An *enhanced* autosegmental-metrical theory (AM$^+$) facilitates phonetically transparent prosodic annotation

Laura Dilley[1], *Mara Breen*[2]

[1]Michigan State University, USA
[2]Mount Holyoke College, USA
ldilley@msu.edu, mbreen@mtholyoke.edu

Abstract

The application of autosegmental theory to a variety of tonal systems in the past 40 years has spurred a wealth of insight. The application of this theory to intonation languages spawned a family of transcription approaches known as Tones and Break Indices, or ToBI. This general theoretic framework, informed by metrical stress theory, has come to be known as autosegmental-metrical (AM) theory. Yet, a number of weaknesses and limitations have been noted, both at the theoretical and empirical levels.

We argue that core limitations in the traditional AM theoretic approach can be traced clearly to a failure to consistently and transparently encode syntagmatic relationships in phonology. Building on the core insights of traditional AM theory, and drawing on empirical evidence about cognitive representations for pitch from phonetics, music cognition, music theory, and cognitive neuroscience, we propose a new theoretic approach, termed enhanced AM theory, or AM$^+$. This proposal offers a theoretical clarification of syntagmatic elements in phonology. It is shown that attributing both syntagmatic and paradigmatic properties to tones provides a unifying account of multiple outstanding challenges in tone and intonation research that have not yet found a satisfactory explanation.

Index Terms: autosegmental-metrical theory, intonational phonology, tone, F0 turning points

1. Introduction

Some 40 years have passed since core theoretic ideas about suprasegmental characteristics of languages were proposed as groundbreaking, well-cited Ph.D. dissertations at Massachusetts Institute of Technology (MIT) by Goldsmith [1], Pierrehumbert [2], and Liberman [3]– hereafter G76, P80, and L75, respectively – which formed the core ideas in what has come to be known as autosegmental-metrical (AM) theory. These theories spurred development of transcription tools known as Tones and Break Indices (ToBI) [4, 5], which have been applied to many languages.

Building on this body of work, 40+ years phonology and phonetics have contributed a core of important knowledge about how tonal aspects of languages work. Several ideas stand out. One idea that has received support is the notion that tones are autonomous from segmental structures but temporally coordinated with them [6]. Moreover, tones are recognized to be sparse, e.g., they do not occur on every syllable and are connected via F0 interpolations [7, 8]. In addition, tones participate in either pitch accents or edge tones [2, 9]. Starred tones of pitch accents associate with (and unstarred tones flank) metrically prominent positions, while phrase tones associate with constituent edges [9].

A further robust empirical finding has been that 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 [e.g., 6, 10]. More recently, evidence has accrued that abstract tonal structure is better conceived as having indices in terms of perceptual targets involving auditory pitch, as opposed to F0 indices like peaks, valleys and "elbows" [e.g., 11].

2. Historical ideas from autosegmental-metrical (AM) phonology

The foundational theoretic ideas for the above findings which have framed research and empirical study for the past 40 years are the proposals of G76 and P80. These proposals theorized that tones have paradigmatic phonological status, meaning that they are defined relative to the speaker's pitch range. A core observation about how lexical tone languages work is that a single-syllable word can be spoken in isolation with a level tone, and perceivers can recognize the tone [12, 13]. 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 [14, 15].

A further foundational proposal was the metrical stress theory of L75 [3], which led to an understanding of the hierarchical organization of stress. The influence of metrical theory is reflected in the eponymous term autosegmental-metrical (AM) theory. Metrical stress theory was assumed to hold in P80; however, the way in which metrical hierarchies interact with tone was unexplored in that work. In the following section we outline some problems and challenges with the theoretical ideas of G76 and P80.

3. Problems with strictly paradigmatic tonal representations

It is abundantly clear that both paradigmatic aspects as well as syntagmatic aspects of representations are important for tonal systems [13, 16]. Syntagmatic properties have long been thought to be central to tonal representations across languages [17, 18]. There is considerable evidence that cognitive representations of tonal information include syntagmatic relationships in lexical tone languages [19, 20], intonation languages [21], and non-linguistic tonal systems (cf. world musical traditions) [22, 23].

Both G76 and P80 acknowledged the importance of syntagmatic relationships for tonal representations, but prioritized only the capture of paradigmatic aspects in phonology. As it turns out, the assumption of strictly paradigmatic features in phonology was highly problematic. We outline here three core problems with this assumption.

3.1. Complex phonetic rules and mechanisms for tone scaling that didn't work

To supplement this "weak" phonology, it was necessary to invent a "strong" phonetics. P80 proposed rules that map the phonological representation (abstract level tone target sequences) to the phonetic representation (the f0 contour). These rules, comprising a complex set of equations laid out in an entire chapter of P80, were the main mechanism in the "grammar" for scaling the relative F0 heights of tones, one to another. They entailed an assumption of an abstract tone reference line necessary for phonetic scaling of tones, together with a gradient parametric value (which was termed "prominence" but which was equated with F0), along with abstruse parameters *n* and *k*, which lacked a phonetic interpretation.

Pierrehumbert and Beckman [9] later proposed a version of the phonetic module that dispensed entirely with the phonetic rules, instead proposing that paradigmatic tones were scaled with respect to both a high reference line and low reference line, as a function of a parameter again termed "prominence" but which was just F0. A variety of other proposals were put forward which varied with respect to numbers of reference lines, whether reference lines were static or dynamically changed, and whether tones were assumed to be on reference lines or could vary freely with respect to the reference lines, e.g., [24].

There was, furthermore, a serious problem with the phonetic rules in P80: they did not actually restrict syntagmatic relative F0 heights of tones. As demonstrated in Dilley and Brown [21] (pp. 545-548), the rules failed to successfully restrict scaling of L and H tones so that specific claimed F0 contours would correspond to the intended tonal entities. For example, Dilley and Brown show that even for bitonal accents like L+H* and L*+H – uniformly assumed to correspond to rising contours – the rules permitted H tones to fall below adjacent L tones, allowing L+H* and L*+H to map onto falling contours. These problems are not limited to these two accents, but are instead widespread throughout the accounts for tonal sequences of a variety of types.

3.2. Inconsistencies in mapping pitch accents to F0 events

Numerous complications and inconsistencies in the pitch accent inventory can be traced to the piecemeal way in which syntagmatic restrictions were handled in P80. For example, the theoretical distinction between bitonal accents like L+H* and single-tone accents like H* was itself motivated in part as a means of capturing syntagmatic relations (P80, p. 4). This treatment implicitly posited that relative heights of other tones in sequence (for example, L* followed by H*) were unconstrained in their relative heights by phonology – leaving a legacy of inconsistent treatment in ToBI*'s notational conventions regarding which pairs of tones in a sequence "code" for syntagmatic relative tone heights, and which do not. As already noted, the tone scaling rules did not actually restrict the syntagmatic relative heights of the two tones of bitonal pitch accents to surface with the intended F0 contours. Further, floating L tones were posited, which were proposed to never be phonetically realized. These are only a handful of the many inconsistencies in mapping pitch accents to F0 events assumed in P80.

3.3. Complications in when F0 curves correspond to phonetic interpolation versus tones

A third problem had to do with complications in when F0 curves correspond to phonetic interpolation versus tones. The exceptional treatment of the L tone in H*+L accents as a "floating low" tone in P80 in order to codify a syntagmatic relationship among high-toned events necessitated a further, somewhat bizarre, theory-internal complication regarding a proposal for non-monotonic, "sagging" interpolation contours. A lesser-known "bulging interpolation" function H tones stems from the assumption that a "late peak" can sometimes arise on a nonprominent syllable. Dilley and Heffner [8] showed evidence that this contour is categorically distinctive from a H accent with a peak on the stressed syllable. Finally, the inconsistency in whether level-pitched regions are to be characterized as tone copying (cf. G76), multiple tone association, or secondary association has not been resolved.

4. An *enhanced* autosegmental-metrical theory: AM$^+$ ("AM plus")

An "enhanced" autosegmental-metrical theory is proposed here, termed AM$^+$ ("AM plus"). AM$^+$ integrates insights from 40+ years of empirical work in intonational phonology, as well as research in speech perception, music cognition, and cognitive neuroscience. We develop a notational device adapted for linguistic systems that is derived from insights about cognitive representations of non-linguistic tonal information – from auditory streaming studies, music cognition studies, and music theory [22, 23, 25, 26].

A central part of AM$^+$ theory is its assumption that syntagmatic aspects of tone are part of cognitive representations for tonal systems cross-linguistically. AM$^+$ assumes syntagmatic features are part of phonology. AM$^+$ proposes that paradigmatic aspects of tones are also part of cross-linguistic tonal systems, and to be specified lexically in some tonal systems, and post-lexically in others. Syntagmatic aspects of tones, which specify the relationships of tones with one another in sequence, are likewise assumed to be specified in the lexicon in some cases, and to be assigned post-lexically in others. It is proposed that each language draws on a combination of paradigmatic and syntagmatic tonal specifications, where there will be different densities of specification at the lexical or post-lexical levels. In the following we outline some key facets of this new theoretic approach.

4.1. Tones are viewed as abstract *pitch targets*

AM$^+$ conceives of tones in cognitive, abstract terms. In this theory, tones are abstract pitch targets that involve language-specific sensorimotor mappings. Conceiving of tones as abstract pitch targets which instantiate experience-dependent sensorimotor mappings is well-grounded in empirical research from the past two decades in speech perception, music cognition, and cognitive neuroscience [27-30]. To relate concepts of tones in traditional AM theory to AM$^+$, note that a "H" tone which in traditional AM theory was taken to correspond to an F0 peak (cf. P80) can be fundamentally re-expressed as an abstract pitch target that is syntagmatically constrained to be higher in pitch than a tonal target to the left and to the right. Viewed in this way, a syllable which is autosegmentally associated with a H tone naturally maps in most speaking situations to an F0 peak. However, since tonal targets are intrinsically perceptual in nature, other F0 mappings are possible, such as F0 plateaux [21, 31] or variations in the F0 shape as given by e.g., tonal center of gravity [32, 33]. AM^+ thus provides a unifying explanation for observed correspondences between abstract tones and their typical F0 consequences, cf. F0 peaks, valleys and plateaux.

4.2. The feature set is syntagmatic

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]. 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]. The Rhythm and Pitch (RaP) Prosodic Transcription System [34, 35], instantiates the proposals of AM^+ theory. RaP includes the symbols **H**, **L** and **E**, which capture the syntagmatic relationship borne by a tone, T_n, with respect to a previous tone, T_{n-1}; boldface type will be used for RaP symbols to distinguish them from ToBI* notations (in this section, for MAE-ToBI in particular). RaP's **H** designates a tone that has feature specification [-same, +higher] and is phonetically higher than the previous tone. **L** designates a tone that has feature specification [-same, -higher] and is phonetically lower than the previous tone. **E** designates a tone with feature specification [+same] which is phonetically equal in pitch compared with the previous tone. In RaP, the features [+/- same] and [+/- higher] are specified for pairs of adjacent tones, T_{n-1} and T_n, on an AM^+ grid tier. An AM^+ grid tier is a hybrid concept which generalizes across notions of a metrical grid row [36] and an autosegmental tier.

The notation T_n / T_{n-1} is adopted to represent a pair of adjacent tones on an AM+ grid tier that is constrained by a given syntagmatic feature; the entity on the right of the "/" is the referent entity. For example, T_n / T_{n-1} = [-same, +higher] means that T_n is *higher than* T_{n-1}; phonetically, this corresponds to a rise. By extension, a *reciprocal relationship* exists between two tones captured through the relationality of this expression. A rise in forward-time is just a fall in reverse-time, which is captured by a sign change when the referent entity is in the past, e.g., T_{n-1} / T_n = [-same, -higher]; this is termed the *Reciprocal Property*.

4.3. Paradigmatic aspects of tonal representations reduce to syntagmatic features

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

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. 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]. 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 [19]. This account appears to fit well the case of Hausa, for which syntagmatic relative heights of H tones in HL sequences distinguishes statements from questions [37].

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 [23]. We propose that [+/-small], like [+/-high], is only specified for [-same]. A language with five level tones – Extra High, High, Mid, Low, Extra Low (EH, H, M, L, EL) – could be described as in Table 1.

Table 1: *Five level "paradigmatic" tone specifications derived from a set of syntagmatic features.*

Lexical phonological specification	**Phonetic interpretation**
T_{EH} / r = [-same, +higher, -small]	*substantially higher* than the mean pitch
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_{EH} / r = [-same, -higher, -small]	*substantially lower* than the mean pitch

4.4. The metrical grid and AM+ grid tiers

RaP and AM^+ theory elaborate productively on the relationship between hierarchical metrical structure and tonal associations. AM^+ and RaP adopt the starred tone notation "*" used previously to describe tones which autosegmentally associate with a metrically prominent syllable. Metrically prominent syllables are marked in RaP with **x** (moderate prominence) or **X** (strong prominence), where the latter would occupy a higher grid tier position than the former. Importantly, AM^+ theory proposes that starred tones which associate with prominent metrical positions propagate upward to be represented in positions of adjacency on higher grid tiers. Following the idea of traditional metrical grid formalisms, higher levels of AM_+ grid tiers entail adjacency of elements that occupy them. The significance of this is that on higher grid tiers, nonadjacent tones may be specified for syntagmatic featural relationships lexically or post-lexically. This allows an account of tone register phenomena, e.g., downstep, downdrift, and upstep [24, 38]. A metrical account is consistent with a growing body of evidence of metrical interactions in a variety of languages with very different tonal systems [39, 40].

4.5. Notational conventions in AM+ and RaP

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.

4.5.1. *Strictly monotonic interpolation functions.*

Interpolation functions are strictly monotonic, ensuring that all turning points are coded as tones. Multiple studies have demonstrated evidence against P80's proposal that certain F0 turning points are not tones but rather reflexes of exceptional non-monotonic interpolation functions [7, 8, 41].

4.5.2. *Tones and timing slots*

An aspect of AM$^+$ theory which notably increases phonetic transparency is that every tone must be associated with a timing slot. This effectively disallows floating tones and multiple associations between a single tone and more than one timing slot (cf. tone spread or secondary association). Like ToBI*, RaP allows multiple tones to be associated with a syllable.

4.5.3. *Meaningful pitch range differences*

AM+ and RaP accounts for ToBI's much-studied distinction between H* vs. L+H* [42]. First, note that RaP repurposes ToBI's "**!**" symbol to indicate a small pitch interval ([+small]). RaP then captures the contours as **L+ !H*** (for ToBI's H*) vs. **L+ H*** (for ToBI's L+H*).

4.5.4. *Phrase edges*

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. The phrase-initial tone thus redundantly corresponds to the reciprocal (via the Reciprocal Property) of the tone in second position in the phrase; this is noted in RaP with "**:**". 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**.

4.5.5. *Sparse tonal representation*

Consistent with a sparse tonal representation, adjacent syntagmatic features are required to have different featural specifications. As a result, when two rising intervals – [-same, +higher] – are adjacent to one another, one of them must be [-small] and the other [+small]. Phonologically, adjacent syntagmatic features of [+same] are thus banned for T_1 T_2 T_3. Phonetically, this corresponds to a slope change, with a tone – starred or unstarred – indicated at the locus of the slope change. As a consequence of these assumptions, there are no sequences like **E* +E**, **E* E+** or **E* E**, meaning that P80's assertion that strings of L* accents may give rise to a low, flat pitch is not supported in the present theory.

5. Conclusions

Numerous problems exist with traditional AM theory and ToBI. These serious problems included complex phonetic rules for tone scaling that didn't work [21], inconsistencies in mapping pitch accents to F0 events, and complications in when F0 curves corresponded to phonetic interpolation versus tones. It was shown that these problems can all be traced to a failure to clearly and consistently codify syntagmatic aspects of tone

An enhanced autosegmental-metrical theory was introduced here, termed *AM$^+$ theory* ("AM plus theory") as an alternative. AM+ offers a simpler, more compact theory with fewer unsupported elements which results in substantial improvements in phonetic transparency. AM$^+$ accounts for a number of outstanding tonal phenomena in intonational and tonal phonology. It integrates evidence from 25+ years of research across disciplines, in phonetics, music cognition/theory, and cognitive neuroscience [22, 23, 25-30].

AM+ offers the novel proposal that tones are themselves endowed with both syntagmatic and paradigmatic properties. Through a novel syntagmatic notational device, paradigmatic aspects of tone are shown to be formally reduced to syntagmatic features. Further, it is shown that when paradigmatic tonal aspects are appropriately expressed, it is possible to obtain syntagmatic aspects "for free".

AM$^+$ is implemented with the Rhythm and Pitch (RaP) annotation system, which offers a phonetically transparent alternative to ToBI*. Its phonetic transparency makes RaP a useful starting point for developing the International Prosodic Alphabet (IPrA) [43]. RaP has been implemented as a full annotation system, with a publically available set of interactive training materials , and a corpus of RaP-labeled speech [44]. 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 [35].

AM+ is a theory which retains the best insights of traditional AM approaches, while affording new insights, as well as considerable improvement in phonetic transparency. RaP and AM+ are informed by 40+ years of research in phonetics, phonology, music cognition, and cognitive neuroscience. We hope that researchers will embrace paradigm change by moving toward AM+ and a phonetically transparent system like RaP, in the interests of fostering future discovery in tone and intonation research.

6. Acknowledgements

We wish to thank Stefanie Shattuck-Hufnagel, Jonathan Barnes, Bob Ladd, Kenneth N. Stevens, Nanette Veilleux, Donca Steriade, Michael Kenstowicz and Victor Manfredi for their encouragement and support. Also thanks to Fred Cummins, Mari Riess Jones, Amalia Arvaniti, Ani Patel, James McQueen, Oliver Niebuhr, Elisabeth Selkirk, Janet Pierrehumbert, and Mary Beckman for good discussions and for their influences on the authors. We further thank Madeleine McAuley, Leif McAuley, Jason Cristini, and Dominic Cristini for their support.

7. References

[1] J. Goldsmith, "Autosegmental phonology," Ph.D. dissertation, MIT, Cambridge, MA, 1976.

[2] J. Pierrehumbert, "The phonology and phonetics of English intonation," Ph.D. dissertation, MIT, Cambridge, MA, 1980.

[3] M. Liberman, "The intonation system of English," Ph.D. Ph.D. dissertation, MIT, Cambridge, MA, 1975.

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

[5] M. Beckman and J. Hirschberg, "The ToBI annotation conventions. Technical report, The Ohio State University and AT&T Bell Laboratories, unpublished manuscript.," 1994.

[6] D. R. Ladd, *Intonational Phonology*, 2nd ed. Cambridge: Cambridge University Press, 2008.

[7] D. R. Ladd and A. Schepman, ""Sagging transitions" between high accent peaks in English: Experimental evidence," *Journal of Phonetics,* vol. 31, pp. 81-112, 2003.

[8] L. C. Dilley and C. Heffner, "The role of f0 alignment in distinguishing intonation categories: Evidence from American English," *Journal of Speech Sciences,* vol. 3, no. 1, pp. 3-67, 2013.

[9] J. Pierrehumbert and M. Beckman, *Japanese tone structure*. Cambridge, MA: MIT Press, 1988.

[10] M. D'Imperio, B. Gili Fivela, and O. Niebuhr, "Alignment perception of high intonational plateaux in Italian and German," in *Proceedings of the International Conference on Speech Prosody,* Chicago, US, 2010.

[11] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel, "Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology," *Laboratory Phonology,* vol. 3, no. 2, pp. 337-383, 2012.

[12] C.-Y. Lee, "Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study," *Journal of the Acoustical Society of America,* vol. 125, pp. 1125-1137, 2009.

[13] G. Peng, C. Zhang, H.-Y. Zheng, J. Minnett, and W. S.-Y. Wang, "The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems," *Journal of Speech, Language, and Hearing Research,* vol. 55, pp. 579-595, 2012.

[14] J. Bishop and P. A. Keating, "Perception of pitch location within a speaker's range: Fundamental frequency, voice quality, and speaker sex," *Journal of the Acoustical Society of America,* vol. 132, no. 2, pp. 1100-1112, 2012.

[15] D. N. Honorof and D. H. Whalen, "Perception of pitch location within a speaker's F0 range," *Journal of the Acoustical Society of America,* vol. 117, no. 4, pp. 2193-2200, 2005.

[16] A. Cutler, D. Dahan, and W. van Donselaar, "Prosody in the comprehension of spoken language: A literature review," *Language and Speech,* vol. 40, pp. 141-201, 1997.

[17] R. Jakobson, C. G. M. Fant, and M. Halle, "Preliminaries to speech analysis." Cambridge: MIT, 1952, p.^pp. Pages.

[18] J. Cole, "Prosody in context: A review," *Language, Cognition and Neuroscience,* vol. 30, no. 1-2, pp. 1-31, 2015.

[19] D. Odden, "Tone: African languages," in *The Handbook of Phonological Theory,* J. Goldsmith, Ed.: Blackwell, 1995, pp. 444-475.

[20] P. C. M. Wong and R. L. Diehl, "Perceptual normalization for inter- and intra-talker variation in Cantonese level tones," *Journal of Speech, Language, & Hearing Research,* vol. 46, pp. 413-421, 2003.

[21] L. C. Dilley and M. Brown, "Effects of pitch range variation on F0 extrema in an imitation task," *Journal of Phonetics,* vol. 35, pp. 523-551, 2007.

[22] E. M. Burns, "Intervals, scales, and tuning," in *The Psychology of Music*, D. Deutsch, Ed. 2nd ed. San Diego: Academic Press, 1999, pp. 215-264.

[23] A. D. Patel, *Music, Language, and the Brain*. Oxford University Press, 2010.

[24] H. Truckenbrodt, "Upstep and embedded register levels," *Phonology,* vol. 19, pp. 77-120, 2002.

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

[26] E. E. Hannon and L. J. Trainor, "Music acquisition: effects of enculturation and formal training on development," *Trends in Cognitive Sciences,* vol. 11, no. 11, pp. 466-472, 2007.

[27] F. H. Guenther and G. Hickok, "Role of the auditory system in speech production," in *Handbook of Clinical Neurology. Vol. 129: The Human Auditory System: Fundamental Organization and Clinical Disorders*, M. J. Aminoff, F. Boller, and D. Swaab, Eds.: Elsevier, 2015, pp. 161-175.

[28] S. Chen, H. Liu, Y. Xu, and C. R. Larson, "Voice F0 responses to pitch-shifted voice feedback during English speech," *Journal of the Acoustical Society of America,* vol. 121, no. 2, pp. 1157-1163, 2007.

[29] T. A. Burnett, M. B. Freedland, C. R. Larson, and T. C. Hain, "Voice F0 responses to manipulations in pitch feeback," *Journal of the Acoustical Society of America,* vol. 103, no. 6, pp. 3153-3161, 1998.

[30] L. H. Ning, C. Shih, and T. M. Loucks, "Mandarin tone learning in L2 adults: A test of perceptual and sensorimotor contributions," *Speech Communication,* vol. 63, pp. 55-69, 2014.

[31] R.-A. Knight, "The shape of nuclear falls and their effect on the perception of pitch and prominence: peaks vs. plateaux," *Language and Speech,* vol. 51, no. 3, pp. 223-244, 2008.

[32] O. Niebuhr, *Perzeption un kognitive Verarbeitung der Sprechmelodie: Theoretische Grundlagen und empirische Untersuchungen* (Language, Context, and Cognition). Berlin: Walter de Gruyter, 2007.

[33] M. D'Imperio, "The role of perception in defining tonal targets and their alignment," Ph.D. Ph.D. dissertation, The Ohio State University, 2000.

[34] L. C. Dilley and M. Brown, "The RaP (Rhythm and Pitch) Labeling System, Version 1.0," 2005.

[35] M. Breen, L. C. Dilley, J. Kraemer, and E. Gibson, "Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch)," *Corpus Linguistics and Linguistic Theory,* vol. 8, no. 2, pp. 277-312, 2012.

[36] M. Halle and W. Idsardi, "General properties of stress and metrical structure," in *The Handbook of Phonological Theory,* J. A. Goldsmith, Ed., 1995, pp. 403-441.

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

[38] S. Inkelas, W. R. Leben, and M. Cobler, "The phonology of intonation in Hausa," in *Proceedings of the 16th Annual Meeting of NELS*, Amherst, 1986: GLSA, University of Massachusetts.

[39] V. Manfredi, "Spreading and downstep: prosodic government in tone languages," in *The Phonology of Tone: The Representation of Tonal Register*, H. van der Hulst and K. Snider, Eds. Berlin, New York: Mouton de Gruyter, 1993, pp. 133-184.

[40] B. Hayes, *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press, 1995.

[41] L. C. Dilley, D. R. Ladd, and A. Schepman, "Alignment of L and H in bitonal pitch accents: Testing two hypotheses," *Journal of Phonetics,* vol. 33, no. 1, pp. 115-119, 2005.

[42] M. Breen, E. Fedorenko, M. Wagner, and E. Gibson, "Acoustic correlates of information structure," *Language and Cognitive Processes,* vol. 25, no. 7-9, pp. 1044-1098, 2010.

[43] J. H. Hualde and P. Prieto, "Towards an International Prosodic Alphabet (IPrA)," *Laboratory Phonology: Journal of the Association for Laboratory Phonology,* vol. 7, no. 1, pp. 1-25, 2016.

[44] M. Breen, L. C. Dilley, M. Brown, and E. Gibson, "Rhythm and Pitch (RaP) Corpus," ed. Philadelphia: Linguistic Data Consortium, 2018.