Author Response to the Commentary: Multiple Layers of Meanings Can Be Linked to Surface Prosody without Direct Mapping

Yi Xu, Santitham Prom-on, and Fang Liu

PENTA Is Not a Direct Mapping Model

We are delighted to see Pierrehumbert's characterization of parallel encoding and target approximation (PENTA) as a third-generation model of prosody and intonation. Indeed, much of the refinement PENTA may potentially bring to our understanding of prosody has benefited from knowledge gained from empirical research since the earlier models. One of the key insights from empirical findings is that surface prosodic forms, such as F0 peaks, valleys, elbows, whole contours, and so on, cannot be mapped to underlying units, be it tone, stress, pitch accents, or prominence. This insight is instrumental in the conceptualization of PENTA and is expressed explicitly in the presentation of the model. Figure 11r.1 is a reproduction of the schematic of PENTA, now with the addition of optional mappings (indicated by curved arrows) to various underlying levels that are more direct than those assumed in the model. Also added is a representation (the cloud on the far left) of all the meanings that could potentially, but not necessarily, be conveyed by speech. As indicated by the crosses, surface prosody (solid curve on the far right) not only cannot be mapped directly to meanings (longest curved arrow), but also cannot be directly linked to communicative functions, encoding schemes, underlying articulatory targets, or even the target parameters. In fact, at least three degrees of separation were recognized when PENTA was first proposed: articulatory implementation, target assignment, and parallel encoding (Xu 2004a, 2004b). In other words, the very premise of PENTA is that surface "phonetic outcomes" are not mapped directly to meanings. Of course, it is not enough to just point out the mismatches between meaning and phonetic outcomes. PENTA is about how meanings can be ultimately mapped to surface prosody through specific connection mechanisms so that there are no missing conceptual links. This means that each of the three degrees of separation needs to be explicitly represented in the model. Very broadly, as shown in figure 11r.1, meanings are first conventionalized into communicative functions, each having an encoding scheme that has been developed through many rounds of conversational interactions. The encoding schemes of all functions work in parallel to jointly determine a single sequence of targets. These targets are then articulatorily implemented through nonoverlapping sequential target approximation to generate continuous surface acoustic events.

This conceptualization indeed deviates from what Pierrehumbert, in her commentary, calls "modern linguistic theories" of prosody in various ways. In particular, two ideas offered by PENTA, which are mentioned in the main essay of this chapter, are worth recapitulating. The first is that the function-form relation, as formulated by de Saussure (1916), needs a major refinement. The second is that parametric

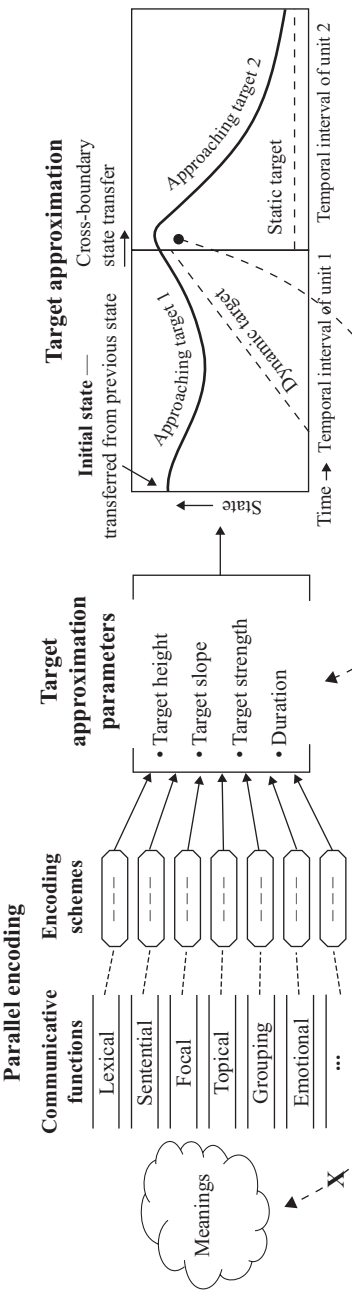


Figure 11r.1

representations should replace symbolic representations as the final link to surface phonetics. These points are elaborated in this response.

Why Function First?

De Saussure's (1916) notion that linguistic units are unities of signified and signifier does not make it clear what to do if there are uncertainties about both the signifier and the signified. This vagueness has not been a major problem for segmental phonemes because their function is relatively straightforward: to differentiate words. Thanks to people's strong intuition about words, the only major uncertainty is whether a particular segment does or does not distinguish certain words in a particular language. In prosody, both the form of the contrastive units and their functions are often ambiguous, as can be seen in the lack of consensus on both after decades of research. It is thus tempting, and has been tried many times, to first develop a descriptive account of easily observable surface prosodic features such as peaks, valleys, shapes, contours, and overall trends (Bolinger 1986; Crystal 1969; Grabe, Kochanskiand, and Coleman 2007; 't Hart, Collier, and Cohen 1990) with the hope that their meaning associations can be determined by further research. Likewise, units such as pitch accents, phrase accents, and boundary tones were originally summarized from "observed features of F0 contours" without explicit association with meanings, as is made clear in Pierrehumbert (1980, 59). Although there have been later efforts to link them to pragmatic meanings such as truth condition and common ground (Pierrehumbert and Hirschberg 1990), the proposed prosodic units remain primarily defined by their forms, as is evident from the fact that transcription of pitch tracks is used as a major means of prosody analysis (Silverman et al. 1992).

What is overlooked in these approaches is that this is *not* how segmental phonemes are determined. While it is true, as Pierrehumbert notes in her commentary, that "each language has a relatively small inventory of phonological units" ("Introduction"), whether a particular segment should be considered as a phoneme has to be determined by whether it serves to make any specific lexical contrasts rather than whether it sounds sufficiently different from other segments (Swadesh 1934). In other words, a highly specific functional contrast is the primary determinant of the phonemic status of the segment.

What may have made the segmental phonology different from prosody is what is known as *duality of patterning* (Hockett 1960), which is the essence of phonology as a bottleneck that, as Pierrehumbert notes, "helps the language learner to acquire a large vocabulary by allowing articulatory and perceptual patterns exhibited in one word to be reused in other words" ("Introduction"). Here the key word is the *reuse* of the same phoneme in different words, for example, the vowel /i/ in *bin*, *pin*, and *tin*, and the consonants /b/ and /n/ in *bin*, *ban*, and *bun*. Note, however, that the reuse is within the same function, that is, lexical contrast. An appropriate comparison in prosody would be the reuse of on-focus expansion and postfocus compression (PFC) of pitch range in foci at different sentence locations (Xu, Chen, and Wang 2012. But the reuse of the same phonetic feature would not work across functions. It would be hard to claim, for example, that because a postfocus high (H) tone has the same pitch level as a prefocus low (L) tone, the [low] feature is shared between the focus function and the lexical function. In other words, it is unlikely that there is a function-independent phonological /Low/ floating around in its own right, because the [low] is only relative to other tones within the same lexical contrast function.

As recognized by Hockett (1960), duality of patterning is due to heavy crowding in the lexical contrast function, as the number of words that need to be encoded massively exceeds the number of possible distinct segmental categories. Prosody, in contrast, confronts a different kind of crowding, that is, each prosodic dimension, for example, F0, is shared by many functions: lexical, focal, phrasal, topical, sentential, attitudinal, emotional, social-indexical, and so on. To make things worse, the identity and nature of these functions are not clear, given the lack of reference in the form of words, either spoken or written. Faced with this difficulty, PENTA-based research has followed a function-first principle that goes beyond the simple function-form relation envisaged by de Saussure. That is, the task of prosody modeling is to find out whether a particular set of meanings has been conventionalized into a communicative function, and what the encoding scheme of this function is like in terms of how the various prosodic dimensions are utilized to encode its internal categories. Following this principle, observable prosodic forms are always treated as a secondary property, that is, a means of encoding the function-internal categories. This is why PENTA-based studies never use prosodic transcription as a method of prosodic analysis.

Hypothesis Testing by Controlled Experiments

Identifying communicative functions and their encoding schemes is by no means a trivial task. The multiple degrees of separation depicted in figure 11r.1 means that not only are surface acoustic events not directly mapped to meanings, but also no two adjacent levels are linearly related to each other to allow analysis by inversion, that is, deriving the underlying form directly from surface properties. Starting from the right end of figure 11r.1, target approximation, implemented as a generative model in the form of quantitative target approximation (Prom-on, Xu, and Thipakorn 2009), cannot be mathematically inversed to derive the underlying targets. So our modeling work has always used analysis-by-synthesis to estimate the underlying targets (Prom-on, Xu, and Thipakorn 2009; Xu and Prom-on 2014). And even with this approach, the quality of the target estimation is correlated with the size of the training corpus. This means that it is simply impossible to derive authentic underlying targets from single utterances.

Moving leftward to the link between underlying targets and the encoding schemes, any single target is the end result of joint contributions by multiple encoding schemes, which makes it impossible to derive all the contributing encoding schemes from an estimated target, no matter how accurate the estimation may be. Even within an encoding scheme, a large portion of it consists of conventions that stipulate arbitrary context-sensitive assignment of the target parameters (referred to by Pierrehumbert as "language-specific constraints"). For Mandarin, for example, the low tone would assume a rising-tone–like target if it is followed by another low tone. This means that even if a contour is correctly recognized as related to a rising tone, the underlying morpheme could be either one with the low tone or with the rising tone. For English, as found in Liu et al. (2013), whether a stressed syllable is assigned a high or low-rising target depends on its position in word, focus status, and the modality (question or statement) of the sentence. This again means that it is impossible to derive individual functions even from the estimated targets.

Finally, as indicated at the far left of the figure, not all possible meanings have conventionalized functions. It is therefore impossible to know, a priori, whether a potential meaning, no matter how useful it may seem (e.g., truth condition and common ground), can be mapped to a specific encoding scheme. For example, seven different

types of focus have been suggested in Gussenhoven (2007). But so far, not even the two most obviously different types, namely, information focus and contrastive focus, have been demonstrated to be consistently distinct from each other in their prosodic realizations (Hanssen, Peters, and Gussenhoven 2008; Hwang 2012; Katz and Selkirk 2011; Kügler and Ganzel 2014; Sityaev and House 2003).

In the face of so many levels of indirect and nonunique mappings, the only viable method of discovering whether a potential meaning has developed a conventionalized function, and what the encoding scheme of that function is like, is hypothesis testing by controlled experiments. In this paradigm, both the function and the encoding schemes are treated as hypothetical, and experiments designed to systematically manipulate the functional content are performed. In the end, it is the outcome of the experiments, which often requires multiple studies, that can inform us, with various levels of certainty, of the presence of a function and the internal structure of its encoding scheme. It is with this approach, for example, that it is determined that the most salient encoding feature of prosodic focus is PFC of pitch range and intensity in many languages and that PFC is nevertheless fully absent in many other languages (Xu, Chen, and Wang 2012).

Even with controlled experiments, however, there is an issue of whether function- or form-defined units should be the target of testing. For example, when pitch accent is targeted in some controlled studies (e.g., Grabe et al. 2000; Shue et al. 2010; Turk and White 1999), the method of elicitation is the same as those used in studies of focus, that is, question-answer or negation paradigms (Cooper, Eady, and Mueller 1985; Eady and Cooper 1986; Liu et al. 2013; Patil et al. 2008; Wang and Xu 2011; Xu and Xu 2005). Due to the presumption of pitch accents as phonological units, these studies either examine phonetic properties of the focused words only or treat those of postfocus components as due to phrase accent or boundary tones that are independent of the nuclear pitch accents.

From the perspective of the function-first principle, pitch accents are merely a phonetic property, as they are identified by the presence of local F0 peaks, valleys, or movements that sound and/or look prominent, which may or may not be due to focus. For example, a prominent F0 peak may occur at the beginning of an utterance even in the absence of an initial focus (Wang and Xu 2011). Or, a prominent pitch movement may occur near the end of a sentence, which would, by definition, be treated as a nuclear pitch accent. But both production and perception studies have shown that these peaks would neither be always intended nor perceived as a sentence-final focus (Cooper, Eady, and Mueller 1985; Rump and Collier 1996; Xu and Xu 2005). Furthermore, focus may not always be marked by an F0 peak more prominent than that in a neutral-focus sentence, as found in Turkish (Ipek 2011). This is not surprising, because the presence of PFC (which is attributed to deaccenting and/or an L-phrase accent in the autosegmental-metrical [AM] theory) already enables successful perception of focus (Ipek 2011; Rump and Collier 1996; Xu, Xu, and Sun 2004). Focus, therefore, is empirically attested as a communicative function marked by multiple phonetic cues, including on-focus increase of pitch range, intensity, and duration, and postfocus reduction of pitch range and intensity (Xu 2011), with a temporal domain that expands even across a silent phrasal pause within a sentence (Wang, Xu, and Ding 2018). In contrast, pitch accent, even when seemingly obvious, is only one of such cues, which may not even be the most critical cue, because the presence of an F0 peak later in the utterance would effectively block the perception of an early focus (Rump and Collier 1996). It would therefore be difficult for PENTA to equate focus with nuclear accent in the phrase, as suggested in Pierrehumbert's commentary.

By the same token, boundary tone, as a cue to sentence modality (question versus statement), is also only one of the phonetic markers of the contrast, rather than being a phonological unit in its own right. For American English, at least, the marking of modality involves not only a sentence-final F0 rise or fall, but also a drastic raising or lowering of postfocus F0 register (treated as due to an independent phrase accent in the AM theory), and a change of target height and target slope of all stressed syllables throughout the sentence (Liu et al. 2013).

Economy of Representation and Degrees of Freedom

The kind of controlled experiments involved in typical empirical studies, however, can go only so far as identifying the functions and the gross patterns of their encoding schemes. To be able to account for the full details of surface prosody, a further step is needed to establish a form of representation that can generate real speech–like continuous prosodic events. This ultimate goal is attempted in PENTA through parametric representation. In this regard, however, PENTA is often criticized for being uneconomical in representation (see Arvaniti, chapter 1, this volume; Arvaniti and Ladd 2009, 2015), given its insistence on (i) pitch target for every syllable even if it is unstressed or bearing the neutral tone, and (ii) full specification of all targets in terms of not only target height (register), but also target slope and target strength, with no allowance for any underspecifications. But we fully agree with Pierrehumbert's remark that "the human cognitive system can learn very detailed patterns and often represents them with a great deal of redundancy" ("Conclusion"). The redundancy is not only in terms of the multiple cues for any specific communicative function, as we've discussed, but also in terms of detailed continuous trajectories that carry massive variability due to articulatory mechanisms, dialectical differences, and idiosyncrasies of individual speakers.

The solution to the redundancy problem explored in the PENTA approach, as detailed in the main essay of this chapter, is model-based parametric representation. *Model-based* means that the representation is meaningful only with respect to a specific computational model. *Parametric* means that targets are specified by numerical parameters rather than symbolic features. The representation of F0, for example, is by numerical specifications of target height, target slope, and target strength, as shown in figure 11r.1. The parameter values are obtained neither by transcription nor by direct acoustic measurement, but by training the computational model on real speech data. Depending on the nature of the training data, the learned targets can be language-, dialect-, or speaker-specific. Our computational studies so far have shown that the approach is able to generate pitch contours that are both natural sounding and functionally contrastive (Prom-on, Thipakorn, and Xu 2009; Xu and Prom-on 2014). And our pilot results based on speech corpora that are less well controlled than typical experimental data have also been encouraging.

Overall, whether a representation is sufficiently economical cannot be measured by the number of representational units assumed by a theory, but by the total specifications needed to generate detailed continuous prosodic events that resemble those of natural speech. If a unit is specified only in terms of H or L, as is the case with pitch accents, phrase accent, and boundary tones, somewhere down the line, there have to be specifications of the exact pitch height, the onset time and offset time of the unit, and how exactly the unit is connected to adjacent units. If underspecification is assumed, sooner or later there has to be a mechanism to generate surface acoustics for

the underspecified units. Without including all these specifications, it is impossible to compare degrees of freedom between different models.

Another way of assessing the economy of a model is to see how many redundant parameters are required. PENTA uses only three free parameters: height, slope, and strength of targets. None of them is redundant, because they are all independently motivated. *Target height* is motivated by its universal recognition; *target slope* is motivated by the consistency of final velocity in dynamic tones (Wong 2006; Xu 1998); and *target strength* is motivated by the sluggish realization of a mid target in the neutral tone in Mandarin (Chen and Xu 2006) and unstressed syllable in English (Xu and Xu 2005). In comparison, the equivalent of target strength in the Fujisaki and the task dynamic models (stiffness) is mostly fixed (Fujisaki 1983; Saltzman and Munhall 1998) and so is largely redundant. On the other hand, the temporal domain of target approximation is fixed to the entire syllable in PENTA (Xu and Prom-on 2015), so that there are virtually no temporal degrees of freedom. This also contrasts with the Fujisaki model (Fujisaki 1983) and articulatory phonology/task dynamic model (Browman and Goldstein 1992; Saltzman and Munhall 1989), where the onset and offset of the commands and gestural scores are free parameters, which means many more degrees of freedom in the temporal domain than PENTA. Given that the AM theory has no strict specifications of tonal alignment, it would also face the problem of degrees of freedom in the temporal domain.

Conclusion

PENTA is part of an effort to develop a new way of conceptualizing the mapping between meanings and continuous acoustic signals in speech, starting from the prosodic aspect. The multifold complexity of prosody has forced us to go back to the first principles to reconsider the phonetic-phonology interface in light of the function-form dichotomy. As a result, PENTA is one of the most indirect models of prosody, as it explicates multiple degrees of separation between meaning and continuous surface prosody. At the same time, it also insists that there be no broken links in the theoretical conceptualization of prosody and intonation and has implemented this tenet by proposing specific connection mechanisms in its computational implementation. What has also emerged from this effort is that model-based parametric representation could be the key to understanding not only the mapping of meaning to continuous phonetic output, but also how the acquisition of speech production is achieved (Xu and Prom-on 2014, 2015).

References

Arvaniti, A., and D. R. Ladd. 2009. "Greek wh-Questions and the Phonology of Intonation." *Phonology* 26 (1): 43–74.

Arvaniti, A., and D. R. Ladd. 2015. "Underspecification in Intonation Revisited: A Reply to Xu, Lee, Prom-on and Liu." *Phonology* 32:537–541.

Bolinger, D. 1986. *Intonation and Its Parts: Melody in Spoken English*. Palo Alto: Stanford University Press.

Browman, C. P., and L. Goldstein. 1992. "Articulatory Phonology: An Overview." *Phonetica* 49:155–180.

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.

Cooper, W. E., S. J. Eady, and P. R. Mueller. 1985. "Acoustical Aspects of Contrastive Stress in Question-Answer Contexts." *Journal of the Acoustical Society of America* 77:2142–2156.

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

de Saussure, F. 1916. "Nature of the Linguistics Sign." In *Cours de linguistique générale*, edited by C. Bally and A. Sechehaye, 66–70. New York: McGraw-Hill.

Eady, S. J., and W. E. Cooper. 1986. "Speech Intonation and Focus Location in Matched Statements and Questions." *Journal of the Acoustical Society of America* 80:402–416.

Fujisaki, H. 1983. "Dynamic Characteristics of Voice Fundamental Frequency in Speech and Singing." In *The Production of Speech*, edited by P. F. MacNeilage, 39–55. New York: Springer-Verlag.

Grabe, E., G. Kochanski, and J. Coleman. 2007. "Connecting Intonation Labels to Mathematical Descriptions of Fundamental Frequency." *Language and Speech* 50:281–310.

Grabe, E., B. Post, F. Nolan, and K. Farrar. 2000. "Pitch Accent Realization in Four Varieties of British English." *Journal of Phonetics* 28:161–185.

Gussenhoven, C. 2007. "Types of Focus in English." In *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, edited by C. Lee, M. Gordon and D. Büring, 83–100. New York: Springer.

Hanssen, J., J. Peters, and C. Gussenhoven. 2008. "Prosodic Effects of Focus in Dutch Declaratives." In *Proceedings of Speech Prosody*, edited by P. A. Barbosa, S. Madureira, and C. Reis, 609–612.

Hockett, C. F. 1960. "The Origin of Speech." *Scientific American* 203:88–96.

Hwang, H. K. 2012. "Asymmetries between Production, Perception and Comprehension of Focus Types in Japanese." In *Proceedings of Speech Prosody 2012*, edited by Q. Ma, H. Ding, and D. Hirst, 326–329.

Ipek, C. 2011. "Phonetic Realization of Focus with no On-Focus Pitch Range Expansion in Turkish." In *Proceedings of the Seventeenth International Congress of Phonetic Sciences*, edited by Wai-Sum Lee and Eric Zee, 140–143.

Katz, J., and E. Selkirk. 2011. "Contrastive Focus vs. Discourse-New: Evidence from Phonetic Prominence in English." *Language* 87 (4): 771–816.

Kügler, F., and S. Genzel. 2014. On the Elicitation of Focus: Prosodic Differences as a Function of Sentence Mode of the Context? *Proceedings of the 4th International Symposium on Tonal Aspects of Languages*, edited by C. Gussenhoven, Y. Chen, and D. Dediu, 71–74.

Liu, F., Y. Xu, S. Prom-on, and A. C. L. Yu. 2013. "Morpheme-Like Prosodic Functions: Evidence from Acoustic Analysis and Computational Modeling." *Journal of Speech Sciences* 3 (1): 85–140.

Patil, U., G. Kentner, A. Gollrad, F. Kügler, C. Féry, and S. Vasishth. 2008. "Focus, Word Order and Intonation in Hindi." *Journal of South Asian Linguistics* 1:55–72.

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

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

Prom-on, Y., S. Xu, and B. Thipakorn. 2009. "Modeling Tone and Intonation in Mandarin and English as a Process of Target Approximation." *Journal of the Acoustical Society of America* 125:405–424.

Rump, H. H., and R. Collier. 1996. "Focus Conditions and the Prominence of Pitch-Accented Syllables." *Language and Speech* 39:1–17.

Saltzman, E. L., and K. G. Munhall. 1989. "A Dynamical Approach to Gestural Patterning in Speech Production." *Ecological Psychology* 1:333–382.

Shue, Y.-L., S. Shattuck-Hufnagel, M. Iseli, S.-A. Jun, N. Veilleux, and A. Alwan. 2010. "On the Acoustic Correlates of High and Low Nuclear Pitch Accents in American English." *Speech Communication* 52 (2): 106–122.

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*, edited by J. J. Ohala, T. Nearey, B. Derwing, M. Hodge, and G. Wiebe, 867–870.

Sityaev, D., and J. House. 2003. "Phonetic and Phonological Correlates of Broad, Narrow and Contrastive Focus in English." In *Proceedings of the Fifteenth International Congress of Phonetic Sciences*, edited by D. Recasens, M.-J. Solé, 1819–1822.

Swadesh, M. 1934. "The Phonemic Principle." *Language* 10:117–129.

't Hart, J., R. Collier, and A. Cohen. 1990. *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

Turk, A. E., and L. White. 1999. "Structural Influences on Accentual Lengthening." *Journal of Phonetics* 27:171–206.

Wang, B., and Y. Xu. 2011. "Differential Prosodic Encoding of Topic and Focus in Sentence-Initial Position in Mandarin Chinese." *Journal of Phonetics* 39 (4): 595–611.

Wang, B., Y. Xu, and Q. Ding. 2018. "Interactive Prosodic Marking of Focus, Boundary and Newness in Mandarin." *Phonetica* 75 (1): 24–56.

Wong, Y. W. 2006. "Realization of Cantonese Rising Tones under Different Speaking Rates." In *Proceedings of Speech Prosody 2006*, edited by R. Hoffmann and H. Mixdorff, PS3-14-198.

Xu, Y. 1998. "Consistency of Tone-Syllable Alignment across Different Syllable Structures and Speaking Rates." *Phonetica* 55:179–203.

Xu, Y. 2004a. "The PENTA Model of Speech Melody: Transmitting multiple Communicative Functions in Parallel." In *Proceedings of From Sound to Sense: 50+ Years of Discoveries in Speech Communication*, edited by J. Slifka, S. Manuel and M. Matthies, C-91–96.

Xu, Y. 2004b. "Transmitting Tone and Intonation Simultaneously: The Parallel Encoding and Target Approximation (PENTA) Model." In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, edited by B. Bel and I. Marlien, 215–220.

Xu, Y. 2011. "Post-Focus Compression: Cross-Linguistic Distribution and Historical Origin." In *Proceedings of the Seventeenth International Congress of Phonetic Sciences*, edited by Wai-Sum Lee and Eric Zee, 152–155.

Xu, Y., S.-W. Chen, and B. Wang, B. 2012. "Prosodic Focus with and without Post-Focus Compression (PFC): A Typological Divide within the Same Language Family?" *Linguistic Review* 29:131–147.

Xu, Y., and S. Prom-on. 2014. "Toward Invariant Functional Representations of Variable Surface Fundamental Frequency Contours: Synthesizing Speech Melody via Model-Based Stochastic Learning." *Speech Communication* 57:181–208.

Xu, Y., and S. Prom-on. 2015. "Degrees of Freedom in Prosody Modeling." In *Speech Prosody in Speech Synthesis—Modeling, Realizing, Converting Prosody for High Quality and Flexible speech Synthesis*, edited by K. Hirose and J. Tao, 19–34. Berlin: Springer.

Xu, Y., and C. Xu. 2005. "Phonetic Realization of Focus in English Declarative Intonation." *Journal of Phonetics* 33 (2): 159–197.

Xu, Y., C. X. Xu, and X. Sun. 2004. "On the Temporal Domain of Focus." In *Proceedings of International Conference on Speech Prosody*, edited by K. Hirose, 81–84.

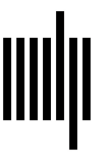
Prosodic Theory and Practice

Edited by: Jonathan Barnes, Stefanie Shattuck-Hufnagel

Citation:

Prosodic Theory and Practice
Edited by: Jonathan Barnes, Stefanie Shattuck-Hufnagel
DOI: 10.7551/mitpress/10413.001.0001
ISBN (electronic): 9780262543194
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

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

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

Library of Congress Cataloging-in-Publication Data is available.
Names: Barnes, Jonathan, 1970– editor. | Shattuck-Hufnagel, Stefanie, editor.
Title: Prosodic theory and practice / edited by Jonathan Barnes and Stefanie Shattuck-Hufnagel.
Description: Cambridge, Massachusetts : The MIT Press, 2022. | Includes bibliographical references and index.
Identifiers: LCCN 2021000764 | ISBN 9780262543170 (paperback)
Subjects: LCSH: Prosodic analysis (Linguistics)
Classification: LCC P224 .P739 2022 | DDC 414/.6—dc23
LC record available at https://lccn.loc.gov/2021000764