Jonathan Barnes and Stefanie Shattuck-Hufnagel (eds.) (2022). *Prosodic theory and practice.* Cambridge, MA: MIT Press. Pp. ix + 453.

Kristine M. Yu 🔘

Department of Linguistics, University of Massachusetts Amherst, 01003, United States
E-mail: krisyu@linguist.umass.edu

1. Overview

In Barnes & Shattuck-Hufnagel's own words (pp. vii–viii), their edited volume *Prosodic theory and practice* arose in response to the following desire:

> What would be ideal, we felt, would be if someone were to compile, in a single published resource, compact and accessible presentations of each major approach currently influential in the realm of prosodic theory and practice (e.g., in the formation of transcription systems, corpus development, etc.). Each chapter would lay out, in its developers' own words, that theory's central goals and assumptions, its strengths (what it does well), and also its weaknesses (what it is not able, or indeed not designed, to do). These chapters then could serve both as works of reference for established scholars and, perhaps more importantly, as tutorials for students just entering the field.

Each contributing author was given a common list of questions to address (enumerated on pages 1–2) about phonological representations, mapping prosodic forms to meaning, phonetic realisation, prosodic typology, psychological reality and prosodic transcription. The resulting book is a thought-provoking collection of eleven chapters by influential thinkers on spoken language prosody. A special feature of the book is its inclusion of critical commentaries responding to some of the chapters, as well as responses to those commentaries by the original chapter authors. The presence of these commentaries and responses – as well as the editors' exhortation to the authors to be explicit about a common list of key (and contentious) issues – has resulted in a constructive work that opens up and clarifies the conversation between researchers in spoken language prosody.

A number of the contributions engage heavily with the autosegmental-metrical (AM) theory of intonation, which is perhaps the most dominant phonological theory underlying much current intonation research. It emerged from Pierrehumbert's (1980) doctoral dissertation on (Mainstream American) English intonation. Chapter 1

comprises Arvaniti's introduction of AM theory and developments, Grice's commentary questioning foundational assumptions of AM theory and Arvaniti's response.[1] Chapter 4 contains Jun's introduction to the ToBI intonational transcription conventions (which emerged from the backdrop of AM theory), commentary by Breen and Dilley that introduces their alternative RaP transcription system and Jun's response. Ladd provides additional thoughts on the practice of using ToBI in Chapter 6.

Some chapters help contextualise AM theory and/or ToBI. Nolan's Chapter 9 is a retrospective on the British school of intonation analysis – one of the most influential sources of ideas about intonational phonology leading up to AM theory – and includes discussion of the British school's relation to AM theory. Chapter 11 introduces the PENTA model (Xu 2005) and includes a commentary by Pierrehumbert that reflects on AM theory; the chapter continues an ongoing conversation about the PENTA model and AM theory starting with Arvaniti & Ladd (2009). The Fujisaki model (Fujisaki & Hirose 1984) and Grønnum's Chapter 2 intonational model of Copenhagen Danish exemplify so-called 'overlay' or 'superpositional' models that have been a traditional point of contrast with the 'linear' tone sequencing of AM theory (Ladd 1995). Hirst's Chapter 3 discusses INTSINT (Hirst & Di Cristo 1998), a narrow phonetic transcription system often presented as a foil to ToBI.

Other chapters showcase different approaches to defining the primitives of prosodic models and mappings from prosodic categories to F0 contours. Krivokapić's Chapter 5 explicates prosody in Articulatory Phonology (Browman & Goldstein 1987) and includes commentary by Turk, who sketches principles of an alternative model based on 'phonology-extrinsic timing'. Niebuhr's Chapter 8 on the Kiel Intonation Model (Kohler 1991) and Chapter 11 on the PENTA model (Xu, Prom-on and Liu) share a 'function first' approach that centres meaning distinctions (including focus, as well as traditionally paralinguistic notions) as the basis for establishing prosodic categories. And like Chapter 11, Hirst's Chapter 3 sections on Momel, Mertens's Chapter 7 on the Prosogram and Chapter 10 on the PaintE model (Schweitzer, Möbius, Möhler and Dogil) exhibit different approaches to parameterising F0 contours. In the rest of the review, I provide some guidance on ways that readers with different interests might engage with the volume in §2 and make final comments in §3.

2. Choose-your-own-adventure guide

Depending on a reader's background and interests, they may find different chapters of particular interest. I make some suggestions for different possible 'choose-your-own-adventure' paths here; this also conveniently provides a way to highlight some of the themes interwoven across chapters.

2.1. *I already know something about AM theory and ToBI and want to think critically about them*

As already mentioned in §1, this volume is well-suited for you! Arvaniti's Chapter 1 presentation of AM theory provides many pointers to references about various

---

[1]The chapters are arranged in alphabetical order by author's last name.

unresolved issues and challenges for AM theory. Grice's commentary highlights the problems with assuming a hard division between pitch accents and edge tones and chaining pitch accents to prominence and edge tones to boundary demarcation. Together with Chapter 1, Jun's Chapter 4 introduction to ToBI, Dilley and Breen's commentary and Jun's response, Ladd's Chapter 6 and Pierrehumbert's Chapter 11 commentary reveal that AM theory is not monolithic and that different researchers have different understandings of what the assumptions of AM theory are. For instance, Ladd points out that exploring the consequences of allowing unspecified boundary tones (Ladd 1983; Gussenhoven 1984, 2016) in intonational grammar stopped as 'the original ToBI tonal transcription system acquired the status of received analysis' (p. 248). This is because the particular phonological analysis of Mainstream American English (MAE) in Pierrehumbert (1980) and Beckman & Pierrehumbert (1986) from which ToBI-influenced intonational analyses emerged assumed that right-edge boundary tones are obligatory. Jun's contributions in Chapter 4 and Ladd's Chapter 6 also disentangle the relation between AM theory and (MAE-)ToBI and highlight biases in prosodic analysis that have crept in due to the evolution of ToBI (Beckman *et al.* 2005) – from a set of transcription conventions designed to tag prosodic structures of interest across large speech corpora of English – to the current day in which 'ToBI has come to refer to a general framework for prosodic transcription systems based on phonological properties' (Jun, p. 151). §1 also points out other chapters that help contextualise AM theory and ToBI.

2.2. I want to get a survey of the transcription systems that are out there

If you are interested in intonational fieldwork and transcription systems that have been deployed across (varieties of) languages, then besides the Chapter 4 introduction to ToBI, you can check out IViE in Nolan's contribution (§9.4) and INTSINT in Hirst's contribution (§3.6). The International Prosodic Alphabet (IPrA) briefly mentioned in Chs. 1 and 4, which has been designed to 'make more transparent comparisons across languages possible' (Jun, p. 169) as a way to remedy the language-specificity of ToBI category labels, should also be of interest. IViE (Grabe *et al.* 1998) has been used in transcribing varieties of Englishes of the British Isles and assumes a phonological analysis of English within AM theory influenced by the British school (Gussenhoven 1984) that allows unspecified boundary tones and restricts bitonal pitch accents to be left-headed. INTSINT has been used to annotate intonational patterns in a range of languages (Hirst & Di Cristo 1998) and was designed 'to provide… something along the lines of a narrow transcription using the IPA. Like the IPA, it was intended that INTSINT could be used for preliminary descriptions of intonation patterns, even for languages that had not previously been described' (p. 136). Hirst contrasts this intent with ToBI's, which he says 'presupposes that the inventory of intonation patterns for the language being described has already been established' (p. 136). Yet, much pilot intonational fieldwork is conducted with ToBI-influenced analyses, see e.g. Jun & Fletcher (2014), as discussed in Chapter 4. If you are interested in using ToBI for fieldwork, Jun's §4.5 and Ladd's Chapter 6 caveats are helpful to take into consideration.

The contrast drawn between INTSINT and ToBI points to another choice point for transcription: at the moment, are you interested in annotating surface-level changes in F0 contours, or in annotating contrastive prosodic events? INTSINT is designed for surface-level changes, while ToBI is designed for annotating phonological contrast. Other transcription systems for explicitly annotating surface-level properties of F0 contours introduced in Barnes and Shattuck-Hufnagel's volume besides INTSINT include Grønnum's system for intonational transcription of Danish in Chapter 2, Polytonia (Mertens's §7.4) and RaP (Dilley and Breen's Chapter 4 commentary). Jun's Ch.4 commentary also briefly mentions PoLaR (Ahn *et al.* 2021). Differences in whether these transcriptions encode F0 movements as an atomic category or F0 with respect to the speaker's range highlight that annotations of even surface-level F0 properties require you to make a choice about what is important to keep track of and are thus never 'theory-neutral'.

If you have stronger hypotheses about what is important to keep track of, then transcription systems that annotate (phonologically) contrastive events should be of interest. In ToBI-based transcriptions, for instance, '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' (Jun, p. 162). ToBI, Grønnum's system for intonational transcription of Danish (§2.4.14), IViE, PROLAB (based on the Kiel Intonation Model (KIM), introduced in Niebuhr's Chapter 8) and RaP all have annotations for classifying prominence levels of some kind,[2] as well as boundary demarcation. IViE (p. 343–44), PROLAB (p. 301), and RaP (p. 192) provide labels for prominence perceived by the annotator that are divorced from F0 movement consistent with a pitch accent. Jun (p. 170) also mentions Rapid Prosodic Transcription (RPT), which crowd-sources prominence and juncture strength labelling.

PROLAB recognises the most prominence categories and has a more refined set of perceptually-motivated categories for phrase boundaries, and in general, deserves special mention if you are brainstorming about transcription system design. All of KIM's phonological distinctions (and thus, PROLAB's atomic symbols) are motivated by perceptual experiments that identify an associated intonational meaning distinction (§8.3).[3] If you're interested in annotating conversations, check out PROLAB's phrase-edge F0 contour distinctions based on attitudes towards dialogue partners and turn-taking distinctions for boundaries. PROLAB also encodes different kinds of 'emphasis', such as 'positive intensification' or 'confessing' (§8.4.7). Finally, unlike all other transcription systems mentioned in this section, PROLAB (and KIM) pays attention to holistic F0 shape patterns, rather than local F0 turning points, e.g., §8.4.4. (Grønnum's system also annotates global properties of the F0 contours: 'gradual decline over several stressed syllables, for instance, from mid to low, is marked with arrows' (p. 110).)

---

[2] Getting a grip on what 'prominence' might actually be is another issue that I (and the volume, largely) sidestep (see Gussenhoven 2015 for discussion of it).

[3] However, Niebuhr notes 'The current KIM is inconsistent insofar as it sets up phonological categories based on groups of listeners and the intonational meanings they identify. But it then defines these phonological categories by acoustic rather than perceptual features' (p. 310).

One could certainly adapt and augment ToBI (and other transcription systems) based on the language and phenomena of interest – as Jun writes, 'depending on the goal of the researchers, what to label on what tier can be different' (p. 161). But being exposed to the range of kinds of distinctions made by the banquet of transcription systems surveyed in Barnes and Shattuck-Hufnagel's volume is invaluable for opening a reader's mind to possibilities (and caveats), while providing templates for getting started.

2.3. *I want a quantitative, implemented model mapping between F0 contours and categories of some kind*

Perhaps you are designing an automated text-to-speech synthesis (TTS) system. Schweitzer et al. have a helpful note explaining the role of an intonation model in TTS sytems (§10.2.1, pp. 352–3). In addition, a number of book sections discuss models that have been implemented (although not necessarily widely) in TTS systems. These include: the discussion of the Fujisaki model in the Introduction, Arvaniti's section on interpolation in AM theory (§1.3.3), discussion of the phonetic implementation rules of AM theory in Dilley and Breen's Chapter 4 commentary and Pierrehumbert's Chapter 11 commentary (e.g. p. 419), Chapter 3 on F0 parameterisation with Momel and mapping between Momel and INTSINT, Chapter 8 on KIM (though there is no discussion of quantitative phonetic implementation) and Chapter 10 on PaintE. The PENTA model (Chapter 11) also has a way of synthesising F0 contours (§11.3.1). If you already have a speech corpus and are interested in automated F0 curve parameterisation extraction and/or prosodic annotation, then you can check out Hirst's Chapter 3 on Momel and INTSINT, and Mertens's Chapter 7 on the Prosogram and Polytonia, as well as the discussion of the Fujisaki model in the Introduction, Schweitzer *et al.*'s Chapter 10 on PaintE, and Chapter 11 on the Target Approximation (TA) component of PENTA (these last three have no automated annotation component).

More generally, Barnes and Shattuck-Hufnagel's volume invites you to consider articulatory, acoustic, and perceptual considerations in parameterising the F0 contours of (sequences of) local tonal events. Quantitative F0/pitch parameterisations explained in some detail in the book are summarised in Table 1.[4] As explicitly noted by Hirst (p. 135) and exemplified by Schweitzer et al. (Chapter 10), these parameterisation algorithms are 'theory-friendly' in the sense that they can be used to relate properties of F0 curves to your prosodic categories of choice.

2.4. *I want to be exposed to different approaches to intonational meaning*

Niebuhr (Chapter 8) and Xu et al. (Chapter 11) both include discussion of experimental design for establishing meaning contrasts. You'll want to take a look at Arvaniti's explication of Pierrehumbert & Hirschberg's (1990) approach to English intonation, which imbues AM-theoretic primitives of different H and L tone types with pragmatic meaning and composes them to 'specify a particular relationship between the propositional

---

[4]Other quantitative algorithms discussed but not covered in detail include: the Fujisaki model inspired by the physiological control of F0, Task Dynamics (Chapter 5, based on the coordination of articulatory gestures), and AM theory's target-and-interpolation acoustic rules.

Table 1. *Summary of F0/pitch parameterisation algorithms motivated and explained in some detail in the book.*

Model	Basis for parameterisation	Summary
(PEN)TA §11.2.1, §11.3	Articulatory constraints on speed of F0 movement, coarticulatory transitions between tones	Underlying F0 target trajectory parameters and rate of target approximation estimated for each syllable over sequence of syllables
Momel §3.5.4	Acoustic F0 anchor points and interpolation	Anchor F0 points of F0 curve estimated in curve fitting of quadratic splines to continuous 'macromelodic' F0 contour
PaintE §10.2.2	Acoustic peak/valley shapes, interaction of pitch perception with spectral change in time normalisation choice	F0 peak/valley shape parameters estimated in 3-syllable window centred on syllable associated to pre-specified tonal event
Prosogram §7.2	Perceptual thresholds for pitch movement, interaction of pitch perception with spectral change	Discontinuous sequence of pitch levels and movements perceived over syllabic nuclei (high sonority regions), from which various 'prosodic features' (§7.3) can be extracted

content of their utterance and the mutual beliefs of the discourse participants' for the tune (p. 47). Niebuhr's Chapter 8 on KIM describes another compositional approach where meaning primitives are more holistic, e.g. phrase-initial and pitch accent F0 contour shapes, see Figure 8.4. Moreover, KIM takes a big-tent approach to meaning: 'KIM does not differentiate between so-called linguistic and paralinguistic meanings in order to give the former priority over the latter' (p. 287). PENTA's (Chapter 11) approach to meaning is similarly big tent, but not as finely articulated, nor compositional. The discussion in Chapter 11 between Xu *et al.* and Pierrehumbert about focus and prosodic meaning should be of interest. Finally, you could take a look at Nolan's Chapter 9 on the primacy of the 'nucleus' in determining intonational meaning in the British school (including fn. 7, as well as p. 69 in Grice's commentary, for perspectives on confusion in current day definitions of nuclear accent), and take note of Grønnum's idea of ascribing different sentence types to different overall slopes over the F0 peaks in an utterance (Figure 2.10).

3. Final commentary

I greatly appreciate the stimulating collection that Barnes and Shattuck-Hufnagel have put together. The volume's contributors have done an admirable job in responding to the questions given to them (including – helpfully – sometimes simply stating that their model doesn't have much to say about some issue) and being explicit about goals and assumptions, strengths and weaknesses. As a whole, the volume thus empowers the reader to think about foundational questions of the field beyond a particular

theoretical orientation, as intended. Readers will undoubtedly find the book to be an invaluable starting point for their research, wherever they might be in the spectrum spanning 'theory' and 'practice' referred to in the book title.

'Practice' is meant to refer to applications like 'the formation of transcription systems' and 'corpus development' (p. vii). But another important sense in which the book is illuminating about 'practice' is in the practice of theory, i.e. the way people have understood certain theoretical assumptions has drifted and diverged over time (§1 and the first part of §2). The reminders in Chapters 1 and 4 that ToBI was originally designed only to supplement recordings and pitch tracks by tagging structures of interest – not replace them – left a strong impression on me, considering the dangers raised by Ladd in Chapter 6: 'In my view, our understanding of intonational phonology is actually still fairly primitive, and a standard transcription that purports to be based on a correct phonological analysis prematurely closes off avenues of investigation and theoretical debate' (p. 250).[5]

The plurality of the 'practice' of AM theory revealed in the volume – both in the attention to the evolution of AM theory and ToBI as well as the multiple and conflicting definitions of fundamental concepts from different contributors[6] – make the volume stimulating for someone experienced with intonational phonology, but likely confusing for a newcomer trying to learn about AM theory or ToBI from scratch. While I didn't notice any obvious typos, and while figures and symbols are crisp and clear, ToBI transcriptions are also used with very little introduction, and the volume does not seem to have supplementary audio recordings and annotations for examples used in the book. Thus, while I think the volume would be perfect for assigning chapters to more advanced students to have a classroom discussion about different theoretical perspectives or transcription systems, it would not be a good place to begin without supplementation and guidance. However, I do think that a newcomer could pick and choose individual chapters to get a first introduction to some transcription system of interest, without the expectation that they necessarily would be able to straightforwardly apply or adapt it, depending on the chapter. Happily, the book is open access and individual chapters can be downloaded at https://doi.org/10.7551/mitpress/10413.001.0001.

As a snapshot of the current state of the field, the book also reveals relatively unexplored topics and connections to be made. While the theories and transcription systems discussed have been developed on the basis of a small number of languages (especially West Germanic), Grice's commentary in Chapter 1 and Grønnum's Chapter 2 on Danish exemplify how exposure to the diverse prosodic systems of the world's languages is essential for developing our understanding. Additionally, while the book title refers to 'prosodic' theory and practice, most of the volume focuses narrowly on post-lexical intonational melodies and pays little attention to topics such as lexical tone (beyond Mandarin in §5.3 and Chapter 11), grammatical/morphosyntactic tone, syntax–prosody mapping, more recent advances in phonological theory on prosodic trees and prosodic weight, or prosodically conditioned segmental allophony. The phenomena examined concern almost exclusively F0 and pitch contours, with little

---

[5]I have experienced an instance of this 'closing off' in intonational fieldwork myself; see Yu (to appear: §12.4.2).

[6]For instance, Arvaniti (p. 28) seems to assume that pitch accents are necessarily associated to stressed syllables (or feet), while Jun (p. 158) does not.

attention to voice quality beyond F0 or segmental phenomena beyond the role of spectral contrast in pitch perception. Major exceptions are Niebuhr's contribution (§8.6) and Krivokapić's Chapter 5 discussion of how Articulatory Phonology provides ways to jointly investigate tones and segments. Other areas ripe for future work include the phonology of tonal association (e.g. pp. 76–77), the psycholinguistics of prosodic planning and processing (touched on the most in Chapter 5), and computational modelling of learning (§11.4).

References

Ahn, Byron, Nanette Veilleux, Stefanie Shattuck-Hufnagel & Alejna Brugos (2021). PoLaR annotation guidelines (version 1.0). Posted on OSF at https://osf.io/usbx5/.

Arvaniti, Amalia & D. Robert Ladd (2009). Greek wh-questions and the phonology of intonation. *Phonology* **26**. 43–74.

Beckman, M. & J. Pierrehumbert (1986). Intonational structure in Japanese and English. *Phonology Yearbook* **3**. 255–309.

Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel (2005). The original ToBI system and the evolution of the ToBI framework. In Sun-Ah Jun (ed.) *Prosodic typology: the phonology of intonation and phrasing*. Oxford: Oxford University Press, 9–54.

Browman, Catherine P. & Louis Goldstein (1987). Tiers in articulatory phonology, with some implications for casual speech. *Haskins Laboratories Status Report on Speech Research* **SR-92**. 1–30.

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

Grabe, Esther, Francis Nolan & Kimberley J. Farrar (1998). IViE: a comparative transcription system for intonational variation in English. In *The 5th International Conference on Spoken Language Processing*. ISCA, paper 0099.

Gussenhoven, Carlos (1984). *On the grammar and semantics of sentence accents*. Dordrecht, The Netherlands: Foris.

Gussenhoven, Carlos (2015). Does phonological prominence exist? *Lingue e linguaggio* **14**. 7–24.

Gussenhoven, Carlos (2016). Analysis of intonation: the case of MAE_ToBI. *Laboratory Phonology* **7**. 10.

Hirst, Daniel & Albert Di Cristo (eds.) (1998). *Intonation systems: a survey of twenty languages*. Cambridge: Cambridge University Press.

Jun, Sun-Ah & Janet Fletcher (2014). Methodology of studying intonation: from data collection to data analysis. In Sun-Ah Jun (ed.) *Prosodic typology II: the phonology and phonetics of intonation and phrasing*, chapter 16. Oxford: Oxford University Press, 493–519.

Kohler, K. J. (1991). Prosody in speech synthesis: the interplay between basic research and TTS application. *JPh* **19**. 121–138.

Ladd, D. Robert (1983). Phonological features of intonational peaks. *Lg* **59**. 721–759.

Ladd, D. Robert (1995). "Linear" and "overlay" descriptions: an autosegmental-metrical middle way. In Kjell Elenius & Peter Branderud (eds.) *Proceedings ICPhS 95 Stockholm*, volume 2. 116–123.

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

Pierrehumbert, J.B. (1980). *The phonology and phonetics of English intonation*. PhD dissertation, Massachusetts Institute of Technology.

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

Yu, Kristine M. (To appear). Samoan intonation and challenges for autosegmental-metrical theory. In Sun-Ah Jun & Sameer ud Dowla Khan (eds.) *Prosodic typology III*. Oxford: Oxford University Press.