

31

The Role of Prosodic Durational Variation in the Temporal Coordination of Utterances

SALEH GHADANFARI AND LAURENCE WHITE

31.1 Introduction

Models of speech timing must reflect the mechanisms by which speakers communicate linguistic structure to listeners through systematic durational variations (e.g., Klatt, 1976; van Santen and Shih, 2000; White, 2002, 2014). Such models refer to theories of prosodic structure and to some notion of hierarchically organised prosodic constituents, such as syllables, word, phrases, and so on (e.g., Nespor and Vogel, 1986; Selkirk, 1986). Furthermore, some accounts of observed durational patterns specifically propose direct temporal influences between higher and lower prosodic constituents, for example, between syllables and some form of stress-delimited feet (e.g., O'Dell and Nieminen, 1999; Port, 2003), whereby, for example, the number of syllables within the higher-level constituent directly influences the duration of the lower-level constituent (e.g., Lehiste, 1972). Critical debates remain, however, over the degree to which prosodic constituents are strictly hierarchical and over the nature of the timing constraints that such hierarchical relations impose on speech production (e.g., Shattuck-Hufnagel and Turk, 1996; Fletcher, 2010; White and Malisz, 2020).

31.2 Coupled Oscillator Models and Temporal Compression Effects

Arguing against isochronous timing principles in (then extant) notions of 'rhythm class', Dauer (1983) reported positive relationships between inter-stress interval duration and the number of inter-stress syllables for a range of languages, whether, at the time, categorised as 'stress-timed' or 'syllable-timed'.¹

¹ The concepts of 'rhythm class' and 'stress-timing' versus 'syllable-timing' are now widely asserted to be spurious, indeed misleading – see White and Malisz (2020) for a review – and are only used here to contextualise earlier research.

Reanalysing Dauer's (1983) data using linear regression, Eriksson (1991) explicitly modelled inter-stress duration as a function of the number of syllables in an inter-stress interval: $I = a + nb$, where a is the intercept, b is the slope of the regression line, and n is the number of syllables in the inter-stress interval. Eriksson reported that the slope, representing the additional duration due to each new syllable in an inter-stress interval, was similar across languages (approximately 100 ms). He also commented on systematic linguistic differences in the intercept of the regression line: this value clustered around 200 ms in English and Thai (then so-called stress-timed languages), and at 100 ms in ('syllable-timed') Spanish, Greek, and Italian.

Eriksson (1991) asserted that the 'natural interpretation' of intercept value is that it refers to the extra duration of stressed syllables (relative to unstressed syllables) in the inter-stress interval (what we refer to here as the magnitude of durational stress contrast). However, he also observed that the intercept value does not in itself capture the locus of this additional duration, raising the possibility that linguistic variation in intercept values could (alternatively) indicate variable compression of syllables somewhere in the inter-stress interval; that is, the residual intercept durational value could be underpinned by inverse relationships between the number of syllables in the interval and their average duration. In such an account, inter-stress interval duration is a function both of duration added by each new syllable ('syllable effect') and syllabic compression due to the composition of the inter-stress interval ('inter-stress effect'; but see, for example, van Santen, 1997, and White, 2014, for arguments against the syllabic compression interpretation).

Following Eriksson's second, syllabic compression interpretation of his regression models of cross-linguistic inter-stress-interval duration, O'Dell and Nieminen (1999) attempted to capture the hypothesised timing influences on these intervals by positing two interacting oscillators, representing two levels of the prosodic hierarchy: the syllabic oscillator and the inter-stress (or stress-foot) oscillator. These oscillators are proposed to have their own natural frequencies, with the syllabic oscillator higher in frequency than the inter-stress oscillator. Importantly for the generation of observed durational patterns, the oscillators are proposed to interact with each other via a coupling function. As such, the coupled oscillators settle into stable frequency patterns in which the frequency of the faster oscillator is an integer multiple of the frequency of the slower oscillator (see Windmann, 2016). Figure 31.1 shows a schematic representation of a 1:2 ratio of the syllable oscillator to the inter-stress oscillator representing a stable state coupling. It may also be noted that these oscillators are not associated with neural or physiological processes in O'Dell and Nieminen's purely mathematical models, but there are obvious parallels with accounts of the synchronisation of the temporal structure of speech to endogenous neural oscillations (see, for example, Chapters 3 and 5).

According to O'Dell and Nieminen (1999), languages differ in which oscillator dominates as regards timing. In so-called stress-timed languages, the inter-stress oscillator would be the most dominant; thus, as the number of syllables increases in a stress group, the inter-stress oscillator tends to preserve its natural frequency and imposes frequency (and thus durational) changes on the syllabic oscillator. The opposite would be true in so-called syllable-timed languages.

In O'Dell and Nieminen's (1999) model, the relative oscillator strength parameter, r , can be estimated as the ratio of the intercept a (which reflects stress-level timing influence) to the slope b (which reflects the duration due to additional syllables in the inter-stress interval); thus: $r = a/b$

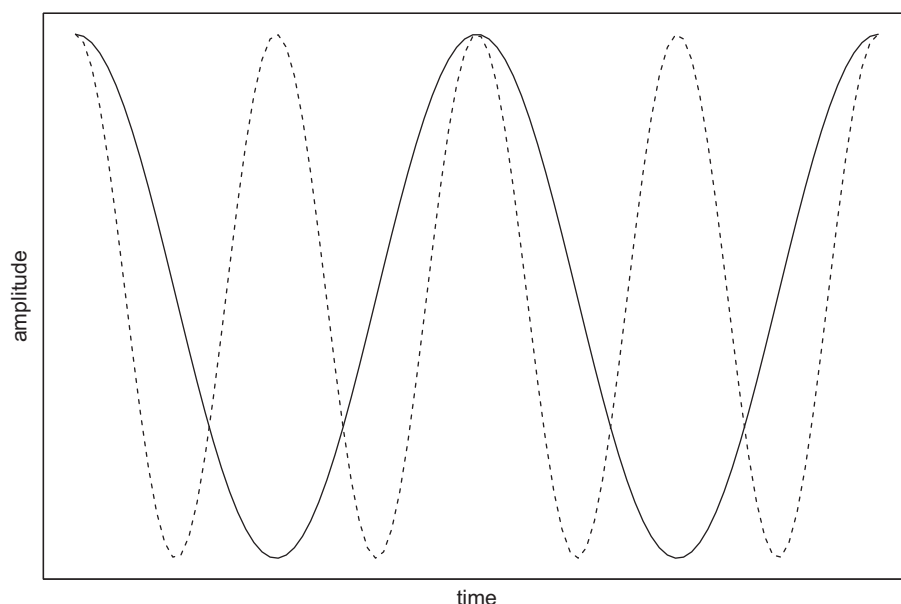


Figure 31.1 Schematic representation of coupled oscillators.

Stable state between syllabic oscillator (dashed line) and inter-stress oscillator (solid line), where the frequency of the syllabic oscillator is an integer multiple of the frequency of the inter-stress oscillator (here, for illustrative purposes only, a 1:2 ratio).

b. If $r > 1$, the stress oscillator dominates, whereas if $r \leq 1$, the syllabic oscillator dominates. O'Dell and Nieminen (1999) applied the oscillator strength parameter to Dauer's (1983) data, as reanalysed in Eriksson (1991), with the addition of data from Finnish. The r parameter value ($r = alb$) classified languages in accordance with Eriksson (1991).

O'Dell and Nieminen (2009) discuss 'polysyllabic shortening', the postulated inverse relationship between the number of syllables in some constituent and the duration of syllables therein (e.g., Lehiste, 1972; see, for example, Guba et al., 2023, for a recent study on Modern Standard Arabic). O'Dell and Nieminen take polysyllabic shortening (across the inter-stress interval) as a reflection of the interaction between syllabic and inter-stress oscillators. Evidence for polysyllabic shortening is reported, for example, in Kim and Cole (2005), where stressed syllable durations were shorter as the size of the stress group increased in English (see also, for example, Lehiste, 1972, regarding word-level polysyllabic shortening). Importantly, however, the coupled oscillators model does not hinge on the assumption of isochronous speech units; rather, compressibility effects only reflect hierarchical nesting, that is, the influence of higher prosodic units on the timing of lower prosodic units and vice versa (see Malisz et al., 2016; White and Malisz, 2020).

Despite some success in coupled oscillator modelling of such timing effects, it has been argued that observed temporal compression, as is implied by polysyllabic shortening, may be reinterpreted in terms of localised lengthening effects (e.g., Beckman, 1992; White, 2002; White and Turk, 2010). For example, Port (1981) reported polysyllabic shortening of stressed

syllables such as *dib* in nonsense sequences like *dib ... dibber ... dibberly*. It is important to note, however, that all tokens in this study were realised as the only new material in a fixed carrier phrase: ‘*I said [target word] again*’. In this (English language) context, the targets will clearly carry phrasal stress (pitch accent), which causes lengthening of constituents within the stressed word (e.g., Cambier-Langeveld and Turk, 1999; Turk and White, 1999). The degree of phrasal-stress lengthening of lexically stressed syllables has been shown to vary inversely with word length, with some of the additional length being evidenced on unstressed syllables in disyllabic and trisyllabic words (Turk and White, 1999; White and Turk, 2010). Thus, what may appear as polysyllabic shortening can be reinterpreted as due to the redistribution of phrasal-stress lengthening according to word length (White, 2002, 2014; see Beckman, 1992, for similar observations with regard to polysyllabic shortening and phrase-final lengthening).

Thus, whilst the coupled oscillators model captures hypothesised timing influences between prosodic units, the implied compressibility effects may not be supported by empirical observations. Rather, prosodic influence on speech timing primarily entails lengthening effects at domain heads (i.e., prominent constituents, such as stressed syllables and pitch-accented words) and edges (i.e., boundaries between prosodic constituents), with distribution and magnitude varying according to language-specific characteristics (for reviews, see Fletcher, 2010; White, 2014; White and Malisz, 2020).

There remain, however, aspects of the coupled oscillators model that appear potentially useful in accounting for timing patterns in circumscribed speech contexts, such as in synchronised speech or speech cycling, as discussed below. Next, we consider how phase relations between hierarchical temporal units may offer an account of certain forms of observed timing variation.

31.3 Temporal Coordination between Different Rhythmic Timescales

Our aim in this section is to show that temporal phase relations may modulate the interaction between hierarchical units in synchronised speech or other constrained speech tasks, in particular, speech cycling (Cummins and Port, 1998). Furthermore, cross-linguistic variation in the performance of such tasks may be informative about the localised timing effects that are evident in natural speech.

Regarding synchronised speech, Cummins (2003) showed that when two speakers read a text together, they synchronise their speech very effectively, often with minimal time lag (between 40 and 60 ms). Cummins further showed that the effect of synchronisation is not the result of one speaker following the speech rate of the other, as there was no consistent leader. Further work suggested that synchronisation is based on a range of suprasegmental sources of information, including fundamental frequency and amplitude envelope modulation, and is not wholly dependent on speech intelligibility (Cummins, 2009); these findings are interpreted as evidence for acoustically based ‘entrainment’ between speakers talking synchronously, although definitions of entrainment vary and typically involve phase-resetting between coupled systems of oscillators (see Obleser and Kayser, 2019, for a discussion), which are not precisely defined in Cummins’ account (see Chapter 14 for a discussion of the nature of entrainment with regard to speech).

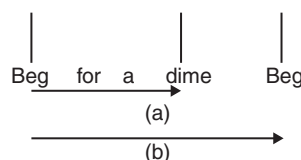


Figure 31.2 Schematic representation of speech cycling task.

Interval *a*, defined as the interval from the first stressed syllable to the final stressed syllable, is divided by interval *b* – the PRC – to calculate the phase angle of the final stress. Here, the final stress is the second stress of the phrase; in some speech cycling tasks, there are three or more stressed syllables per phrase.

Speech cycling represents another case of temporal coordination between a speaker and an external stimulus (e.g., Port et al., 1995). In speech cycling tasks, speakers repeat phrases to coordinate with metronome beeps, typically starting each new repetition of a phrase synchronously with a beep. The interval between repetition onsets is called the phrase repetition cycle (PRC). It is shown that acoustically salient points, namely stressed vowel onsets, tend to lie at certain privileged phases within the PRC (Figure 31.2). These phases typically divide the PRC into simple integer ratios, such as 1:3, 1:2, and 2:3, reflecting metrical structure within the PRC. These simple ‘phase angles’ (‘harmonic phases’) are said to be attractors for prominence in the PRC that emerge from task constraints, specifically, repeating sentences at a constant period; thus, the organisation of stress beats at privileged time intervals within the PRC reflects a hierarchical structure wherein the lower-level prosodic units, that is, stressed vowel onsets, are nested within a higher-level unit, that is, the PRC (Cummins and Port, 1998).

Languages appear to vary in the propensity for speakers to align prominent syllables at metrically important points within the PRC. For example, Cummins (2002) asked English, Spanish, and Italian speakers to read sentences with two stressed syllables, each followed by an unstressed syllable, and to align the first stressed syllable to a high-tone beep and the second stressed syllable to a low-tone beep. English speakers found the task of metrical coordination easy to perform and showed close and consistent alignment with simple harmonic phases. On the other hand, Italian and Spanish speakers found it more difficult, even after more than 30 minutes of practice, and phase alignment was not close to simple harmonic phase angles. Cummins’ explanation of the easier performance of English speakers referred to the greater salience of stress feet in English than in Italian and Spanish.

Another (alternative or complementary) explanation for these reported cross-linguistic differences in the propensity for speech cycling (Cummins, 2002) may lie in variation in the magnitude of durational contrast between strong and weak syllables. English is known to have a high durational contrast between stressed and unstressed syllables, in part due to a substantial lengthening of (lexically and phrasally) stressed syllables and vowel reduction of unstressed syllables (e.g., Oller, 1973; Klatt, 1976). On the other hand, the degree of durational contrast between strong and weak syllables in Italian and Spanish appears lower (than in, for example, English), with lower stress-related lengthening, especially for Spanish, and limited vowel reduction in unstressed syllables (Grabe and Low, 2002; White and Mattys, 2007). The placement of metronome beeps at simple phases – in particular, at the desired phrase onset – leads to the

emergence of prominence attractors (Cummins and Port, 1998); therefore, the close alignment of stressed syllables to regular metronome beats is more natural in a language such as English, where stress contrast is high, than in languages with lower stress contrast (e.g., Spanish and Italian). The lower stress contrast in the latter languages tends to make prominences (stressed syllables) less acoustically salient and thus implies less compelling coordination of prominences with attractors; in other words, strong stress contrast affords temporal coordination with regular metronome beats. (Note: this interpretation by no means denies the obvious fact that stress per se is highly salient for speakers of languages such as Spanish, which actually has many more stress-based minimal word pairs than English, but simply suggests that there may be less natural affordance for acoustically based temporal coordination of prominences.)

Zawaydeh et al. (2002) compared the speech cycling performance of speakers of American English to those of Jordanian Arabic, finding that the English speakers tended to align stressed syllables closer to a simple phase of 1:2 than Arabic speakers (who tended to have later alignment). The vowel reduction of unstressed syllables in Jordanian Arabic is of lower magnitude than English; that is, stressed and unstressed syllables have more similar durations in Jordanian Arabic (Vogel et al., 2017); thus, by the same argument as for Spanish and Italian (above), there is a lower affordance for alignment of stressed syllables to simple phases in Jordanian Arabic.

Similarly, Ghadanfari (2022) found dialectal differences in speech cycling between two varieties of Kuwaiti Arabic, Bedouin and Hadari. Specifically, there were smaller phase alignment differences between heavy and light stressed syllables for Hadari speakers than Bedouin speakers. Vowel duration analysis showed that Hadari had greater unstressed syllable reduction than Bedouin. Ghadanfari's interpretation was that unstressed syllable reduction in Hadari leads to stronger stress contrast, which affords more consistent alignment with respect to the PRC.

Ghadanfari (2022) further showed that speech rate mediates temporal coordination: at shorter metronome periods, where speech rate is likely to be faster, stressed syllables were closer to simple phase angles. This potentially reflects changes in relative durations at faster speaking rate – such as the compression of unstressed syllables – promoting the harmonic alignment of stressed syllables (and see below regarding between-language variation in speech rate influences on phase angles).

Another factor that may promote alignment to simple phases in speech cycling is phonetic compressibility: as discussed by Klatt (1973), segments may vary in duration according to context but tend to have a minimum duration below which they may not be compressed. The minimum duration of specific segments will depend not only on the manner and place of articulation but also on perceptual factors, that is, what is required to phonetically distinguish sounds within a specific language's phonemic inventory (see White, 2014, for discussion). Compressibility may also relate to prominence, in particular, whether languages have a high degree of durational stress contrast, with reduction and shortening, or even deletion, of unstressed vowels or whole syllables (such as in Standard Southern British English: e.g., Beckman, 1996).

Tajima (1999) examined how phase alignment in English and Japanese was affected by manipulation of metronome rate, from slow to fast. English speakers demonstrated consistent alignment of stressed vowel onsets with simple phase angles across different metronome rates, whilst Japanese speakers showed alignment of prominent syllables to incrementally distinct phase angles as metronome rate increased. It is plausible that the more consistent phase alignment in

English may be facilitated through relative tolerance of unstressed syllable compression with increasing metronome rate. Note that, regarding the impact of speech rate on temporal coordination, Kuwaiti Arabic dialects behaved more like Japanese than like English, given that phase alignment changed with increased rate (Ghadanfari, 2022). In Kuwaiti Arabic, however, increased rate did not lead to qualitative changes in phase alignment, as for Japanese; rather, in Kuwaiti Arabic, alignment moved closer to a harmonic phase angle with increasing rate. The compressibility of Kuwaiti unstressed syllables may thus be intermediate between those of Japanese and English, although the range of rate variation was higher in the Japanese task (>10 metronome rates (Tajima, 1999) versus three metronome rates in the Kuwaiti task (Ghadanfari, 2022)).

In summary, speech cycling studies show differences between languages, and between dialects of a particular language, in temporal coordination, as evidenced by their differential propensity to generate simple, consistent phase angles of stressed syllable alignment within the phrase repetition angle. As discussed, a plausible interpretation of these cross-linguistic and cross-dialectal differences relates to variation in durational stress contrast. From a top-down perspective, more durationally marked (and thus acoustically more salient) stressed syllables (in languages such as English) are more strongly attracted to metrically stable positions in repeated phrases. From a bottom-up perspective, languages or dialects allowing substantial compression of unstressed syllables (for example, English, in contrast with, for instance, Spanish or Japanese) provide more scope for consistent phase alignment of the stressed syllables, regardless of perturbations due to phonetic content of phrases or variation in metronome rates.

31.4 Temporal Coordination in Natural Dialogues

The nature of temporal coordination in speech cycling, that is, the division of the PRC into simple phases, is clearly specific to the constrained task demands. However, the interaction between interlocutors in natural dialogue has also been suggested to reflect patterns of temporal coordination, as the timing characteristics of a dialogue partner's speech are proposed to influence the timing of turn-taking of the other interlocutors (e.g., Wilson and Wilson, 2005). In this final section, we consider potential commonalities in the timing factors influencing temporal coordination, particularly speech rate and local durational cues, in speech cycling and natural dialogues.

Research on temporal coordination in natural dialogues has focused on the fluent timing of turn transitions between speakers (e.g., Wilson and Zimmerman, 1986; Couper-Kuhlen, 1993; Benuš, 2009). The reportedly minimal gaps (suggested to average 200 ms across languages – for example, Stivers et al., 2009; Heldner and Edlund, 2010) and relatively rare overlaps in interlocutors' turns implies adaptation to the current speaker's rate and anticipation of their utterance termination (Wilson and Wilson, 2005). Moreover, Wilson and Wilson suggested a predictive mechanism in conversational turn-taking by which listeners entrain to the syllable oscillation rate of the speaker; thus, to avoid overlap, listeners coordinate the onset of their turns in anti-phase relation to the speaker's syllable rate.

A related, but distinct, perspective on turn-transition timing was provided in studies of overlapped (i.e., interrupted) conversational transitions (Włodarczak et al., 2012a, 2012b) using corpora of spontaneous speech of American English, German, and French. They first

observed that the initiation of an interrupting turn showed a bias to occur at the end of a vowel-to-vowel (VTV) interval in the preceding speech. Using the normalised pairwise variability index (nPVI; Low et al., 2000) of VTV duration, they found that more regular VTV timing was associated – for English dialogues – with this predominant pattern of a late interruption point in the VTV interval (Włodarczak et al., 2012a). Reinforcing their conclusions with analyses of French and German dialogues, they interpret this pattern as evidence that coordination of turn-timing is underpinned by temporal entrainment between speakers (Włodarczak et al., 2012b). They additionally argue that such entrainment is, in particular, governed by the salient recurrence of the perceptual centre of syllables (p-centre) (Morton et al., 1976; Marcus, 1981).

With regard to durational cues to utterance boundaries, it has long been demonstrated that localised final lengthening may contribute to the salience of phrase or utterance endings (e.g., Price et al., 1991). It appears surprising, therefore, that Hoogland et al. (2023) found that inter-speaker intervals in question-answer sequences in Dutch and English were *longer* with longer final rhymes. However, they also reported an interaction with articulation rate of the preceding utterance: thus, at faster rates, inter-speaker intervals were *shorter* when the final rhyme was relatively long. The interpretation of Hoogland et al. of this interaction related to potential listener entrainment to foregoing speaking rate: as segments are shorter at faster speaking rate, the *relative* length of phrase-final segments is boosted (see Dilley and Pitt, 2010; Reinisch et al., 2011; Morrill et al., 2015), thus providing a more salient cue to question termination. Thus, local timing cues to turn-ending may potentially be mediated by listener entrainment to the foregoing utterance, at least insofar as it is required to develop expectations regarding segment duration.

31.5 Summary, Future Research, and Conclusions

The review presented here considers various influences on speakers' temporal coordination patterns in artificial tasks, such as speech cycling, and in natural dialogue turn-taking. Speech cycling indicates cross-linguistic variability in speakers' propensity to coordinate the occurrence of stressed syllables within an external cycle, with some languages' greater length of stressed syllables and greater compressibility of unstressed syllables both potentially contributing to the more consistent alignment of stresses with simple phase angles of the PRC (e.g., Tajima, 1999; Cummins, 2002; Ghadanfari, 2022).

Regarding natural dialogue turn-taking, existing accounts point to relative consistency of *mean* turn-transition time between languages (e.g., Stivers et al., 2009). There are relatively few analyses of the factors that influence the variation of turn-transition time around the reported mean; however, those that have analysed these point to influences of local timing factors towards the ends of utterances, potentially interacting with foregoing articulation rate (e.g., Włodarczak et al., 2012a, 2012b; Hoogland et al., 2023). How cross-linguistic variation in local timing effects serves to mediate fluent turn-timing has been little explored to date, not least due to the difficulties of comparing between spontaneous corpora of distinct languages, elicited using different methods and often for distinct research goals.

This review has suggested that findings from artificial tasks, such as speech cycling, may point to influences on coordination in natural conversation. One potential focus of research

relates to cross-linguistic and cross-dialectal differences in structural factors influencing coordination. It has been established that speech rate affects the interpretation of local timing cues that signal structure (e.g., word and phrase boundaries). Reinisch et al. (2011) showed that the perception of stress is modulated by speaking rate, with faster foregoing utterance rate increasing the likelihood of listeners perceiving stress contrast. As languages and dialects differ in the magnitude of durational stress contrast, it is worth investigating if speech rate variation differentially influences stress perception between languages and between dialects.

Future cross-linguistic research could likewise benefit from experimentally controlled tasks that probe the influence of specific timing cues on temporal coordination. For example, paradigms requiring coordination of speech with movement (e.g., Allen, 1972a, 1972b; Rathcke et al., 2021) offer a means of limiting cross-linguistic task variation in the interests of discerning how native language experience affects listeners' temporal coordination behaviour. Likewise, artificial language tasks can manipulate timing cues whilst keeping segmental stimuli consistent between languages (e.g., White et al., 2020).

It is obvious that speakers of all languages are skilled at coordinating the flow of conversation. Artificial speech-based coordination tasks may be an effective means of unpicking the diverse cues that listeners use to achieve such interactional fluency.

Box 31.1 Chapter Overview

Summary

Languages vary in their distribution and realisation of prominent versus less prominent syllables, and in the magnitude of local timing processes such as phrase-final lengthening. These timing differences influence speakers' performance on temporally constrained artificial tasks such as speech cycling, and may have implications for the coordination of natural conversation.

Implications

It is obvious that speakers of all languages are skilled at coordinating the flow of conversation, but the mechanisms by which this temporal coordination is achieved remain unclear. Artificial speech-based coordination tasks may represent an effective means of unpicking the diverse cues that listeners use to achieve such interactional fluency.

Gains

Making cross-linguistic comparisons of speech timing is challenging given the diversity of structural and realisational differences. The degree to which speakers consistently coordinate prominent syllables within an externally imposed cycle in laboratory tasks is informative about the magnitude of stronger versus weaker syllable contrasts, and – potentially – about temporal coordination in conversation.

REFERENCES

- Allen, G. D. (1972a). The location of rhythmic stress beats in English: An experimental study I. *Language and Speech*, 15, 72–100.
- Allen, G. D. (1972b). The location of rhythmic stress beats in English: An experimental study II. *Language and Speech*, 15, 179–195.
- Beckman, M. E. (1992). Evidence for speech rhythms across languages. In Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 457–463). Amsterdam: IOS Press.
- Beckman, M. E. (1996). When is a syllable not a syllable? In T. Otake and A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp. 95–123). Berlin: de Gruyter.
- Beňuš, Š. (2009). Variability and stability in collaborative dialogues: Turn-taking and filled pauses. *Proceedings of Interspeech 2009*, pp. 796–799.
- Cambier-Langeveld, T., and Turk, A. E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics*, 27(3), 255–280.
- Couper-Kuhlen, E. (1993). *English Speech Rhythm: Form and function in everyday verbal interaction*. John Benjamins.
- Cummins, F. (2002). Speech rhythm and rhythmic taxonomy. *Proceedings of Speech Prosody 2002*, Aix-en-Provence, pp. 121–126. <https://doi.org/10.21437/SpeechProsody.2002-17>
- Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics*, 31(2), 139–148.
- Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37(1), 16–28.
- Cummins, F., and Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26(2), 145–171.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51–62.
- Dilley, L. C., and Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21, 1664–1670.
- Eriksson, A. (1991). Aspects of Swedish speech rhythm. Doctoral dissertation, University of Göteborg.
- Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, and F. E. Gibbon (Eds.), *The handbook of phonetic sciences*, second edition (pp. 521–602). Hoboken, NJ: Wiley-Blackwell.
- Ghadanfari, S. (2022). Hierarchical timing in varieties of Kuwaiti Arabic. Doctoral dissertation, Newcastle University.
- Grabe, E., and Low, E. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and Natasha Warner (Eds.), *Papers in laboratory phonology VII* (pp. 515–546). Berlin and New York: Mouton de Gruyter.
- Guba, M. N. A., Mashaqba, B., and Huneety, A. (2023). Polysyllabic shortening in Modern Standard Arabic. *Journal of Semitic Studies*, LXVIII(2), 759–770.
- Heldner, M., and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38, 555–568.
- Hoogland, D., White, L., and Knight, S. (2023). Speech rate and turn-transition pause duration in Dutch and English spontaneous question-answer sequences. *Languages*, 8, 115.
- Kim, H., and Cole, J. (2005). The stress foot as a unit of planned timing: Evidence from shortening in the prosodic phrase. *Proceedings of Interspeech 2005*, Lisbon, pp. 2365–2368. <https://doi.org/10.21437/Interspeech.2005-37>
- Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America*, 54, 102–1104.

- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59(5), 1208–1221.
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51(6B), 2018–2024.
- Low, E. L., Grabe, E., and Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43(4), 377–401.
- Malisz, Z., O'Dell, M., Nieminen, T., and Wagner, P. (2016). Perspectives on speech timing: Coupled oscillator modeling of Polish and Finnish. *Phonetica*, 73(3–4), 229–255.
- Marcus, S. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception and Psychophysics*, 30(3), 247–256.
- Morrill, T., Baese-Berk, M., Heffner, C., and Dille, L. (2015). Interactions between distal speech rate, linguistic knowledge, and speech environment. *Psychonomic Bulletin & Review*, 22, 1451–1457.
- Morton, J., Marcus, S., and Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, 83(5), 405–408.
- Nespor, M., and Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
- Obleser, J., and Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, 23(11), 913–926.
- O'Dell, M. L., and Nieminen, T. (1999). Coupled oscillator model of speech rhythm. Proceedings of the XIVth International Congress of Phonetic Sciences 2, American Institute of Physics, pp. 1075–1078.
- O'Dell, M. L., and Nieminen, T. (2009). Coupled oscillator model for speech timing: Overview and examples. *Nordic Prosody: Proceedings of the 10th Conference*, Helsinki, pp. 179–190. <https://doi.org/10.3726/978-3-653-00335-2/19>
- Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54(5), 1235–1247.
- Port, R. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America*, 61(1), 262–274.
- Port, R. (2003). Meter and speech. *Journal of Phonetics*, 31(3), 599–611.
- Port, R., Cummins, F., and Gasser, M. (1995). A dynamic approach to rhythm in language: Toward a temporal phonology. arXiv. <https://doi.org/10.48550/arXiv.cmp-lg/9508007>
- Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90(6), 2956–2970.
- Rathcke, T., Lin, C., Falk, S., and Bella, S. (2021). Tapping into linguistic rhythm. *Laboratory Phonology*, 12(1), 1–32.
- Reinisch, E., Jesse, A., and McQueen, J. M. (2011). Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech*, 54, 147–165.
- Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, 3, 371–405.
- Shattuck-Hufnagel, S., and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2), 193–247.
- Stivers, T., Enfield, N.J., Brown, P., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106, 10587–10592.
- Tajima, K. (1999) Speech rhythm in English and Japanese: Experiments in speech cycling. Doctoral dissertation, Indiana University.
- Turk, A., and White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27(2), 171–206.
- van Santen, J. P. H. (1997). Segmental duration and speech timing. In Y. Sagisaka, N. Campbell, and N. Higuchi (Eds.), *Computing prosody: Computational models for processing spontaneous speech* (pp. 225–249). New York: Springer-Verlag.

- van Santen, J. P. H., and Shih, C. (2000). Suprasegmental and segmental timing models in Mandarin Chinese and American English. *Journal of the Acoustical Society of America*, 107(2), 1012–1026.
- Vogel, I., Athanasopoulou, A., and Pincus, N. (2017). Acoustic properties of prominence and foot structure in Arabic. In H. Ouali (Ed.), *Perspectives on Arabic linguistics XXIX: Papers from the Annual Symposium on Arabic Linguistics, Milwaukee, Wisconsin, 2015* (pp. 55–88). John Benjamins. <https://doi.org/10.1075/sal.5.04vog>
- White, L. (2002). English speech timing: A domain and locus approach. Doctoral dissertation, University of Edinburgh.
- White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication*, 63–64, 38–54.
- White, L., and Malisz, Z. (2020). Speech rhythm and timing. In C. Gussenhoven and A. Chen (Eds.), *Oxford handbook of language prosody* (pp. 167–182). Oxford: Oxford University Press.
- White, L., and Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35(4), 501–522.
- White, L., and Turk, A. E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, 38(3), 459–471. <https://doi.org/10.1016/j.wocn.2010.05.002>
- White, L., Benavides-Varela, S., and Mády, K. (2020). Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues? *Journal of Phonetics*, 81, 100982.
- Wilson, M., and Wilson, T. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, 12(6), 957–968.
- Wilson, T. P., and Zimmerman, D. H. (1986). The structure of silence between turns in two-party conversation. *Discourse Processes*, 9(4), 375–390.
- Windmann, A. (2016). Optimization-based modeling of suprasegmental speech timing. Doctoral dissertation, Bielefeld University.
- Włodarczak, M., Simko, J., and Wagner, P. (2012a). Temporal entrainment in overlapped speech: Cross-linguistic study. Proceedings of Interspeech 2012, Portland, OR, pp. 615–618. <https://doi.org/10.21437/Interspeech.2012-186>
- Włodarczak, M., Simko, J., and Wagner, P. (2012b). Syllable-boundary effect: Temporal entrainment in overlapped speech. Proceedings of Speech Prosody 2012, Shanghai, China, pp. 611–614. <https://doi.org/10.21437/SpeechProsody.2012-154>
- Zawaydeh, B. A., Tajima, K., and Kitahara, M. (2002). Discovering Arabic rhythm through a speech cycling task. In D. Parkinson and E. Benmamoun (Eds.), *Perspectives in Arabic linguistics XIII–XIV* (pp. 39–58). Amsterdam: John Benjamins.