Commentary on Chapter 5: Time to Consider Nonoscillatory, Phonology-Extrinsic Timing Approaches to Prosody in Speech Motor Control

A. E. Turk

Introduction

Articulatory phonology/task dynamics (hereafter AP/TD) is the theory of speech production that provides the most comprehensive account of how prosodic structure influences the surface phonetic form of an utterance (Browman and Goldstein 1992; Browman and Goldstein 1995; Goldstein, Byrd, and Saltzman 2006; Saltzman and Munhall 1989; Nam, Goldstein, and Saltzman 2010; Saltzman et al. 2008). In this chapter, Krivokapić describes the mechanisms AP/TD uses to model coordination among speech articulators and a wide range of prosodic influences on coordination patterns. In this approach, prosodic gestures (*point-attractor oscillators*, i.e., damped oscillators that approximate a target and go no further) are used to model boundary-related lengthening, prominence-related lengthening, lexical tone, and pitch accent.

A second mechanism, a hierarchy of limit cycle planning oscillators (nondamped oscillators that actually oscillate), is used to model coordination and polysyllabic shortening. Because each gesture is assumed to be associated with a limit cycle planning oscillator, which freely oscillates, interoscillator coupling (entrainment) mechanisms can be used to model coordination among all types of prosodic gestures and segmental gestures and to model relationships among levels in a hierarchy of syllable, foot, and phrase oscillators. As Krivokapić shows, AP/TD is impressive in its coverage, and it is unrivaled among existing models of speech production in its accounts of temporal and prosodic phenomena.

The AP/TD view has been developed over many years and presented in many separate research articles, each dealing with a specific aspect of prosodic modeling. A major contribution of Krivokapić's essay is that it presents a synthesized view of the current AP/TD approach, drawn from a number of published sources. As a result, her review will be extremely useful to anyone working within the AP/TD approach, as well as to others working in other approaches who wish to understand it better.

In this commentary, I focus on one particular aspect of the theory: AP/TD's coupled oscillator approach to polysyllabic shortening (see next section), with some additional comments on the overall approach to speech timing. This aspect has been selected because it is a critical core assumption of the theory, which I argue is not compatible with a range of findings in the motor control literature. The additional comments address two further core assumptions: that gestures are coordinated with respect to their onsets, and that speech timing is controlled by phonology-intrinsic timing mechanisms. The first of these further assumptions is addressed in the "Challenges to AP/TD's Onset-Based Coordination" section, which presents evidence that motivates an alternative to onset-based coordination: coordination based on a goal-related part of movement (often the movement endpoint). The second is addressed in the "Challenges to AP/TD's Phonology-Intrinsic Timing" section, which presents evidence that motivates an alternative to phonology-intrinsic timing: phonology-extrinsic, general-purpose timing mechanisms. The discussion in these two sections is necessarily short; see Turk and Shattuck-Hufnagel 2020) for a more detailed treatment.

Accounts of Poly-subconstituent Shortening: A Hierarchy of Syllable, Foot, and Phrase Oscillators versus a Means of Signaling Word Boundaries

As Krivokapić described, AP/TD's approach to speech timing is based on gesture onset coordination via coupling between sets of planning oscillators for syllables, feet, and phrases. In this approach, the relative timing of gesture onsets is controlled via stable entrainment relationships (in-phase and antiphase) of the hierarchy of gestural coupling oscillators. The lengthening associated with prosodic prominence and boundaries is accounted for using π and μ_T gestures, which stretch the default gestural activation intervals with which they overlap.

The main motivation for the hierarchy of coupled syllable, foot, and phrase oscillators in AP/TD has come from findings of poly-segmental and polysyllabic shortening, where more subconstituents in a larger constituent (e.g., segments or syllables in a cross-word foot) results in durational compression of these segments or syllables. The duration of the larger constituent is shorter than expected based on the number of its subconstituent segments and/or syllables and their duration when produced in isolation. The constituent whose subconstituents undergo compression is often proposed to be a cross-word foot, that is, an interstress interval[1] that can, but does not have to, include multiple words or word fragments. On this view, for example, [*bake avo-*] in *bake avocadoes* and [*bake*] in *bake apples* would both be considered cross-word feet.

English and Swedish have been claimed to have polysyllabic shortening within cross-word feet. Campbell (1988, cited in Williams and Hiller 1994), Eriksson (1991), Williams and Hiller (1994), Kim and Cole (2005), and Kim (2006) have shown shorter stressed syllable durations when additional syllables follow the stressed syllable within a cross-word foot, even though the effect size is small (e.g., 10 to 15 percent per additional syllable, in Williams and Hiller's 1994 study). Moreover, the total duration of the cross-word foot increases linearly with additional syllables and segments (see Dauer 1983 for a summary of such findings from the literature). Although compression does not yield cross-word foot isochrony, these findings suggest a *tendency* toward isochrony that might be due to periodic control at this level. As Classe (1939, 87) puts it, the surface shortening patterns might arise from a "rhythmic tendency," which "has to contend with other factors which obscure its effects"; for example, as proposed in O'Dell and Nieminen (1999), Barbosa (2007), and AP/TD (Saltzman et al. 2008), from the interaction of coupled suprasegmental planning oscillators at multiple levels of prosodic constituency (syllable, cross-word foot, and phrase), which yields tendencies toward isochrony on the surface.

However, the interpretation of these findings as unambiguous evidence for syllable, cross-word foot, and phrase oscillators is not conclusive for two reasons. First, in most cases, the researchers did not control for the locations of word and/or phrase boundaries in their experimental materials. Thus, it is difficult to be sure whether apparent shortening in cross-word feet might be a spurious effect of word- or phrase-final lengthening. Word- and especially phrase-final syllables are known to be longer than non–word-final syllables. Thus, in corpus studies, it is possible that most monosyllabic cross-word feet correspond to single words that might exhibit word- or phrase-final lengthening; final lengthening might therefore account for the difference in duration between monosyllabic (e.g., *bake*) versus polysyllabic (e.g., *bake avo-*) cross-word feet, without requiring a separate oscillator-based polysyllabic shortening mechanism (Windmann, Šimko, and Wagner 2015).

Second, the interpretation of findings as evidence for syllable, foot, and phrase oscillators is made difficult because alternative possibilities for the constituents that might be responsible for the compression effects were often not considered. For example, whether polysyllabic shortening might operate in intervals other than cross-word feet, such as words or word-based constituents. With respect to this second question, Kim (2006) tested whether shortening occurs within words, rather than cross-word feet, for one of two speakers in the study, but did not test the possibility that it occurs within word-based content + function word units, as opposed to within interstress intervals. If the polysyllabic shortening effects described are spurious, then the main motivation for using suprasegmental oscillators to model such effects would be removed. Moreover, if polysyllabic shortening operates within intervals other than those proposed in AP/TD (that is, in units other than the syllable, cross-word foot, and phrase), then the theory would need to be modified to accommodate these different units.

Windmann, Šimko, and Wagner (2015) tested for these possibilities in a corpus study of 5.5 hours of automatically segmented broadcast speech produced by fifty-three speakers of British English. Like others in the literature, they found a clear compression effect of number of syllables on syllable duration in cross-word feet defined on the basis of word-level stress, especially for syllables that were phrasally prominent. However, this effect was due almost entirely to word-final lengthening; when the position of the measured syllable with respect to the word boundary was taken into account, that is, when separate analyses were conducted for word-final and nonfinal syllables, polysyllabic shortening effects within cross-word feet largely disappeared. This result suggests that compression mechanisms for syllables might not be needed as long as word- and phrase-final lengthening mechanisms are available.

A few effects consistent with polysyllabic shortening within cross-word feet nevertheless remained after position with respect to word boundaries was controlled for. For example, Windmann, Juraj, and Wagner (2015) found longer word-final syllables before a following word-initial stressed syllable than before a following word-initial unstressed syllable; this occurred both for unstressed syllables when preceded by a stressed syllable and for stressed syllables themselves. Effects of this type, which have been called *stress-adjacent lengthening effects*, have been found elsewhere, by, for example, Fowler (1977); also Rakerd, Sennett, and Fowler (1987); van Lancker, Kreiman, and Bolinger (1988); and Fant, Kruckenberg and Nord (1991), discussed in White (2014).

These effects are consistent with polysyllabic shortening within cross-word feet, but it is still unclear which type of constituent is responsible. If the unstressed syllable that followed the word boundary consisted of a function word (e.g., *a, us, the, him*), and the effect did not occur when the following unstressed syllable formed part of a following word, the results would be consistent with polysyllabic shortening within a word-based constituent that includes a following function word (sometimes called a clitic group) rather than within a cross-word foot that includes word fragments. It is true that, either way, these findings suggest that there are durational effects additional to boundary-related lengthening and prominence effects that need to be accounted for in some way. Critically, however, it is still unclear whether the observed effects involve cross-word feet that may contain word fragments, as proposed by AP/TD and Kim (2006), or whether they involve a word-based constituent such as the word-based clitic group.

One preliminary study that did test for the clitic group, word, and cross-word foot as possible domains of polysyllabic shortening found more support for word-based prosodic units (words and clitic groups) than for cross-word feet that may contain word fragments. Shattuck-Hufnagel and Turk (2011) contrasted these three types of candidate units in metrically regular poetic contexts (i.e., reciting limericks), where cross-word foot periodicity involving word fragments would be most likely to surface if they play a role in speech. They found that the rhyme interval durations of, for example, *bake* in *baking apples, bake us apples, bake an apple*, and *bake us an apple* were reliably shorter than *bake* in *bake apples*, for all three participants in their study. This result is consistent with polysyllabic shortening either within metrically defined inter-stress intervals (from the stressed onset syllable in *bake* to the stressed onset of *apples*), or with polysyllabic shortening within word-based clitic groups (including the content word *bake* plus the function words *us* and/or *an*).

The same study reported additional results that help to resolve the question of word-based versus metrically based constituents as the domain of polysyllabic shortening. Comparison of the rhyme interval duration of *bake* in, for example, *bake elixirs* and *bake avocadoes* versus in, for example, *baking, bake us, bake an*, and *bake us an* showed that the rhyme interval of *bake* was significantly shorter in the latter, word-based constituents. This suggests that word-based clitic groups were more influential than metrically based cross-word feet (such as *bake e-* and *bake avo-*). Only one of the three participants showed shorter rhyme interval durations for the metrically based constituents. These results suggest a stronger role for word-based constituents as compared to cross-word feet containing word fragments. They put polysyllabic shortening primarily in the domain of word-based structure, rather than of structure based on cross-word feet, suggesting that the timing control system based on syllables, feet, and phrases proposed in AP/TD would need to be modified to accommodate word-based constituents close in size to the lexical word.

A separate set of findings provides a further challenge to the use of oscillators to implement these shortening effects. White and Turk (2010) and Turk and Shattuck-Hufnagel (2000) found that polysyllabic shortening does not occur in all contexts; instead it occurs more often in phrasally stressed words, where it is also greatest in magnitude. These results challenge an oscillator-based implementation system because they suggest that even in the same utterances, not all words are affected by the suprasegmental compression mechanism. This reduces the motivation for a periodicity-based compression mechanism. Instead, these results support the idea that rather than being a reflection of a periodic control mechanism tending toward surface isochrony, polysyllabic shortening may be one of a set of mechanisms that speakers use to signal the locations of word-based prosodic constituent boundaries in an utterance (Turk 2012).

This section has described some of the challenges to AP/TD's oscillator-based approach to modeling systematic variation in speech timing that are raised by its account of poly-subconstituent shortening. Additional challenges to other aspects of oscillator-based timing control are described in the following two sections, which address, respectively, (i) the assumption of coordination based on movement onsets and (ii) the assumption of duration lengthening using phonology-intrinsic clock-slowing gestures.

Challenges to AP/TD's Onset-Based Coordination, by Evidence of Coordination Based on Movement Endpoints

AP/TD's interplanning-oscillator-coupling approach to coordination is based on gesture onsets. In this approach, the relative timing of gesture onsets is controlled via stable entrainment relationships (in-phase and antiphase) of gestural coupling oscillators. The lengthening associated with prominence and boundaries is accounted for using π and μ_T gestures, which stretch the default gestural activation intervals with which they overlap. This approach is challenged by findings suggesting that coordination can be based on parts of movement other than the onset, notably the part of movement that is most "behaviourally meaningful,"often the endpoint (Shaffer 1982; Semjen 1992). For example, Perkell and Matthies (1992) found that lip protrusion timing in relation to voicing onset was much more consistent at the end of the protrusion movement, as compared to a point near the beginning of protrusion. (See also Leonard and Cummins 2011 for related results for speech-accompanying gestures; Gentner, Grudin, and Conway 1980 for typing; Bootsma and van Wieringen 1990 and Katsumata and Russell 2012 for hitting balls; and Billon, Semjen, and Stelmach 1996 and Spencer and Zelaznik 2003 for finger-tapping.)

These findings do not find an obvious account in AP/TD because it does not have a representation of the time of gestural target approximation, that is, of mass-spring settling time. This is because in AP/TD the time of target approximation is an emergent property of mass-spring (gestural) stiffness. Although the time of gestural target approximation/mass-spring gestural settling corresponds to a fixed proportion of a gestural planning oscillator cycle at a default speaking rate, this correspondence changes when a μ_T gesture stretches the gestural planning oscillator cycles at particular locations in an utterance, and at fast speaking rates, when gestural planning oscillator cycles are shortened. Findings of movement-endpoint–based coordination motivate alternative approaches that make it possible to refer to the (surface) timing of movement endpoints, and in which endpoint times serve as reference points for coordination (Shaffer 1982, Semjen 1992). For example, Lee's (1998, 2009) general tau theory proposes that endpoint coordination can be accomplished if movement tau, that is, the time remaining to movement endpoint attainment at the current movement rate, is kept in constant proportion either to the tau of another movement, or to the tau of an internally generated tau guide. This mechanism ensures synchronous endpoint achievement, as long as two movements (or a movement with the tau guide) are tau-coupled before the end of the movement.

In sum, evidence for movement coordination at goal-related parts of movements, e.g., endpoints, presents challenges to the onset-based coordination mechanisms of AP/TD. The second set of challenges is raised by AP/TD's use of a phonology-intrinsic clock rather than a general-purpose timekeeping mechanism.

Challenges to AP/TD's Phonology-Intrinsic Timing, by Evidence for General-Purpose Timekeeping Mechanisms

This section discusses the fact that AP/TD's account of prosodically governed duration lengthening relies on the slowing of a phonology-intrinsic clock by π/μ_T gestures and contrasts this mechanism with the use of general-purpose, phonology-extrinsic timekeeping mechanisms that meter out units of solar (i.e., surface) time. As noted herein and as described by Krivokapić, in AP/TD the lengthening associated with prominence and boundaries is accounted for using π and μ_T gestures, which stretch the default gestural activation intervals with which they overlap in time. This reliance on the adjustment of default timing is challenged by several sets of findings, including (i) findings showing that timing variability is greater for longer duration intervals, (ii) findings relating to prominence and boundaries suggestive of constraints on surface durations, and

(iii) the fact that π and μ_T gestures warp phonological time in relation to solar time, in a nonlinear manner across an utterance.

In relation to the first of these findings, many investigators have shown that timing variability for a movement grows with its duration. Results of this kind have been reported for nonspeech movements by Treisman (1963); Gibbon (1977); Schmidt et al. (1979); Rosenbaum and Patshnik (1980a, 1980b); Wing (1980); Hancock and Newell (1985); Wearden (1991); Ivry and Corcos (1993); Ivry and Hazeltine (1995); Spencer and Zelaznik (2003); Merchant et al. (2008); and Merchant, Zarco, and Prado (2008); see Malapani and Fairhurst (2002) for a review. For speech movements, see studies by Byrd and Saltzman (1998); Edwards, Beckman and Fletcher (1991); Remijsen and Gilley (2008); Chen (2006); Nakai et al. (2012); and Lefkowitz (2017).

These findings are challenging for AP/TD, because in that framework, activation intervals are stretched not by adding extra time to the activation intervals, but instead by a "clock slowing" mechanism. That is, the μ_T gestures slow the planning + suprasegmental ensemble clock during intervals overlapped by the μ_T gesture. In this approach, an interval that has been stretched by a μ_T gesture therefore has the same number of AP/TD planning + suprasegmental ensemble clock units as a corresponding interval that has not been stretched by a π or μ_T gesture. The π/μ_T approach therefore does not provide an account of the greater variability for longer surface duration intervals, because these longer surface duration intervals are not longer in AP/TD planning + suprasegmental ensemble clock time. The findings appear to instead require the representation of surface durations in solar time units, where longer duration intervals contain more surface time intervals (that is, they are longer in solar time) than shorter duration intervals. Greater variability of these longer duration intervals can be explained by a "noisy" timekeeper, in which variability correlates with interval duration (Gallistel 1999; Gallistel and Gibbon 2000; Jones and Wearden 2004; Shouval et al. 2014). It is not clear how AP/TD can account for these findings, because surface durations are only emergent in this approach, and are not represented.

With respect to the second point, findings suggestive of constraints on surface durations in prominent and boundary-adjacent positions provide another challenge to the use of π/μ_T gestures. For example, in Northern Finnish, a quantity language in which short versus long quantities are not signaled by vowel quality but by vowel duration, phonemically short vowels show smaller amounts of prominence- and boundary-related lengthening than phonemically long vowels do (Nakai et al. 2009; Nakai et al. 2012; Remijsen and Gilley 2008). These findings suggest that speakers of these languages avoid large amounts of lengthening on phonemically short vowels to maintain their surface durational distinction with phonemically long vowels. These findings are difficult to explain if (as in AP/TD) surface durations cannot be represented, because prominence- and boundary-related lengthening are emergent properties of default activation interval durations plus π/μ_T adjustments, without the possibility of representing the surface duration results as a goal.

With respect to the third point, AP/TD's slowing of the phonology-intrinsic clock in prosodically prominent and phrase-final positions, and at slow speaking rates, results in a lack of linear correspondence between AP/TD's planning + suprasegmental clock time and solar time. Because AP/TD does not make use of phonology-extrinsic timekeepers, this "time-warping" may have unintended and undesirable consequences. This may occur, for example, in cases where singers or speakers need to interact with external events, as when singing to instrumental music (where matching events in solar time is critical) or timing an oral presentation to end at a particular point in solar time.

Taken together, these findings present a challenge to AP/TD's use of phonology-intrinsic timing mechanisms and motivate the development of models in which speech timing is controlled instead by general-purpose timing mechanisms that are extrinsic to the phonology.

Conclusion

While AP/TD has successfully modeled many aspects of segmental coordination and prosody, as demonstrated in Krivokapić's essay, there are a number of lines of evidence that are difficult to reconcile with its coupled oscillator approach to coordination; (i) its default-adjustment approach to longer durations at prosodic boundaries and in prominent positions; and (ii) its system of syllable-, cross-word foot- and phrase-level oscillators. These observations highlight the importance of developing alternative models of speech production and testing the predictions of these models against predictions of AP/TD. Turk and Shattuck-Hufnagel (2020) provide additional evidence that supports phonology-extrinsic timing-based approaches. Considerable work will be required to bring such models to the state of experimentally tested implementation that has made the AP/TD approach so influential and widely discussed.

Note

1. Note that the term *cross-word foot* does not specify the type of stress foot, that is, the level of prominence that delimits the foot. For example, Abercrombie's (1991) cross-word feet were delimited by "chest pulses" and were envisioned to be delimited by phrasal prominences. In AP/TD, cross-word feet are defined on the basis of word-level stress. See also Williams and Hiller (1984), who performed tests of polysyllabic shortening within feet of different types.

References

Abercrombie, D. 1991. "Some Functions of Silent Stress." In *Fifty Years in Phonetics*, edited by D. Abercrombie. Edinburgh: Edinburgh University Press.

Barbosa, P. A. 2007. "From Syntax to Acoustic Duration: A Dynamical Model of Speech Rhythm Production." *Speech Communication* 49 (9): 725–742.

Billon, M., A. Semjen, and G. E. Stelmach. 1996. "The Timing Effects of Accent Production in Periodic Finger-Tapping Sequences." *Journal of Motor Behavior* 28 (3): 198–210.

Bootsma, R. J., and P. C. van Wieringen. 1990. "Timing an Attacking Forehand Drive in Table Tennis." *Journal of Experimental Psychology: Human Perception and Performance* 16 (1): 21–29.

Browman, C. P., and L. Goldstein. 1992. "Articulatory Phonology: An Overview." *Phonetica* 49 (3–4): 155–180.

Browman, C. P., and L. Goldstein. 1995. "Dynamics and Articulatory Phonology." In *Mind as Motion*, edited by R. F. Port and T. v. Gelder, 175–193. Cambridge, MA: MIT Press.

Byrd, D., and E. Saltzman. 1998. "Intragestural Dynamics of Multiple Prosodic Boundaries." *Journal of Phonetics* 26:173–199.

Campbell, W. N. 1988. "Foot-Level Shortening in the Spoken English Corpus." Paper presented at the Seventh Forensic Anthropology Society of Europe Symposium, Edinburgh, UK, September 1988.

Chen, Y. 2006. “Durational Adjustment under Corrective Focus in Standard Chinese.” *Journal of Phonetics* 34 (2): 176–201.

Classe, A. 1939. *The Rhythm of English Prose*. Oxford: Basil Blackwell.

Dauer, R. M. 1983. “Stress-Timing and Syllable-Timing Re-Analysed.” *Journal of Phonetics* 11:51–62.

Edwards, J., M. E. Beckman, and J. Fletcher. 1991. “The Articulatory Kinematics of Final Lengthening.” *Journal of the Acoustical Society of America* 89 (1): 369–382.

Eriksson, A. 1991. “Aspects of Swedish Speech Rhythm.” PhD diss., University of Gothenburg.

Fant, G., A. Kruckenberg, and L. Nord. 1991. “Stress Patterns and Rhythm in the Reading of Prose and Poetry with Analogies to Music Performance.” Paper presented at the Music, Language, Speech, and Brain International WennerGren Symposium, Stockholm.

Fowler, C. A. 1977. *Timing Control in Speech Production*. Bloomington: Indiana University Linguistics Club.

Gallistel, C. R. 1999. “Can a Decay Process Explain the Timing of Conditioned Responses?” *Journal of the Experimental Analysis of Behavior* 71:264–271.

Gallistel, C. R., and J. Gibbon. 2000. “Time, Rate, and Conditioning.” *Psychological Review* 107 (2): 289–344.

Gentner, D. R., J. Grudin, and E. Conway. 1980. *Finger Movements in Transcription Typing*. Technical Report 8001. San Diego: Center for Human Information Processing, University of California, San Diego.

Gibbon, J. 1977. “Scalar Expectancy Theory and Weber’s Law in Animal Timing.” *Psychological Review* 84 (3): 279–325.

Goldstein, L., D. Byrd, and E. Saltzman. 2006. “The Role of Vocal Tract Gestural Action Units in Understanding the Evolution of Phonology.” In *Action to Language via the Mirror Neuron System*, edited by M. A. Arbib, 215–249. Cambridge: Cambridge University Press.

Hancock, P. A., and K. M. Newell. 1985. “The Movement Speed-Accuracy Relationship in Space-Time.” In *Motor Behavior: Programming, Control, and Acquisition*, edited by Herbert Heuer, U. Kleinbeck, and K.-H. Schmidt, 153–185. Berlin: Springer-Verlag.

Ivry, R., and D. M. Corcos. 1993. “Slicing the Variability Pie: Component Analysis of Coordination and Motor Dysfunction.” In *Variability and Motor Control*, edited by K. Newell and D. Corcos, 415–447. Champaign, IL: Human Kinetics Publishers.

Ivry, R. B., and R. E. Hazeltine. 1995. “Perception and Production of Temporal Intervals across a Range of Durations: Evidence for a Common Timing Mechanism.” *Journal of Experimental Psychology—Human Perception and Performance* 21 (1): 3–18.

Jones, L. A., and J. H. Wearden. 2004. “Double Standards: Memory Loading in Temporal Reference Memory.” *Quarterly Journal of Experimental Psychology* 57B (1): 55–77.

Katsumata, H., and D. M. Russell. 2012. “Prospective versus Predictive Control in Timing of Hitting a Falling Ball.” *Experimental Brain Research* 216 (4): 499–514.

Kim, H. 2006. “Speech Rhythm in American English: A Corpus Study.” PhD diss., University of Illinois at Urbana-Champaign.

Kim, H., and J. Cole. 2005. “The Stress Foot as a Unit of Planned Timing: Evidence from Shortening in the Prosodic Phrase.” In *Proceedings of Interspeech/Eurospeech 2005*, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, September 4–8, 2005; ISCA Archive. http://www.isca-speech.org/archive/interspeech_2005.

Lee, D. N. 1998. "Guiding Movement by Coupling Taus." *Ecological Psychology* 10 (3–4): 221–250.

Lee, D. N. 2009. "General Tau Theory: Evolution to Date." *Perception* 38:837–858.

Lefkowitz, L. M. 2017. "Maxent Harmonic Grammars and Phonetic Duration." PhD diss., UCLA.

Leonard, T., and F. Cummins. 2011. "The Temporal Relation between Beat Gestures and Speech." *Language and Cognitive Processes* 26 (10): 1457–1471.Malapani, C., and S. Fairhurst. 2002. "Scalar Timing in Animals and Humans." *Learning and Motivation* 33:156–176.

Merchant, H., W. Zarco, R. Bartolo, and L. Prado. 2008. "The Context of Temporal Processing Is Represented in the Multidimensional Relationships between Timing Tasks." *PLoS One* 3 (9): e3169.

Merchant, H., W. Zarco, and L. Prado. 2008. "Do We Have a Common Mechanism for Measuring Time in the Hundreds of Millisecond Range? Evidence from Multiple-Interval Timing Tasks." *Journal of Neurophysiology* 99:939–949.

Nakai, S., S. Kunnari, A. Turk, K. Suomi, K., and R. Ylitalo. 2009. "Utterance-Final Lengthening and Quantity in Northern Finnish." *Journal of Phonetics* 37:29–45.

Nakai, S., A. E. Turk, K. Suomi, S. C. Granlund, R. Ylitalo, and S. Kunnari,. 2012. "Quantity Constraints on the Temporal Implementation of Phrasal Prosody in Northern Finnish." *Journal of Phonetics* 40:796–807.

Nam, H., L. Goldstein, and E. Saltzman. 2010. "Self-Organization of Syllable Structure: A Coupled Oscillator Model." In Approaches to Phonological Complexity, edited by François Pellegrino, E. Marisco, and I. Chitoran, 99–328. Berlin: Mouton de Gruyter.

O'Dell, M. L., and T. Nieminen. 1999. "Coupled Oscillator Model of Speech Rhythm." In *Proceedings of the Fourteenth International Congress of Phonetic Sciences*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey, vol. 2, 1075–1078. College Park, MD: American Institute of Physics.

Perkell, J. S., and M. L. Matthies. 1992. "Temporal Measures of Anticipatory Labial Coarticulation for the Vowel /u/: Within-Subject and Cross-Subject Variability." *Journal of the Acoustical Society of America* 91 (5): 2911–2925.

Rakerd, B., W. Sennett, and C. A. Fowler. 1987. "Domain-Final Lengthening and Foot-Level Shortening in Spoken English." *Phonetica* 44:47–155.

Remijsen, B., and L. Gilley. 2008. "Why Are Three-Level Vowel Length Systems Rare? Insights from Dinka (Luanyjang Dialect)." *Journal of Phonetics* 36:318–344.

Rosenbaum, D. A., and O. Patashnik. 1980a. "A Mental Clock Setting Process Revealed by Reaction Times." In *Tutorials in Motor Behavior*, edited by G. E. Stelmach and J. Requin, 487–499. Amsterdam: North-Holland.

Rosenbaum, D. A., and O. Patashnik. 1980b. "Time to Time in the Human Motor System." In *Attention and Performance*, vol. 8, edited by R. S. Nickerson. Hillsdale, NJ: Erlbaum.

Saltzman, E. L., and K. G. Munhall. 1989. "A Dynamical Approach to Gestural Patterning in Speech Production." *Ecological Psychology* 1 (4): 333–382.

Saltzman, E., H. Nam, J. Krivokapić, and L. Goldstein. 2008. "A Task-Dynamic Toolkit for Modeling the Effects of Prosodic Structure on Articulation." In *Proceedings of the Speech Prosody 2008 Conference*, edited by P. A. Barbosa, S. Madureira, and C. Reis, 175–184.

Schmidt, R. A., H. Zelaznik, B. Hawkins, J. S. Frank, and J. T. Quinn. 1979. "Motor-Output Variability: A Theory for the Accuracy of Rapid Motor Acts." *Psychological Review* 86 (5): 415–451.

Semjen, A. 1992. "Determinants of Timing in Serial Movements." In *Time, Action and Cognition*, vol. 66, edited by F. Macar, V. Pouthas, and W. J. Friedman, 247–261. Berlin: Springer.

Shaffer, L. H. 1982. "Rhythm and Timing in Skill." *Psychological Review* 89 (2): 109–122.

Shattuck-Hufnagel, S., and A. Turk. 2011. "Durational Evidence for Word-Based vs. Prominence-Based Constituent Structure in Limerick Speech." In *Proceedings of the Seventeenth International Congress of Phonetic Sciences*. Hong Kong. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/index.htm.

Shouval, H. Z., M. G. Hussain Shuler, A. Agarwal, and J. P. Gavornik. 2014. "What Does Scalar Timing Tell Us about Neural Dynamics?" *Frontiers in Human Neuroscience* 8:438.

Spencer, R. M. C., and H. N. Zelaznik. 2003. "Weber (Slope) Analyses of Timing Variability in Tapping and Drawing Tasks." *Journal of Motor Behavior* 35 (4): 371–381.

Treisman, M. 1963. "Temporal Discrimination and the Indifference Interval: Implications for a Model of the Internal Clock." *Psychological Monographs* 77 (13): 1–31.

Turk, A. 2012. "The Temporal Implementation of Prosodic Structure." In *The Oxford Handbook of Laboratory Phonology*, edited by A. C. Cohn, C. Fougeron, and M. K. Huffman, 242–253. New York: Oxford University Press.

Turk, A. E., and S. Shattuck-Hufnagel. 2000. "Word-Boundary-Related Duration Patterns in English." *Journal of Phonetics* 28:397–440.

Turk, A., and S. Shattuck-Hufnagel. 2020. *Speech Timing: Implications for Theories of Phonology, Phonetics, and Speech Motor Control*. New York: Oxford University Press.

Van Lancker, D., J. Kreiman, and D. Bolinger. 1988. "Anticipatory Lengthening." *Journal of Phonetics* 61:339–347.

Wearden, J. H. 1991. "Do Humans Possess an Internal Clock with Scalar Timing Properties?" *Learning and Motivation* 22:59–93.

White, L. 2014. "Communicative Function and Prosodic Form in Speech Timing." *Speech Communication* 63–64:38–54.

White, L., and A. E. Turk. 2010. "English Words on the Procrustean Bed: Polysyllabic Shortening Reconsidered." *Journal of Phonetics* 38:459–471.

Williams, B., and S. M. Hiller. 1994. "The Question of Randomness in English Foot Timing: A Control Experiment." *Journal of Phonetics* 22:423–439.

Windmann, A., J. Šimko, and P. Wagner. 2015. "Polysyllabic Shortening and Word-Final Lengthening in English." In *Proceedings of Interspeech* 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6–10; ISCA Archive. http://www.isca-speech.org/archive/interspeech_2015.

Wing, A. M. 1980. "The Long and Short of Timing in Response Sequences." In *Tutorials in Motor Behavior*, edited by G. E. Stelmach and J. Requin, 469–484. Amsterdam: North-Holland.

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 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