo-Croatian (Godjevac 2005).

7. Because the undershoot allophones of an L boundary tone are predictable from the words tier in J_ToBI, both wL% and %wL symbols are removed in X-JToBI (Maekawa et al. 2002), a revised version of J_ToBI.

8. If the AP-initial segment is either an aspirated or a tense consonant, the tone is H. Otherwise, the tone is L.

9. The GToBI had the same number of prosodic units as in MAE_ToBI, but the GToBI developers proposed not to label a BI lower than 3 and proposed to label two types of mismatch cases, labeled 2r and 2t. BI 2r refers to rhythmic break with tonal continuity, and BI 2t refers to tonal break with rhythmic continuity.

10. In the original ToBI, filler words (such as *uh* and *um*, also known as filled pauses) were not treated as a type of disfluency. But this and other types of disfluency could be added in other ToBI systems.

11. When a labeler is sure about the degree of juncture at a certain word boundary but the tonal realization does not match the corresponding degree of juncture, a mismatch label (2 or #m, discussed in section 4.3.3) should be used.

12. Languages can cue a word's prominence by marking both the head and the edge of the word, and this type of language is categorized as a head-/edge-prominence language (examples are Georgian, Bengali, French, and Japanese). The tonal inventory of these languages includes a starred tone as well as a boundary tone of a word or AP (see Jun 2014a).

13. In C-ToBI, the domain of the foot is reflected in the BI tier by aligning the foot-medial weakened syllable with a BI 0, while the foot-final syllable is aligned with a BI 1. However, the domain of prominence is not linked with the BI tier, reflecting the unknown relationship between the domain of prominence and the prosodic hierarchy. The developers of C-ToBI state that the architecture of independent parallel tiers "highlights one of the principal advantages of the looser structure of the original ToBI framework" (Wong, Chan, and Beckman 2005, 293).

14. The hypothesis was proposed to explain cross-linguistic differences in attachment preferences. Fodor claimed that the difference is due to implicit default prosody assigned to the syntactic structure of each sentence. She predicted that, in the example sentence where the attachment of a relative clause (RC) is ambiguous between the first noun ("the servant") or the second ("the actress"), languages that prefer high attachment (i.e., RC attachment to the first noun) would have a larger juncture after the second noun than after the first noun, and the opposite is true for low attachment preference languages. She also assumed implicit prosody generated in silent reading is equal to explicit prosody generated in reading aloud.

15. The AM model does not include a prosodic unit above the IP, such as the utterance-level unit proposed by some prosodic phonologists (e.g., Nespor and Vogel 1986). The primary reason for this is the absence of distinctive tones or other categorical prosodic events that mark any unit above the IP.

16. In the AM model of English intonation adopted in the MAE_ToBI, an ip is the domain of nuclear pitch accent and the domain of pitch reset. Thus, capturing prominence relation across ips is limited.

17. Examples where ToBI systems assign different symbols for the same phenomenon are easily found among the systems currently available. For example, undershoot has been marked with either *w* (in Japanese and Greek ToBIs) or *u* (in Bangladesh Bengali ToBI), and an upstep has been marked with ^ (in German and Bininj Gun-wok ToBI) or with ¡ (in Spanish and Catalan ToBIs).

18. For example, downstep can be phonological or phonetic across languages. Its status can also vary within a language depending on tone types (e.g., in GRToBI, downstepped H is contrastive for a phrase accent (!H-) or a boundary tone (!H%) but not for a pitch accent (e.g., !H* or !H* + L).

19. In this case, a "ToBI model" would seem to refer to an intonational phonology model that is more transparent and surface-rich than the traditional AM model of intonational phonology.

20. For more information about the workshop, visit http://linguistics.ucla.edu/ipra_workshop/.

21. The RPT methods have been used for Hindi, Russian, French, and Spanish in addition to English (for some cautionary notes on applying the RPT across languages, however, see Gussenhoven 2015). It has also been used with fluent bilinguals and learners of English as L2 (see Cole and Shattuck-Hufnagel 2016).

22. In English ToBI, metrical prominence is also represented at three levels (Beckman & Edwards 1994): NPA, pitch accent, no accent. The NPA is the last pitch accent within an ip. Because it is predictable from the location of pitch accent and an ip boundary tone, NPA is not labeled with a separate symbol in English ToBI. But in languages where NPA is not predictable as in Italian, an NPA tone is marked by a diacritic *n* (e.g., L*+Hn).

References

Ahn, B. 2015. "Giving Reflexivity a Voice: Twin Reflexives in English." PhD diss., UCLA.

Anderson, A., M. Bader, E. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, et al. 1991. "The HCRC Map Task Corpus." *Language and Speech* 34:351–366.

Arbisi-Kelm, T. 2006. "An Intonational Analysis of Disfluency Patterns in the Speech of Stutterers." PhD diss., UCLA.

Armstrong, M. E. 2017. "Accounting for Intonational Form and Function in Puerto Rican Spanish Polar Questions." *Probus* 29 (1): 1–40. https://doi.org/10.1515/probus-2014-0016.

Arnhold, A. 2014. "Prosodic Structure and Focus Realization in West Greenlandic." In Jun 2014b, 216–251.

Arvaniti, A. 2016. "Analytical Decisions in Intonation Research and the Role of Representations: Lessons from Romani." *Laboratory Phonology* 7 (1): 1–43.

Arvaniti, A., and M. Baltazani. 2005. "Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora." In Jun 2005c, 84–117.

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. Broe and J. Pierrehumbert, 119–131. Cambridge: Cambridge University Press.

Atterer, M., and D. R. Ladd. 2004. "On the Phonetics and Phonology of "Segmental Anchoring" of F0: Evidence from German." *Journal of Phonetics* 32: 177–197.

Beckman, M. E. 1986. *Stress and Non-Stress Accent*. Dordrecht, the Netherlands: Foris.

Beckman, M. E. 1996. "The Parsing of Prosody." *Language and Cognitive Processes* 11:17–67.

Beckman, M. E., and G. M. Ayers Elam. 1997. "Guidelines for ToBI Labeling." Unpublished manuscript. Ohio State University. https://www.ling.ohio-state.edu/research/phonetics/E_ToBI/.

Beckman, M. E., M. Díaz-Campos, J. T. McGory, and T. A. Morgan, 2002. "Intonation across Spanish, in the Tones and Break Indices Framework." *Probus* 14:9–36.

Beckman, M. E., and J. Edwards. 1994. "Articulatory Evidence for Differentiating Stress Categories." In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, edited by P. A. Keating, 7–33. Cambridge: Cambridge University Press.

Beckman, M. E., and J. Hirschberg. 1994. "The ToBI Annotation Conventions." Unpublished manuscript. Ohio State University. https://web.archive.org/web/20100622195855/http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/ToBI/ToBI.6.html.

Beckman, M. E., J. Hirschberg, and S. Shattuck-Hufnagel. 2005. "The Original ToBI System and the Evolution of the ToBI Framework." In Jun 2005c, 9–54.

Beckman, M. E., and S.-A. Jun. 1996. "K-ToBI (Korean ToBI) Labelling Convention." Version 2. Unpublished manuscript. Ohio State University and UCLA. https://linguistics.ucla.edu/people/jun/ktobi/k-tobi-V2.html.

Beckman, M. E., and J. Pierrehumbert. 1986. "Intonational Structure in Japanese and English." *Phonology Yearbook* 3:255–309.

Bergmann, A., M. Armstrong, and K. Maday, K. 2008. "Relative Clause Attachment in English and Spanish: A Production Study." In *Proceedings of the Fourth International Conference on Speech Prosody*. https://www.isca-speech.org/archive/sp2008/papers/sp08_505.pdf.

Bishop, J. 2012. "Information Structural Expectations in the Perception of Prosodic Prominence." In *Prosody and Meaning (Interface Explorations*), edited by G. Elordieta and P. Prieto, 239–269. Berlin: Mouton de Gruyter.

Bishop, J. 2013. "Prenuclear Prominence: Phonetics, Phonology, and Information Structure." PhD diss., UCLA.

Bishop, J., and J. Fletcher. 2005. "Intonation in Six Dialects of Bininj Gun-wok." In Jun 2005c, 331–361.

Bishop, J., and G. Kuo. 2016. "Do 'Autistic-Like' Personality Traits Predict Prosody Perception?" Paper presented at LabPhon 15 (Workshop on Personality in Speech), Cornell University, Ithaca, New York, July 17.

Bishop, J., G. Kuo, and B. Kim. 2020. "Phonology, Phonetics, and Signal-Extrinsic Factors in the Perception of Prosodic Prominence: Evidence from Rapid Prosody Transcription." *Journal of Phonetics* 82. https://doi.org/10.1016/j.wocn.2020.100977.

Bolinger, D. L. 1951. "Intonation: Levels versus Configurations." *Word* 7:199–210.

Bolinger, D. L. 1958. "A Theory of Pitch Accent in English." *Word* 14:109–149.

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.

Brown, G., K. Currie, and J. Kenworthy 1980. *Questions of Intonation*. London: Croom-Helm.

Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective*. Lund, Sweden: Gleerup.

Brugos, A., N. Veilleux, M. Breen, and S. Shattuck-Hufnagel. 2008. "The Alternatives (Alt) Tier for ToBI: Advantages of Capturing Prosodic Ambiguity." In *Proceedings of the Fourth International Conference on Speech Prosody*. https://www.isca-speech.org/archive/sp2008/papers/sp08_273.pdf.

Calhoun, S. 2006. "Information Structure and the Prosodic Structure of English: A Probabilistic Relationship." PhD diss., University of Edinburgh.

Calhoun, S. 2012. "The Theme/Rheme Distinction: Accent Type or Relative Prominence?" *Journal of Phonetics* 40:329–349.

Cangemi, F., and M. Grice. 2016. "The Importance of a Distributional Approach to Categoriality in Autosegmental-Metrical Accounts of Intonation." *Laboratory Phonology* 7 (1): 1–20.

Cole, J., T. Mahrt, and J. I. Hualde. 2014. "Listening for Sound, Listening for Meaning: Task Effects on Prosodic Effects on Prosodic Transcription." In *Proceedings of the Seventh International Conference on Speech Prosody*. https://www.isca-speech.org/archive/SpeechProsody_2014/pdfs/165.pdf.

Cole, J., Y. Mo, and S. Baek. 2010. "The Role of Syntactic Structure in Guiding Prosody Perception with Ordinary Listeners and Everyday Speech." *Language and Cognitive Processes* 25:1141–1177.

Cole, J., Y. Mo, and M. Hasegawa-Johnson. 2010. "Signal-Based and Expectation-Based Factors in the Perception of Prosodic Prominence." *Laboratory Phonology* 1:425–452.

Cole, J., and S. Shattuck-Hufnagel. 2016. "New Methods for Prosodic Transcription: Capturing Variability as a Source of Information." *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7 (1): 8. http://dx.doi.org/10.5334/labphon.29.

Cooper, W., and P. Paccia-Cooper. 1980. *Syntax and Speech*. Cambridge, MA: Harvard University Press.

Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Delais-Roussarie, E., B. Post, M. Avanzi, C. Buthke, A. Di Cristo, I. Feldhausen, S.-A. Jun, et al. 2015. "Intonational Phonology of French: Developing a ToBI System for French." In *Intonation in Romance*, edited by Sonia Frota and Pilar Prieto, 63–100. Oxford: Oxford University Press.

Dilley, L. C. 2005. "The Phonetics and Phonology of Tonal Systems." PhD diss., MIT.

Dilley, L. C., M. Breen, M. Bolivar, J. Kraemer, and E. Gibson. 2006. "A Comparison of Inter-Transcriber Reliability for Two Systems of Prosodic Annotation: RaP (Rhythm and Pitch) and ToBI (Tones and Break Indices)." In *Proceedings of Interspeech 2006*. https://www.isca-speech.org/archive/archive_papers/interspeech_2006/i06_1619.pdf.

Dilley, L., and M. Brown, M. 2005. "The RaP (Rhythm and Pitch) Labeling System." Version 1.0. https://pdfs.semanticscholar.org/5f73/1dbcafb2b64da6eb15daa67718866bc74cc9.pdf.

Dilley, L., S. Shattuck-Hufnagel, and M. Ostendorf. 1996. "Glottalization of Word-Initial Vowels as a Function of Prosodic Structure." *Journal of Phonetics* 24:423–444.

Dupoux, E., N. Sebastián-Gallés, E. Navarrete, and S. Peperkamp. 2008. "Persistent Stress 'Deafness': The Case of French Learners of Spanish." *Cognition* 106:682–706.

Escudero, D., L. Aguilar, M. del M. Vanrell, and P. Prieto. 2012. "Analysis of Inter-Transcriber Consistency in the Cat_ToBI Prosodic Labelling System." *Speech Communication* 54:566–582.

Face, T., and P. Prieto. 2007. "Rising Accents in Castilian Spanish: A Revision of Sp_ToBI." *Journal of Portuguese Linguistics* 6:117–146.

Fiscus, J., J. Garofolo, M. Przybocki, W. Fisher, and D. Pallett. 1998. *English Broadcast News Speech (HUB4)*. Philadelphia: Linguistic Data Consortium.

Fletcher, J., and J. Harrington. 1996. "Timing of Intonational Events in Australian English." In *Proceedings of the Sixth Australian International Conference on Speech Science and Technology*,

edited by P. McCormack and A. Russell, 611–615. Australian Speech Science and Technology Association, Canberra, Australia.

Fodor, J. D. 1998. "Learning to Parse." *Journal of Psycholinguistic Research* 27:285–319.

Fodor, J. D. 2002. "Prosodic Disambiguation in Silent Reading." *Proceedings of the North East Linguistic Society* 32:113–132.

Frota, S. 2014. "The Intonational Phonology of European Portuguese." In Jun 2014b, 7–42.

Frota, S., P. Oliveira, M. Cruz, and M. Vigário. 2015. *P-ToBI: Tools for the Transcription of Portuguese Prosody*. Lisbon: Laboratório de Fonética, CLUL/FLUL. http://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/.

Frota, S., and P. Prieto. 2015. *Intonation in Romance*. Oxford: Oxford University Press.

Garellek, M. 2013. "Production and Perception of Glottal Stops." PhD diss., UCLA.

Gee, J. P., and F. Grosjean. 1983. "Performance Structures: A Psycholinguistic and Linguistic Appraisal." *Cognitive Psychology* 15:411–458.

Godjevac, S. 2005. "Transcribing Serbo-Croatian Intonation." In Jun 2005c, 146–171.

Goldsmith, J. 1976. "Autosegmental Phonology." PhD diss., MIT.

Gordon, M. 2005. "Intonational Phonology of Chickasaw." In Jun 2005c, 301–330.

Grice, M., S. Baumann, and R. Benzmueller. 2005. "German Intonation in Autosegmental-Metrical Phonology." In Jun 2005c, 55–83.

Grice, M., and R. Benzmueller. 1995. "Transcription of German Intonation Using ToBI-Tones—The Saarbrücken System." *Phonus* 1:33–51.

Grice, M., M. D'Imperio, M. Savino, and C. Avesani. 2005. "Strategies for Intonation Labeling across Varieties of Italian." In Jun 2005c, 362–389.

Grice, M., and M. Savino. 2003. "Map Tasks in Italian: Asking Questions about Given, Accessible and New Information." *Catalan Journal of Linguistic* 2:153–180.

Gussenhoven, C. 2015. "Does Phonological Prominence Exist?" *Lingue e Linguaggio* 14:7–24.

Halliday, M. A. K. 1967. *Intonation and Grammar in British English*. The Hague, the Netherlands: Mouton.

Hayes, B. 1989. "The Prosodic Hierarchy in Meter." In *Perspectives on Meter*, edited by P. Kiparsky and G. Youmans, 203–260. New York: Academic Press.

Hirschberg, J., and J. Pierrehumbert. 1986. "Intonational Structuring of Discourse." *Proceedings of the 24th Meeting of the Association for Computational Linguistics*, 126–144.

Hirst, D. 1998. "Intonation in British English." In *Intonation Systems*, edited by D. Hirst and A. Di Cristo, 56–77. Cambridge: Cambridge University Press.

Hirst, D., and A. Di Cristo. 1998. "A Survey of Intonation Systems." In *Intonation Systems*, edited by D. Hirst and A. Di Cristo, 1–44. Cambridge: Cambridge University Press.

Hualde, J., and P. Prieto. 2016. "Towards an International Prosodic Alphabet (IPrA)." *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7 (1): 1–25, especially 5. http://dx.doi.org/10.5334/labphon.11.

Indjieva, E. 2009. "Oirat Tones and Break Indices (O-ToBI). Intonational Structure of the Oirat Language." PhD diss., University of Hawaii.

Ito, K., and S. R. Speer. 2008. "Anticipatory Effect of Intonation: Eye Movements during Instructed Visual Search." *Journal of Memory and Language* 58:541–573.

Jun, S.-A. 1998. "The Accentual Phrase in the Korean Prosodic Hierarchy." *Phonology* 15 (2): 189–226.

Jun, S.-A. 2000. "K-ToBI (Korean ToBI) Labeling Conventions, Version 3.1." *UCLA Working Papers in Phonetics* 99:149–173.

Jun, S.-A. 2005a. "Korean Intonational Phonology." In Jun 2005c, 201–229.

Jun, S.-A. 2005b. "Prosodic Typology." In Jun 2005c, 430–458.

Jun, S.-A., ed. 2005c. *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Jun, S.-A. 2010. "The Implicit Prosody Hypothesis and Overt Prosody in English." *Language and Cognitive Processes* 25:1201–1233.

Jun, S.-A. 2014a. "Prosodic Typology: By Prominence Type, Word Prosody, and Macro-Rhythm." In Jun 2014b, 520–539.

Jun, S.-A., ed. 2014b. *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Jun, S.-A., and J. Bishop. 2015. "Priming Implicit Prosody: Prosodic Boundaries and Individual Differences." *Language and Speech* 58 (4): 459–473.

Jun, S.-A., and J. Fletcher. 2014. "Methodology of Studying Intonation: From Data Collection to Data Analysis." In Jun 2014b, 493–519.

Jun, S.-A., and S. Kim, S. 2004. "Default Phrasing and Attachment Preferences in Korean." In *Proceedings of Interspeech 2004*. https://www.isca-speech.org/archive/archive_papers/interspeech_2004/i04_3009.pdf.

Jun, S.-A., and C. Koike. 2008. "Default Prosody and RC Attachment in Japanese." *Japanese-Korean Linguistics* 3:41–53.

Jun, S.-A., and M. Oh. 2000. "Acquisition of 2nd Language Intonation." *Proceedings of International Conference on Spoken Language Processing* 4:76–79.

Karlsson, A. 2014 "The Intonational Phonology of Mongolian." In Jun 2014b, 187–215.

Khan, S. D. 2008. "Intonational Phonology and Focus Prosody of Bengali." PhD diss., UCLA.

Khan, S. D. 2014. "The Intonational Phonology of Bangladeshi Standard Bengali." In Jun 2014b, 82–117.

Kjelgaard, M. M., and S. Speer. 1999. "Prosodic Facilitation and Interference in the Resolution of Temporary Syntactic Closure Ambiguity." *Journal of Memory and Language* 40:153–194.

Klatt, D. H. 1975. "Vowel Lengthening Is Syntactically Determined in a Connected Discourse." *Journal of Phonetics* 3:129–140.

Ladd, D. R. 1983. "Phonological Features of Intonational Peaks." *Language* 59:721–759.

Ladd, D. R. 1990. "The Metrical Representation of Pitch Register." In *Papers in Laboratory Phonology I*, edited by J. Kingston and M. Beckman, 35–57. Cambridge: Cambridge University Press.

Ladd, D. R. (1996) 2008a. *Intonational Phonology*. 2nd ed. Cambridge: Cambridge University Press.

Ladd, D. R. 2008b. "Review of Prosodic Typology by Jun 2005c." *Phonology* 25 (2): 372–376.

Leben, W. 1973. "Suprasegmental Phonology." PhD diss., MIT.

Leben, W. 1976. "The Tones in English Intonation." *Linguistic Analysis* 2:69–107.

Lee, E.-K., and D. Watson. 2011. "Effects of Pitch Accents in Attachment Ambiguity Resolution." *Language and Cognitive Processes* 26 (2): 262–297. https://linguistics.ucla.edu/people/jun/papers%20in%20pdf/ASAPoster2016LeeandJun.pdf.

Lee, H., and S.-A. Jun. 2016. "Types of Errors in English Prosody Made by Native Korean Speakers." Poster presented at the Acoustical Society of America, Honolulu, Hawaii, December 1. https://linguistics.ucla.edu/people/jun/papers%20in%20pdf/ASAPoster2016YooandJun.pdf.

Lehiste, I. 1973. "Phonetic Disambiguation of Syntactic Ambiguity." *Glossa* 7 (2): 107–121.

Liberman, M. 1975. "The Intonational System of English." PhD diss., MIT.

Liberman, M., and J. Pierrehumbert. 1984. "Intonational Invariance under Changes in Pitch Range and Length." In *Language Sound Structure*, edited by M. Aronoff and R. Oehrle, 157–233. Cambridge, MA: MIT Press.

Lieberman, P. 1967. *Intonation, Perception and Language*. Cambridge, MA: MIT Press.

Maekawa, K., H. Kikuchi, Y. Igarashi, and J. Venditti. 2002. "X-JToBI: An Extended J-ToBI for Spontaneous Speech." In *Proceedings of the Seventh International Conference on Spoken Language Processing*. https://www.isca-speech.org/archive/archive_papers/icslp_2002/i02_1545.pdf.

Mayo, C., M. Aylett, and D. R. Ladd. 1997. "Prosodic Transcription of Glasgow English: An Evaluation Study of GlaToBI." In *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications*. https://www.cstr.ed.ac.uk/downloads/publications/1997/esca2.pdf.

McGory, J. 1997. "Acquisition of Intonational Prominence in English by Seoul Korean and Mandarin Chinese Speakers." PhD diss., Ohio State University.

Mennen, I. 1998. "Stability of Tonal Alignment: The Case of Greek Prenuclear Accents." *Journal of Phonetics* 26 (1): 3–25.

Mennen, I. 2004. "Bi-Directional Interference in the Intonation of Dutch Speakers of Greek." *Journal of Phonetics* 32:543–563.

Mennen, I. 2007. "Phonological and Phonetic Influences in Non-Native Intonation." In *Non-Native Prosody: Phonetic Descriptions and Teaching Practice*, edited by J. Trouvain and U. Gut, 53–76. Berlin: Mouton de Gruyter.

Mennen I. 2015. "Beyond Segments: Towards a L2 intonation Learning Theory." In *Prosody and Language in Contact*, edited by E. Delais-Roussaire, M. Avanzi, and S. Herment, 171–188. Heidelberg, Germany: Springer-Verlag.

Millar, J., J. Vonwiller, J. Harrington, and P. Dermody. 1994. "The Australian National Database of Spoken Language." *Proceedings of the ICASSP* 94:97–100.

Mo, Y., J. Cole, and E. K. Lee. 2008. "Prosody Perception by Naive Listeners: Evidence from a Large Multi-Transcriber Reliability Study." Poster presented at the Eighty-Second Annual Meeting of the Linguistic Society of America, Chicago, IL. January 3–6.

Nespor, M., and I. Vogel. 1986. *Prosodic Phonology*. Dordrecht, the Netherlands: Foris.

O'Brien, M., and U. Gut. 2010 "Phonological and Phonetic Realisation of Different Types of Focus in L2 Speech." In *Achievements and Perspectives in the Acquisition of Second Language*

Speech: New Sounds 2010, edited by K. Dziubalska-Kołaczyk, M. Wrembel, and M. Kul, 205–215. Frankfurt, Germany: Peter Lang.

Ostendorf, M., P. J. Price, and S. Shattuck-Hufnagel. 1995. *The Boston University Radio News Corpus*. Technical Report ECS-95–001. Boston: Boston University.

Ostendorf, M., I. Shafran, S. Shattuck-Hufnagel, L. Carmichael, and W. Byrne. 2001. "A Prosodically Labeled Database of Spontaneous Speech." In *Proceedings of the ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*. https://www.isca-speech.org/archive_open/archive_papers/prosody_2001/prsr_022.pdf.

Palmer, H. E. 1922. *English Intonation with Systematic Exercises*. Cambridge: Heffer.

Peng, S-H., C. Marjorie, C.-Y. Tseng, T. Huang, O. J. Lee, and M. Beckman. 2005. "Towards a Pan_Mandarin System for Prosodic Transcription." In Jun 2005c, 230–270.

Pierrehumbert, J. 1980. "The Phonology and Phonetics of English Intonation." PhD diss., MIT.

Pierrehumbert, J., and M. Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pierrehumbert, J., and J. Hirschberg. 1990. "The Meaning of Intonational Contours in the Interpretation of Discourse." In *Intensions in Communication*, edited by P. R. Cohen, J. Morgan, and M. E. Pollack, 271–311. Cambridge, MA: MIT Press.

Pitt, M. A., K. Johnson, E. Hume, S. Kiesling, and W. Raymond. 2005. "The Buckeye Corpus of Conversational Speech: Labeling Conventions and a Test of Transcriber Reliability." *Speech Communication* 45:89–95.

Price, M. P. J., S. Ostendorf, S. Shattuck-Hufnagel, and C. Fong. 1991. "The Use of Prosody in Syntactic Disambiguation." *Journal of Acoustical Society of America* 60:2956–2970.

Prieto, P. 2014. "The Intonational Phonology of Catalan." In Jun 2014b, 43–80.

Prieto, P., L. Aguilar, I. Mascaró, F. J. Torres-Tamarit, and M. Vanrell. M. 2009. "L'etiquetatge prosòdic Cat_ToBI." *Estudios de Fonètica Experimental* 18:287–309.

Prieto, P., and M. Ortega-Llebaria. 2009. "Do Complex Tones Induce Syllable Lengthening in Catalan and Spanish?" In *Phonetics and Phonology: Interactions and Interrelations*, edited by M. Vigário, S. Frota, and M. J. Freitas, 51–70. Amsterdam: John Benjamins.

Prieto, P., and P. Roseano, eds. 2010. *Transcription of Intonation of the Spanish Language*. Munich: Lincom.

Schafer, A., J. Carter, C. Clifton, and L. Frazier 1996. "Focus in Relative Clause Construal." *Language and Cognitive Processes* 11:135–163.

Schafer, A., S. Speer, and P. Warren. 2005. "Prosodic Influences on the Production and Comprehension of Syntactic Ambiguity in a Game-Based Conversation Task." In *Approaches to Studying World-Situated Language Use*, edited by J. C. Trueswell and M. K. Tanenhaus, 209–225. Cambridge, MA: MIT Press.

Schafer, A., S. Speer, P. Warren, and S. White. 2000. "Intonational Disambiguation in Sentence Production and Comprehension." *Journal of Psycholinguistic Research* 29:169–182.

Schepman, A., R. Lickley, and D. R. Ladd. 2006. "Effects of Vowel Length and 'Right Context' on the Alignment of Dutch Nuclear Accents." *Journal of Phonetics* 34:1–28.

Scott, D. 1982. "Duration as a Cue to the Perception of a Phrase Boundary." *Journal of the Acoustical Society of America* 71:996–1007.

Selkirk, E. 1986. "On Derived Domains in Sentence Phonology." In *Phonology Yearbook 3*, edited by C. Ewen and J. Anderson, 371–405. Cambridge: Cambridge University Press.

Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg. 1992. "ToBI: A Standard for Labeling English Prosody." In *Proceedings of the 1992 International Conference on Spoken Language Processing*. https://www.isca-speech.org/archive/archive_papers/icslp_1992/i92_0867.pdf.

Snedeker, J., and E. Casserly. 2010. "Is It All Relative? Effects of Prosodic Boundaries on the Comprehension and Production of Attachment Ambiguities." *Language and Cognitive Processes* 25:1234–1264.

Speer, S., and A. Foltz. 2015. "Implicit Contrastive Prosody: Individual Differences in Processing and Production." In *Explicit and Implicit Prosody in Sentence Processing: Studies in Honor of Janet Dean Fodor*, edited by L. Frazier and E. Gibson, 263–285. Berlin: Springer.

Speer, S. R., M. M. Kjelgaard, and K. M. Dobroth. 1996. "The Influence of Prosodic Structure on the Resolution of Temporary Syntactic Closure Ambiguities." *Journal of Psycholinguistic Research* 25:249–271.

Stirling, L., J. Fletcher, I. Mushin, and R. Wales. 2001. "Representational Issues in Annotation: Using the Australian Map Task Corpus to Relate Prosody and Discourse Structure." *Speech Communication* 33:113–134.

Thorsen, N. 1980. "A Study of the Perception of Sentence Intonation: Evidence from Danish." *Journal of the Acoustical Society of America* 67:1014–1030.

Thorsen, N. 1983. "Two Issues in the Prosody of Standard Danish." In *Prosody: Models and Measurements*, edited by A. Cutler and D. R. Ladd, 27–38. Heidelberg, Germany: Springer.

Trofimovich, P., and W. Baker. 2006. "Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech." *Studies in Second Language Acquisition* 28 (1): 1–30.

Ueyama, M., and S.-A Jun. 1998. "Focus Realization in Japanese English and Korean English Intonation." *Japanese-Korean Linguistics* 7:629–645.

Venditti, J. 1997. "Japanese ToBI Labeling Guidelines." *Ohio State University Working Papers in Linguistics* 50:127–162.

Venditti, J. 2005. "J_ToBI Model of Japanese Intonation." In Jun 2005c, 172–200.

Wightman, C., S. Shattuck-Hufnagel, P. Price, and M. Ostendorf. 1992. "Segmental Durations in the Vicinity of Prosodic Phrase Boundaries." *Journal of Acoustical Society of America* 91:1707–1717.

Wong, W. Y. P., M. K. M. Chan, and M. E. Beckman. 2005. "An Autosegmental-Metrical Analysis and Prosodic Annotation Conventions for Cantonese." In Jun 2005c, 271–300.

Yoo, H.-J., and S.-A. Jun. 2016. "Distribution of Accentual Phrase-Medial Tones in Seoul Korean." Poster presented at the Acoustical Society of America, Honolulu, Hawaii December 2.

Commentary on Chapter 4: An Enhanced Autosegmental-Metrical Theory (AM$^+$) 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$^+$. 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$^+$-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.[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 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.[2] 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$^+$ 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.[3]

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.[4] 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.[5] 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.[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 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$^+$ Theory and the RaP Transcription System

An "enhanced" AM theory (AM$^+$) is proposed here. AM$^+$ 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$^+$ 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.[8]

Note that by including both paradigmatic and syntagmatic aspects of tone in phonology, AM$^+$ 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$^+$ theory, paradigmatic aspects of tone reduce to syntagmatic feature specifications. This is a core insight of AM$^+$ theory—namely, that paradigmatic tonal features can be formally reexpressed as syntagmatic ones. AM$^+$ 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).[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 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$^+$, 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$^+$ 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$^+$ 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$^+$ provides a unifying explanation for observed correspondences between abstract tones and their typical F0 consequences, as with F0 peaks, valleys, and plateaus.[10]

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$^+$ theory.[11] 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).[12] 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.[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. 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*.[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 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).[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 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 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.[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 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 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 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.[19]

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 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$^+$ 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$^+$ 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).[20] 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 = [–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

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$^+$ and codified in RaP's conventions which enhance phonetic transparency and explanatory power relative to ToBI*:

- *Strictly monotonic interpolation functions*: AM$^+$ 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$^+$ 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$^+$, 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*.[21] 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$^+$ 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_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 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$^+$ 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).[22] RaP has been implemented as a full annotation system, with a publicly available set of interactive training materials[23] 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$^+$ 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 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$^+$ 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 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 Pierrehumbert and Beckman (1988) 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. 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 μ, 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 $n, n+1, \ldots, N$ 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 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.

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-Alone 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.

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.

Liberman, M. 1975. *The Intonation System of English*. PhD diss., MIT.

Liberman, 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.

Author Response to the Commentary: ToBI Is Not Designed to Be Phonetically Transparent

Sun-Ah Jun

Introduction

My essay for this chapter introduced how the tones and break indices (ToBI) transcription system works, and by referring to the ToBI systems of various languages, it discussed the strengths and challenges associated with the ToBI system and some of the recent developments made in response to such challenges. I was asked by the volume editors to present the ToBI system to relative newcomers to the field of prosody so that they can decide which transcription method is the one they are looking for, given their individual goals. I was also asked not to introduce the autosegment-metrical (AM) theory in depth, because this is addressed by Arvaniti (chapter 1, this volume). Therefore, my chapter was not about AM theory in general, but about ToBI in particular, and so AM theory was only described to the extent necessary to understand the ToBI system.

In their commentary, Dilley and Breen argue that AM theory has several serious problems, leading to limitations on ToBI*'s (= all ToBI-like annotation systems) value as a scientific tool. They summarized AM theory before criticizing it and introduced what they call an enhanced AM theory (AM^+). Their chief criticism of AM theory was that it does not include syntagmatic tonal relationships in phonology, which they also cite as a source of problems with ToBI*'s phonetic transparency and consistency (see "Enduring Insights from over Forty Years of Traditional AM Theory").

To fix this perceived problem, they included in AM^+ theory both syntagmatic and paradigmatic relationships of tones in the phonology and presented the rhythm and pitch (RaP) prosodic transcription system, which is an instantiation of their AM^+ theory. In this response, I will argue that Dilley and Breen's criticisms of (standard) AM theory and ToBI* are not fully justified, and in so doing, I will also highlight how AM theory and ToBI* are quite different from AM^+ theory and the RaP system.

The AM Theory versus the AM Model of English Intonation

In their commentary, Dilley and Breen list three main problems with AM theory that they claim are due to the assumption that the phonology lacks syntagmatic restrictions on tones: (i) complex phonetic rules and mechanisms for tone scaling that do not work, (ii) inconsistencies in mapping pitch accents to F0 events, and (iii) complications when F0 curves correspond to phonetic interpolation versus tones. They claimed that these problems in AM theory then translate to problems for systems like ToBI*, particularly the use of such systems for research on prosodic typology (see "Strictly Paradigmatic Phonological Representations"). However, these problems are mainly based on the AM model of English intonation proposed in Pierrehumbert (1980), and it should

be noted that the phonological categories and phonetic rules proposed for English intonation in Pierrehumbert (1980) and later works by Pierrehumbert and colleagues to explain surface F0 variations are specific to English, and thus do not apply equally to AM models of the intonational phonology of other, typologically different languages. What has been adopted from the AM model of English intonation by researchers working on models of other languages are instead the basic principles and assumptions of AM theory, such as:

- Intonation contours are analyzed as a linear sequence of two tonal targets, H (high) and L (low), and their combinations (e.g., bitones such as H + L or L + H).
- Unlike a single tone, a bitone has a close relationship between two adjacent tonal targets aligned to the hosting syllable or mora.
- Tones can have fundamentally different functions (marking prominence relations among words via pitch accents or marking the hierarchical structure of an utterance via boundary tones) and can change pragmatic or semantic meanings or cue syntactic structures.
- F0 heights of tones as well as the alignment of the tone-text matter are used in deciding a tonal category. If any of these differences changes the meaning or the prosodic structure of the utterance, that is, if any of them is distinctive, it should be coded into a tone.
- Pitch range can change throughout an utterance, reflecting the degree of prominence, the informational status of a word, the pragmatic and discourse meaning of a phrase, and the prosodic phrasing of the utterance.
- Prosodic phrasing can be defined by pitch-range reset as well as by the degree of phrase-final lengthening and the presence of a boundary tone.

MAE_ToBI versus ToBI*

It is also important to highlight that English ToBI, or more specifically, MAE_ToBI (Mainstream American English tones and break indices), is not a direct instantiation of the AM model of English intonation proposed by Pierrehumbert (1980). As Beckman, Hirschberg, and Shattuck-Hufnagel (2005) noted, a great deal of literature preceding Pierrehumbert (1980) and Beckman and Pierrehumbert (1986) contributed to the development of the MAE_ToBI. Furthermore, when a ToBI system is developed for a language whose prosodic system is quite different from English (for example, languages like Korean that lack both lexical stress and intonational pitch accent), some of the main principles proposed in MAE_ToBI had to be modified or extended, as mentioned in my essay. Therefore, criticisms of language-specific implementations like MAE_ToBI do not straightforwardly apply to ToBI* generally, and Dilley and Breen's argumentation is rather misleading in this regard.

Difficulties with Using ToBI* for Studying Prosodic Typology

Dilley and Breen claim that AM theory's lack of syntagmatic phonological restrictions on tones has led to difficulties in using ToBI* for prosodic typology, citing Hualde and Prieto (2016). However, the difficulties that Hualde and Prieto discuss are not at all specific to ToBI*, nor are they related to a lack of syntagmatic phonological specifications. Rather, the difficulties that emerge are due to ToBI*'s status as a phonological transcription system. In fact, this was one of the primary challenges to the ToBI system

that is highlighted in my essay: there are inherent limits to the use of such a system for the purposes of typological description. Phonological contrasts are by nature language-specific, and thus the same tonal label can "mean" different things across languages—just as is seen in typological study of segmental sound patterns. Typological study therefore requires consideration of both phonological and phonetic components of the target languages' sound systems, and difficulties will arise if either is ignored. To study prosodic typology, the AM approach to intonation would provide evidence primarily regarding phonological contrasts, and thus we would need to look elsewhere for evidence bearing on the relevant aspects of phonetic realization. Such evidence could come from acoustic data, but, as noted in my essay, it could also come from phonetically transparent labels of the sort being developed for an IPrA (international prosodic alphabet). As a phonological transcription system, ToBI* is neither intended to be nor expected to be phonetically transparent, and so its lack of phonetic transparency is a poor basis for criticism. Indeed, the fact that it is possible to use some ToBI labels in a phonetically transparent way, or to mix phonological and phonetic labels, was cited as one of the weaknesses of ToBI* in my essay. If a transcription system is phonetically transparent—that is, its labels pick out nondistinctive surface prosodic properties—it cannot and should not be used as a phonological transcription system.

Syntagmatic Tonal Relationships in AM Theory

There is no question that we need to consider both paradigmatic and syntagmatic properties of tones in the study of tone and intonation. As rightly pointed out in Dilley and Breen's commentary, Pierrehumbert (1980) and Pierrehumbert and Beckman (1988) did not ignore syntagmatic relationships for underlying tonal representations; although Pierrehumbert (1980) and later works on the AM model of English intonation (including Beckman and Pierrehumbert 1986) defined the L and H tonal targets paradigmatically by referring to the pitch range of a speaker at a given point in an utterance, they did include syntagmatic relationships among tones in phonology. This is, for example, evidenced in their inventory of bitonal pitch accents (e.g., L + H*, L* + H, H + L*) and the addition of an H* + L ("floating L" tone) to capture the downstep relation between two adjacent H tones. In fact, their AM model of English was not the first model to codify the importance of syntagmatic tonal relations to English intonation, as the matter was also central to earlier debate over levels versus configurations (see discussion in Bolinger 1951). Pierrehumbert (1980) and later works made their approach distinct from the earlier research by proposing that underlying tonal representations were simpler and more abstract (by allowing only two tonal levels, instead of four as in Trager and Smith 1951) and by deriving the various nondistinctive surface realizations of underlying tones in the phonetic component of the grammar. As mentioned in my essay, the surface F0 variations, including the relative F0 height of a sequence of singleton pitch accents, are explained by the degree of prominence, a tone's lexical status, and the type of tone sequence. Importantly, these factors cover both paradigmatic (e.g., the top reference line of nuclear H*'s pitch range can be higher than that of non-nuclear H*'s) and syntagmatic relationships (e.g., H% is realized higher than the preceding H-; L% is lower than the L of L + H*) of tones. Thus, a central tenet of AM theory is to allow only distinctive tonal events in the underlying representation. Any f0 events that are nondistinctive, predictable, or unmarked (e.g., a speaker's medium/unmarked F0 level at the onset of an utterance) are outside of the phonological component of the grammar.

Dilley and Breen criticized AM theory for not including syntagmatic tonal relationships in phonology, stating in their "Summary" section: "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." However, as mentioned herein, the AM model of English intonation proposed in Pierrehumbert (1980) and revised in Beckman and Pierrehumbert (1986) explicitly treated syntagmatic tonal relationships as part of the phonology and specified how bitonal pitch accents differ from singleton pitch accents. For example, Beckman and Pierrehumbert specified that "the difference between L* + H and H* involves contrast not only in the timing of peak but also in the F0 level immediately preceding the peak. The L* + H accent has a valley on the stressed syllable which is as low as that for any L* accent whereas the H* accent has no such valley" (259). The MAE_ToBI manual (Beckman and Ayers-Elam 1997) further illustrates in section 2.2 how L + H* is acoustically different from H* or from a sequence of L* and H*. Section 2.6 of the manual also illustrates how L + H* is acoustically different from L* + H, and the alignment difference is distinctive in English, demonstrating that the timing of the F0 rise relative to the accented syllable can be also phonological. Therefore, it is natural and expected that ToBI systems of other languages have included syntagmatic tonal relationships as part of phonology.

AM$^+$ Theory and the RaP Prosodic Transcription System

The details of AM$^+$ theory and the RaP prosodic transcription system, as described in Dilley and Breen's commentary, suggest that the goal of AM$^+$ is to make a prosodic transcription phonetically transparent and to provide a theory for such transcriptions. Indeed, "phonetically transparent" is explicit both in the commentary's title and their note 12 about the AM$^+$ theory validating the INTSINT (International Transcription System for Intonation) approach (Hirst and Di Cristo 1998, and Hirst, chapter 3, this volume). INTSINT is known as a narrow phonetic transcription of F0 contours.

To achieve such a goal, AM$^+$ and RaP are designed to transcribe F0 changes (turning points and slope changes) in terms of symbols representing syntagmatic tonal relationships. However, it seems that phonetic transparency is enhanced in RaP not because the AM$^+$ prioritizes syntagmatic tonal relationships (as Dilley and Breen emphasize throughout their commentary), but because the AM$^+$ theory made its "phonological" component very rich and its categories very fine-grained. The phonological component in their theory includes three features capturing syntagmatic tonal relationships—[same], [higher], and [small][1]—as well as other categories (with a few restrictions on their combination). Because the feature values of [higher] or [small] are specified only with [–same], combinations of these three features will generate five types of syntagmatic relationships between two adjacent tones, represented by five symbols in RaP: **H, !H, E, !L, L**. (Following the convention used in the commentary, I will mark the RaP symbols in bold.) **H** is for a tone that has the feature specifications [–same, +higher], **L** for those with [–same, –higher], **E** for [+same], **!H** for [–same, +higher, +small], and **!L** for [–same, –higher, +small]. In addition to these five symbols for syntagmatic F0 targets, RaP includes the **+** symbol to represent the association of an unstarred tone (on a nonprominent syllable) to an adjacent starred tone (on a prominent syllable). **+** can be put on the right side or to the left side of any unstarred tone symbol (meaning an unstarred tone can come before or after a starred tone), theoretically creating multiple combinations (e.g., **H* +L, H* +E, H+ L*, H+ E*, L* +H, L+ H*, L* +E, E+ L*, E* +H, !H* +H, H+**

!L*). Furthermore, RaP posits three phrase-initial phonological tones, designated by :, namely, **:L H** (omitting + and *) for a rise, **:H L** for a fall, and **:E E** for a level pitch, and it does not restrict the number of phrase-final tones (e.g., **H%, HH%**). Having all of these categories, AM+/RaP would easily allow for the transcription of all possible F0 contours in English or any other language at the phonetic level. At the phonological level, however, it would predict many more tonal contrasts than any one language could have, and at the same time, it would generate many impossible or unobserved F0 contours, suggesting that AM+ has a problem of overgeneration as a phonological model.

To illustrate their system's capacity for phonetic transparency, Dilley and Breen proposed RaP tonal labels for a few cases that pose well-known challenges for pitch accent categories in ToBI. First, they proposed labeling the H* versus L + H* in MAE_ToBI as **L+ !H*** versus **L+ H***, respectively, in RaP. This would allow their system to capture the difference between a small rise to a peak for H* versus a large rise to a peak for L + H* in English. However, these two RaP labels misrepresent the tonal contrast of the two pitch-accent types in English. While it is true that F0 is generally higher for L + H* than H*, it is the presence or absence of the preceding L tone target, not the height of the H target, which is distinctive in English; L + H* requires a low F0 target immediately before the H* target, but H* does not. For H*, F0 rises gradually toward the H* target and no L target should be adjacent to it. In fact, F0 does not have to rise to a peak at all to be labeled as H*; an H* can be preceded by a high F0 (e.g., a "hat pattern"). Labeling ToBI's H* as **L+ !H*** in RaP further suggests that the symbol **!H** refers to a small, local F0 rise toward a single high target, independent of the F0 level of the preceding H tone. That is, RaP would allow **!H** to occur as the first H tone of a phrase, because **!H** does not indicate that the H target is lower than a preceding H as it does in MAE_ToBI. A **!L** symbol is used in RaP to represent an F0 target slightly lower than the preceding tone. Thus the **!H** symbol does not serve the purpose of capturing the global pitch range reduction applied to all post-!H tones as it does in MAE_ToBI. And it is not clear how this phenomenon of global pitch-range reduction (beyond the window of syntagmatic relation between two adjacent tonal targets) can be represented, even on higher grid tiers, in RaP. While RaP could mark lowered pitch range with a specific diacritic before each affected tone, this would not really capture the global effect.

The idea of using **!H** to denote a local tone level for a small rise to a peak compared to a large rise to a peak is also exemplified in their proposal for Greek prenuclear pitch accents, for which they assign **L+ !H* +H**. One might think that this label closely matches what is observed in the surface F0 contour, namely, the low F0 valley before the accented syllable whose F0 is slightly lower than that of the postaccented syllable. However, this would work only if the label represents paradigmatic tonal height, which is not the case in RaP. Furthermore, given the phonological constraints on assigning pitch accent status in the AM+ theory—a metrically prominent syllable without any local pitch change is not considered a pitch accent—the **L+ !H* +H** label in RaP would not match the described surface F0 contour because this label would imply a change in the rising F0 slope during the rise (i.e., from **L+ !H*** to **!H* +H**). However, as described in detail in the literature on Greek intonation (Arvaniti and Baltazani 2005; Arvaniti, Ladd, and Mennen 1998, 2006), the prenuclear pitch accent in Greek is a simple rising tone, with a low F0 valley anchored at the preaccented syllable and a high f0 peak anchored at the postaccented syllable. Because there is no slope change in the middle of the rising F0, this rising contour would be labeled as **L+ H** in RaP (i.e., an **L+** on the pre-accented syllable and an **H** on the postaccented syllable) without any intervening tone. Consequently, the prenuclear pitch accent in Greek would not be represented

as a pitch accent in RaP at all; instead, only the prominence of the accented syllable would be marked (as **x**, for "weak prominence level") on the rhythm tier. In sum, their proposal of L+ !H* +H does not capture the F0 contour of the prenuclear pitch accent in Greek, illustrating that the AM$^+$/RaP system encounters a problem of undergeneration. This problem would disappear if pitch accent were not defined only by pitch "movements" but by a combination of pitch and other acoustic properties such as increased intensity and duration as well as the clarity of segment articulation. This then could also correctly capture the cases where a sequence of three prominent words is realized in the same pitch height.

Finally, RaP is designed to be phonetically transparent and thus make the surface F0 contours more easily recoverable from its labels. Yet the system's emphasis on syntagmatic tonal relationships to the exclusion of all paradigmatic height contrasts may actually hinder recoverability in some cases. For example, consider AM$^+$'s use of the **E** symbol, which refers to a tone having an F0 target equal in height to that of a previous tone. Because **E** only means "equal to previous F0," the symbol **E** itself is not phonetically transparent; we need to look at the following or preceding tone label to know the value of **E**. Thus, **H* +E** means **E** = H, and **E* +L** means **E** = H or !H (i.e., any tone but L), but the interpretation of **E** in a label such as **:E E+ !L*** is not as simple. It would mean a phrase-initial flat F0 followed by a slight lowering of F0. This tells us about the shape of an F0 contour but not about the phonetic value of the F0 plateau, which can be high, mid, or even low. Similarly, because the five tonal symbols in RaP are all representing relative F0 heights in comparison to the previous and sometimes following F0, recovering the F0 values for the symbols would require a much larger utterance domain (or a greater number of symbols) than would be required if using ToBI labels. And the idea that such a wide domain is perceptually necessary is in contradiction to the findings from studies on the perception of pitch targets that Dilley and Breen cite in their commentary (i.e., Bishop and Keating 2012; Honorof and Whalen 2005). It seems unlikely that these problems can be avoided unless tonal labels reflect both paradigmatic and syntagmatic tonal relationships.

Conclusion

Presenting AM$^+$ theory as an "enhanced AM" theory implies that it is a phonological theory. However, a phonological theory of intonation should make predictions about possible or impossible tonal contours of a language. It is not clear how AM$^+$ does this. It is also not clear whether the AM$^+$ model of English intonation can accommodate any fixed inventory of tonal elements or a syntax of their combination. With such a rich set of tonal symbols generated from syntagmatic tonal features and diacritics, AM$^+$ theory would certainly "facilitate" phonetically transparent prosodic annotation. However, as a theory of linguistic pitch, it captures more tonal contrasts than is found in any language, that is, it overgenerates, and in some cases it does not capture distinctive tonal contrasts, that is, it also undergenerates. If a model is only adequate for describing F0 contours by using phonetically transparent labels but that model cannot make predictions about possible or impossible tonal contours, there is no sense in which the model is truly a phonological model. That being the case, it seems that the central difference between AM and AM$^+$ is not the absence versus presence of syntagmatic tonal relationships in phonology—or about the difference in the division of labor between the phonetics and the phonology—but instead whether what is being modeled is even phonology or not.

However, given that RaP is a hybrid transcription system of phonology and phonetics, the issue is not straightforward. Though the "rhythm" aspect of RaP was not described in great detail by Dilley and Breen in their commentary, RaP's rhythm tier encodes the prominence relations among words, as well as the prosodic structure of an utterance—both of which are phonological properties. (These two types of information are labeled on the phonology tier in the PoLaR [points, levels and ranges] annotation system [Ahn, Shattuck-Hufnagel, and Veilleux 2018], which is the most recent annotation system developed to address some of the challenges of the ToBI system.) The pitch tier in RaP, however, captures a mixture of phonological and phonetic properties. In terms of phonological properties, it allows for (i) the transcription of prominence-lending pitch changes (the equivalent to pitch accents in their system, marked with *****) and (ii) the association between unstarred tones and starred tones (marked with **+**), distinguishing a bitonal pitch accent from a singleton pitch accent. In terms of phonetic properties, it allows for (i) transcribing small changes in pitch level (marked by !) in both upward and downward directions relative to the preceding F0 target and (ii) transcribing any changes (in direction or slope) displayed in an F0 contour. These changes are transcribed without any regard to whether they are contrastive (e.g., labeling phrase-initial medium level F0; labeling H* as an F0 rise, **L+ H**). Therefore, it is important to highlight that RaP is not simply a separation of ToBI's tones tier into two tiers, rhythm and pitch. Although RaP's rhythm tier includes information very similar to the marking of prominence and metrical/prosodic structure on ToBI's tones tier, RaP's pitch tier includes tonal labels that are much more phonetic than ToBI's tonal categories. In fact, the tonal labels on RaP's pitch tier are similar to those of INTSINT, in that both systems label phrase-medial F0 changes based on syntagmatic tonal relationships. At the same time, RaP's pitch tier includes more phonological information about F0 than INTSINT does, because the pitch tier represents an association between tones based on their prominence and alignment to a syllable. However, even though RaP's pitch tier is more phonological than INTSINT, it is less phonological than IPrA because RaP's pitch labels annotate finer differences in F0 than IPrA does.

In sum, ToBI* was designed to be a phonological transcription system, using L and H tonal symbols that capture contrastive tonal events marking either prominence relationships among words or the metrical and prosodic structure of an utterance. As such, the goal of a ToBI transcription is not to recover surface F0 contours like a close copy, but to represent and analyze contours in terms of contrastive tonal categories. Moreover, ToBI is not only a tool for a phonological transcription of an utterance's prosody; it is also a powerful tool for studying prosody and intonation and their interaction with various phenomena in any subfield of linguistics. However, because many researchers in recent years have acknowledged the need to transcribe nondistinctive surface tonal events in addition to the distinctive tonal events, developing transcription systems for nondistinctive F0 events should be encouraged, and these transcription systems should be used in addition to ToBI, but not as a replacement for ToBI transcription.

Note

1. Dilley and Breen say that the phonological representations in AM[+] are based on two features, [+/– same] and [+/– higher], but later, they introduced [+small] to capture a slope change in rising or falling F0 in English and Greek intonation contours. Yet both [+/–] feature values of [small] are included in their representation of paradigmatic tone specifications, as shown in their table 4c.1.

References

Ahn, B., S. Shattuck-Hufnagel, and N. Veilleux. 2018. "Polar Annotation Conventions: A Tool for Annotating Prosodic Variation." Poster presented at Experimental and Theoretical Advances in Prosody 4, Amherst, MA, October 11–13.

Arvaniti, A., and M. Baltazani. 2005. "Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora." In *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun, 84–117. New York: Oxford University Press.

Arvaniti, A., D. Ladd, and I. Mennen. 1998. "Stability of Tonal Alignment: The Case of Greek Prenuclear Accents." *Journal of Phonetics* 26:3–25.

Arvaniti, A., D. Ladd, and I. Mennen. 2006. "Tonal Association and Tonal Alignment: Evidence from Greek Polar Questions and Contrastive Statements." *Language and Speech* 49:421–450. http://dx.doi.org/10.1177/00238309060490040101.

Beckman, M., and G. M. Ayers Elam. 1997. "Guidelines for ToBI Labeling." Unpublished manuscript, Ohio State University. https://www.ling.ohio-state.edu/research/phonetics/E_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. Oxford: Oxford University Press.

Beckman, M., and J. Pierrehumbert. 1986. "Intonational Structure of English and Japanese." *Phonology Yearbook* 3:255–309.

Bishop, J., and P. 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 (2): 1100–1112.

Bolinger, D. L. 1951. "Intonation: Levels versus Configurations." *Word* 7 (3): 199–210.

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 (4): 2193–2200.

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

Pierrehumbert, J. 1980. "The Phonology and Phonetics of English Intonation." PhD diss., MIT.

Pierrehumbert, J., and M. Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Trager, G. L., and H. L. Smith. 1951. *An Outline of English Structure*. Norman, OK: Battenberg Press.