



Review

Cite this article: Krivokapić J. 2014 Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Phil. Trans. R. Soc. B* **369**: 20130397. <http://dx.doi.org/10.1098/rstb.2013.0397>

One contribution of 14 to a Theme Issue 'Communicative rhythms in brain and behaviour'.

Subject Areas:
cognition

Keywords:
prosodic structure, prosodic boundaries, speech planning, gestural coordination, manual gesturing, articulatory gestures

Author for correspondence:
Jelena Krivokapić
e-mail: jelenak@umich.edu

Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes

Jelena Krivokapić^{1,2}

¹Department of Linguistics, University of Michigan, 440 Lorch Hall, 611 Tappan St., Ann Arbor, MI 48109-1220, USA

²Haskins Laboratories, 300 George Street No. 900, New Haven, CT 06511, USA

Prosodic structure is a grammatical component that serves multiple functions in the production, comprehension and acquisition of language. Prosodic boundaries are critical for the understanding of the nature of the prosodic structure of language, and important progress has been made in the past decades in illuminating their properties. We first review recent prosodic boundary research from the point of view of gestural coordination. We then go on to tie in this work to questions of speech planning and manual and head movement. We conclude with an outline of a new direction of research which is needed for a full understanding of prosodic boundaries and their role in the speech production process.

1. Introduction

The goal of this review is to investigate the coordination of articulatory gestures at prosodic boundaries and to examine how these coordinations can inform us about prosodic structure and the broader context of language use, such as speech planning. The focus is predominantly on articulatory gestures, but recent work on manual and head gesturing is also examined. This is a necessarily limited selection of research; for reviews focusing on other aspects of prosodic structure (e.g. the acoustic properties of prosodic structure, the syntax–prosody interface, different accounts of prosodic structure that have been proposed) see Shattuck-Hufnagel & Turk [1], Wagner & Watson [2] and Fletcher [3].

The linguistic term *prosody* refers to the suprasegmental structure of the utterance, encoding prominence and phrasal organization [4–7]. Here, the focus is on phrasing and prominence above the word level (see e.g. [6,8,9] for discussion of prosodic properties at the word level). In English, prominence highlights important or new information and is also used for rhythmic purposes.^{1,2} Phrasal organization serves to group words together into appropriate chunks that are used by listeners and speakers for language processing.

The focus of this review is on phrasal organization, specifically on the properties of prosodic boundaries that mark these phrases. An example of prosodic phrasing and the importance of prosodic boundaries in deducing the meaning of these utterances is given in the sentences below. Note that sentence (ii) has two more boundaries, marked by commas, than sentence (i).

- (i) She knew David thought about the present.
- (ii) She knew, David thought, about the present.

The phrases form a prosodic hierarchy, with larger units dominating smaller units. We start with a brief outline of the prosodic hierarchy in §2, describe how the π -gesture model [7] accounts for the temporal properties of prosodic boundaries (§3a), and then in §3b,c turn to more recent studies examining how linguistic gestures at prosodic boundaries are coordinated (e.g. [13–15]). A discussion of the role of pauses in speech planning and of the articulatory properties of pauses is given in §4a,b. We bring these lines of research together

by investigating the coordination of body movements (e.g. manual gesturing) with prosodic structure [16,17] in §4c and by presenting a small study of head and manual gesturing at prosodic boundaries in §4d. In the remainder of the paper, we discuss the implications that gestural coordinations have for our understanding of prosodic boundaries in a broader context. We suggest directions for research that will help address some of the central issues for prosodic theory by taking gestural coordination at boundaries as a starting point.

2. The prosodic hierarchy

Within an utterance, words group together into phrases, which in turn group into larger phrases, forming a prosodic hierarchy. Research approaches differ as to how they define these prosodic units. Within one approach, they are determined based predominantly on syntactic structure [18,19], whereas another approach identifies them based on intonational properties (e.g. [20,21]; see [22,23] for a discussion of these two approaches).

As to factors influencing prosodic structure, it is known that syntactic structure determines prosodic structure to a large extent. Additional factors have been found to influence prosodic phrasing in terms of both occurrence and strength of prosodic boundaries. For example, speech rate (for English [24] for French [25] and for German [26]) and phrase length [27–30] are well known to have an effect on prosodic phrasing.³ More recently, phrase boundaries across an utterance have been shown to influence how individual boundaries are produced or perceived, indicating global effects of prosodic structure on speech production and perception [30–34]. For example, listeners' interpretation of ambiguous sentences depends on the relative strength of prosodic boundaries, i.e. on the strength of one boundary in relation to another [31].

Over the years, a number of different prosodic hierarchies have been proposed (see the overview in [1]; see also [9,35,36] for a discussion on a more gradient approach to the prosodic hierarchy). For English, there are at least two categories above the word level although further additional categories have been proposed. We follow Beckman & Pierrehumbert [20] and refer to these two categories as the intermediate phrase (ip) and intonation phrase (IP). A schematic representation of the hierarchy is given in figure 1. In this model, the IP is the highest prosodic category. Specific tonal properties mark the phrasal units. The ip has at the minimum a nuclear pitch accent. Pitch accents signal prominence, and an ip can have additional pitch accents; the nuclear pitch accent is by definition the last pitch accent in the ip. The end of the ip is marked by a phrase accent. The IP consists of one or more intermediate phrases, and its end is marked by a boundary tone (and, although rarely, the beginning can also be marked with a boundary tone).

In addition to tonal properties, prosodic phrases have characteristic temporal properties. In the acoustic domain, boundaries are marked by lengthening of segments, i.e. segments at the boundary are longer compared with the same segments phrase-medially [37–40]; this phenomenon is also referred to as phrase-final and phrase-initial lengthening. Strong boundaries (IP boundaries) often have pauses in addition to the lengthening. Work in articulation has shown that boundary-adjacent gestures become longer and less overlapped [41–49]. The temporal effects grade in

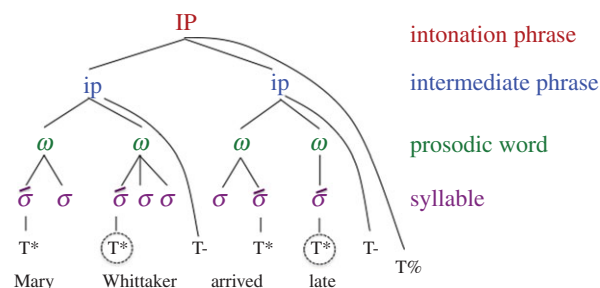


Figure 1. Prosodic structure for English ([20] model). T^* stands for pitch accents, T^- for phrase accents and $T\%$ for boundary tones. The nuclear pitch accent is circled. The structure represents a production with a weak (ip) boundary after 'Whittaker' and a strong boundary (IP) after 'late'. (Online version in colour.)

magnitude, reflecting hierarchical embedding such that prosodically higher categories show more lengthening than lower categories, and distinguishing several degrees of lengthening (e.g. phrase-final lengthening: [30,38,42,44]), and phrase-initial: (e.g. [30,41,42,44,48–50]).

3. Prosodic gestures

(a) The π -gesture model

The π -gesture model provides an account of the properties of prosodic boundaries [7]. While various other conceptualizations of prosodic boundaries have been proposed [18–20] (see reviews in [1,3]), the π -gesture model is discussed here because it clearly defines prosodic boundaries (for a discussion of the lack of definition of boundaries in the literature, see [9,51]). Furthermore, it has explicit temporal properties and allows the examination of the coordination of prosodic events. Finally, the model allows for a structurally more gradient prosodic hierarchy, which is in line with experimental evidence (see discussion in [9,35,52]). The π -gesture model has been developed within the Articulatory Phonology framework [53–55], where the basic phonological unit is a gesture, which specifies a constriction target as its goal (e.g. for alveolar consonants, a tongue tip constriction constitutes the constriction target). Gestures in Articulatory Phonology are both units of information, specifying lexical contrast, and units of action, with specified temporal and spatial information. That is, gestures are lexical units parametrized both phonetically and phonologically, such that there is no need for a translational component that traditionally might be posited as mediating between phonology and phonetics. Byrd & Saltzman [7] propose that prosodic boundaries are viewed as gestures. The prosodic gesture (π -gesture) does not instantiate constriction targets, but instead, the effect of the π -gesture is to slow the time course of the constriction gestures that are active at the same time as the π -gesture. Like constriction gestures, π -gestures extend over a temporal interval. As a consequence of these properties, gestures at boundaries become temporally longer and less overlapped. Stronger prosodic boundaries have a stronger π -gesture activation, which in turn leads to boundary effects becoming stronger. That is, at hierarchically higher boundaries, there is more lengthening and less overlap than at lower boundaries, thus accounting for the empirical findings on the temporal properties of boundaries. As mentioned above, the model also allows for a structurally gradient or

categorical prosodic hierarchy [52]. Byrd [52] discusses that the gradient structure could arise by allowing the π -gesture to have a continuum of activation strength values, whereas the categorical structure could be achieved by allowing a small number of attractors for the activation strength of the π -gesture. Alternatively, the π -gesture could have a continuum of activation strength values, but different tonal properties (phrase accent, boundary tone) would distinguish between different types of prosodic categories (see [30,52] for further discussion). The π -gesture suggested a novel way of viewing prosodic boundaries as inherently temporal, and it makes specific predictions about the behaviour of gestures at boundaries. One of the predictions of the model is that the temporal effects of the boundary extend over a period of time, and that the slowing effects decrease with distance from the boundary, corresponding to the activation shape of the π -gesture, shown in figure 2 (see [7]). The predictions of this model have been born out in a number of articulatory (e.g. [14,30,56]) and acoustic studies (e.g. [57–59] for Hebrew and [60] for Dutch).

(b) Tone gesture and μ -gesture

Recent work has further developed the idea that prosodic structure can be understood in terms of gestures. Saltzman *et al.* [61] have modelled lexical stress using a temporal modulation gesture (μ -gesture), which, like the π -gesture, slows the constriction gestures that are co-active with it. Gao [13] extends the gestural approach to tones in Mandarin. Thus, tones are viewed as dynamical systems [62–64], that, like other gestures, extend in time. They have as their goal tonal targets and are coordinated with constriction gestures [13]. This approach has also been used to investigate Catalan and German pitch accents [65], and boundary tones in Greek [14,66].

These developments allow us to consider the relationship among prosodic events, and between prosodic events and constriction events, in terms of gestures and gestural coordination (where gestural coordination refers to the timing of gestures to each other). We can also address the question of how prosodic events and articulatory constriction events are coordinated with other body movements. Examining these relationships is critical for any theory of prosodic structure, but the Articulatory Phonology framework lends itself to these questions owing to the dynamic properties of the model. Gestural coordination has been implemented in Articulatory Phonology with a coupled oscillator system, in which gestures are associated with oscillators and gestural coordination emerges through the coupling of these oscillators [67–69].

(c) The coordination of prosodic events

The relationship between prominence and boundaries is a crucial area of research, because a model of prosodic structure can be developed only if the interactions between various prosodic events are known. It has been suggested, for example, that the onset of prosodic boundaries might be triggered by lexical stress (this could be achieved by coordinating the π -gesture with the last lexically stressed syllable in a prosodic phrase; see [30,52,66]).⁴ It has also been argued that prominence and phrasing have similar functions cross-linguistically [5,6,9,22,70]. The current discussion is limited to the interaction of boundary and prominence in the

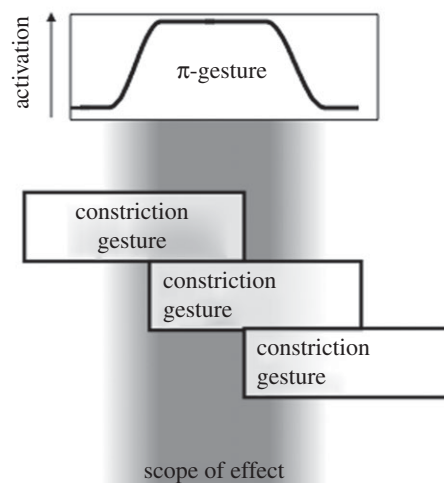


Figure 2. A schematic representation of the π -gesture [7]. The arrow indicates the activation strength of the gesture, and the scope of its effect is indicated by the grey area. Darkness in the area reflects the strength of the effect.

pre-boundary domain, as the relationship between the two phenomena has been less examined in the post-boundary domain (but see [15,71]).

Systematic investigations of the interaction between boundaries and prominence have started only recently. In an acoustic study of English, Turk & Shattuck-Hufnagel [72] examined final lengthening in words with final, penultimate and antepenultimate lexical stress and found that final lengthening starts earlier in the word when the stressed syllable is further away from the boundary. Unstressed syllables between the stressed and the final syllable were not always lengthened, or were lengthened less than the stressed syllable. Rusaw [73] replicated these findings. Similar results were obtained in White's [74] study, but contrary to Turk & Shattuck-Hufnagel [72] and Rusaw [73] the intervening unstressed syllables in his study also lengthened. Byrd & Riggs [15] in an articulatory magnetometer study also found that final lengthening can start earlier when the prominent syllable is earlier in the phrase-final word. In an articulatory magnetometer study of Greek, Katsika [14] found that the onset of final lengthening was during the final syllable in words with final stress, but that final lengthening started earlier when the prominent syllable was earlier in the phrase-final word (consistent with [15,72–74]). Like White [74], she also found that intervening unstressed syllables lengthened. Katsika [14] further examined the onset of the boundary tone in IP boundaries. She found that the boundary tone always occurred during the final vowel, but that it occurred earlier within the vowel if the stressed syllable occurred earlier in the word.

While the results of these studies are not consistent, they all suggest an interaction of prosodic events. Rusaw [73] modelled the interaction between boundary and prominence with an artificial central pattern generator neural network model. In the model, prosodic boundaries increase the effects of prosodic prominence, achieving precisely the findings of Turk & Shattuck-Hufnagel [72] and Rusaw [73]. Within the Articulatory Phonology framework, Byrd & Riggs [15] suggested two possible accounts of their findings. Either the π -gesture shifts towards the prominent syllable, keeping the scope of lengthening constant, or, alternatively, the π -gesture extends towards the pre-boundary prominent syllable,

which would mean that the scope of the π -gesture increases if the prominent syllable is further away from the boundary. In the most comprehensive account of prosodic boundaries to date, Katsika [14,66] argues that the boundary properties arise through the coordination of the μ -gesture, the boundary tone and the π -gesture. In this account, the π -gesture is coordinated with the phrase-final vowel gesture and it is also coordinated, although more weakly, with the temporal modulation gesture of the stressed syllable (the μ -gesture) of the phrase-final word. These coordinations account for the flexible scope of lengthening in that the onset of lengthening will depend on the position of the stressed syllable, i.e. the onset will shift towards the stressed syllable owing to the coordination of the π -gesture with the μ -gesture. The boundary tone gesture is activated when the π -gesture reaches a certain threshold activation level; that is, it will be activated only at strong boundaries (see also [52] for the relationship between π -gesture activation and boundary tone).⁵ The onset of the boundary tone will also, owing to its being triggered at a certain level of activation of the π -gesture, shift with stress: given that the π -gesture will reach its activation earlier if it starts earlier, the boundary tone will also be activated earlier, accounting for the observed boundary tone shift (for further discussion, see [14] and [66]). Thus, Katsika suggests that prosodic boundaries can be conceived of as a set of prosodic gestures that interact in a dynamic way [14,66]. Through this interaction, the temporal and tonal properties that characterize boundaries emerge.

These empirical studies [14,15,72–74], the accounts within the Articulatory Phonology approach [14,15,66] and Rusaw's artificial central pattern generator neural network model [73] all show that prosodic events are interdependent and need to be investigated in a holistic manner for a full understanding of prosodic structure.

4. Cognitive functions of prosodic boundaries

(a) Pauses and speech planning

At boundaries, acoustic segments and articulatory gestures become longer (i.e. phrase-final and phrase-initial lengthening), and strong prosodic boundaries can contain pauses. Pauses are of particular interest because they can inform us about the cognitive function of prosodic boundaries in speech production, specifically about the speech planning processes. In what follows, we review work on pauses in speech planning and continue with an examination of how recent work on articulation during pauses has the potential to illuminate the process of speech planning and the distinction between prosodic boundaries and other types of pauses.

Boundaries mark linguistic organizational units or 'constituents' for both listeners and speakers. Listeners can use prosodic boundaries in sentence disambiguation (e.g. [31,75–80], for an overview, see [81]). In terms of the speaker, starting with work by Goldman Eisler (see [24] for a summary), it has been argued that speakers plan speech at prosodic boundaries. Thus, pause duration and sentence initiation times increase when the upcoming chunk of speech is syntactically more complex [27,82], prosodically more complex [29,30,83]), or when the upcoming chunk of speech is longer in terms of number of syllables [28–30,82,84,85]. Cooper & Paccia-Cooper [86] examined both

pauses and final lengthening and found that upcoming syntactic complexity led to increased final lengthening, even though the pauses were short and did not themselves show lengthening. These findings indicate that pauses, and more generally prosodic boundaries, allow speakers to plan the upcoming chunk of speech; the more planning that chunk requires (owing to its complexity, length, etc.), the longer the pause (and the boundary) is going to be [24,29,30,82–88].

Planning is not the only factor affecting pause duration. As mentioned in §2, it is known that the structure of the utterance *preceding* the pause (and therefore not related to planning) also affects pause duration. This has been found for syntactic structure [86,89], for prosodic structure [29,30], and for phonological length [28–30]. It has been suggested that these effects are owing to the speaker's need to deactivate processed information [29,86,88].

Drawing on the distinction between structural and planning effects, Ferreira [82,87,88,91] suggests that pauses consist of two parts. The first part of the pause is the grammatical part, and the second is the planning, or performance, part. This suggestion is based in part on her findings that boundary-induced final lengthening is determined by the prosodic structure of the preceding phrase, rather than the syntactic complexity of the following phrase, indicating that final lengthening is not affected by speech planning [82,87]. By contrast, pause duration seems to be affected by both the complexity of the preceding phrase and the complexity of the upcoming phrase, indicating that planning takes place here. Thus, Ferreira suggests that final lengthening and pause duration are evidence of different properties, with final lengthening and the first part of the pause reflecting prosodic structure, and the second part of the pause reflecting the planning of the upcoming speech. Ferreira's findings [82,87] about final lengthening not being affected by upcoming material, and therefore not reflecting speech planning, are in contrast with findings from other studies (e.g. [86] see also [72]). It is also known that planning is a continuing process that takes place throughout the sentence [84,90,92,93] and not just at the boundary. Thus, the differential account of the lengthening and pausing phenomena put forth by Ferreira [88,91] might be too strong. Nevertheless, the idea that a boundary can be understood as divided into different 'cognitive' parts is intriguing.

Related to the question of the cognitive function of prosodic boundaries is the distinction between prosodic boundaries and disfluent pauses. Some cases are clear-cut (for example, pauses in read, highly practiced speech, are typically prosodic boundaries; similarly, pauses containing *uh*, *um*, are clearly disfluencies). A disfluency should not count as a boundary as it is an indicator of a breakdown in the planning or production process. However, the fact that planning also takes place at boundaries, and that not every disfluent pause contains *uh* or *um*, makes boundaries and disfluencies often difficult to distinguish. The question arises then whether there are articulatory correlates of the different processes taking place at boundaries. These aspects of pauses cannot be distinguished using acoustic data only (see also [88] on this point), as any articulatory behaviour indicative of these differences in the functions of the pauses will be too minute to observe in the acoustic signal. In order to learn more about the correlates of planning, a better understanding of articulation during prosodic boundaries is needed.

(b) Articulation in pauses

Advances in articulatory measurement techniques, particularly in real-time magnetic resonance imaging (MRI) [94], have made it possible to obtain a full view of the midsagittal vocal tract, which in turn allows for an adequate examination of changes that the vocal tract might be undergoing during pauses. A real-time MRI study of spontaneous speech by Ramanarayanan *et al.* [95,96] has found that the position of the articulators during grammatical pauses (defined as pauses that occur between syntactic constituents and excluding hesitation and word search pauses, in other words, pauses that can be assumed to be prosodic boundaries) show least variability, followed by ready-to-speak postures (i.e. postures of articulators that speakers adopt immediately before starting to speak, see also [97,98], followed by rest positions). As pointed out by Ramanarayanan *et al.* [95], lower variability can be understood as more control over the vocal tract, indicating that pauses at prosodic boundaries are planned, whereas the ready-to-speak pauses are less planned. In another real-time MRI study of spontaneous speech, Ramanarayanan *et al.* [99] found that grammatical pauses, but not ungrammatical ones (such as hesitation pauses), showed a significant decrease in speed of articulators at the pause compared with the pre-pausal region. The speed of articulators increased for the post-pause region for both types of pauses. Ungrammatical pauses also showed larger variability than grammatical pauses. These findings indicate, as Ramanarayanan *et al.* [99] point out, that grammatical, but not ungrammatical, pauses are planned. This work has the potential to indicate detailed vocal tract positioning during boundaries, which in turn could be taken as indicators of speech planning processes. For example, a decrease in the speed of articulators at the pause could be an indicator of the start of a predominantly planning stage in the speech production process, whereas an increase in the speed could be an indicator of a predominantly articulatory stage.

Katsika [14,66] has also examined articulation in pauses. She has identified for Greek a pause posture, which is a specific configuration of the vocal tract during grammatical pauses. The achievement of the pause posture and the end of phonation occur at a stable time from the onset of the boundary tone across all prosodic conditions examined. Katsika suggests that the pause posture is triggered by a π -gesture with a specific, very high activation level (a higher level than the one that triggers boundary tones), and therefore occurs only at strong intonation phrases and at a stable interval from the onset of the boundary tone. The achievement of the pause posture deactivates the glottal gestures (boundary tone and phonation). Katsika's [14,66] study adds a further articulatory component to the study of boundaries.

The studies of Ramanarayanan *et al.* [95,99] and Katsika [14] begin to illuminate articulation during pauses. Research of this kind will be crucial for gaining insights into the nature of prosodic boundaries and has already started informing us as to what constitutes a grammatical boundary and what is less planned articulatory behaviour during silent intervals.

(c) Gesturing and prosodic structure

In addition to research on vocal tract articulation, examining movement of other parts of the body, such as hand, head and torso movements at prosodic boundaries can shed further light on the properties of planning at prosodic boundaries

and of disfluencies. We will refer to these movements as co-speech gestures or gesturing (to be distinguished from the linguistic gestures discussed so far). Gesturing research has typically focused on movement that contributes to the utterance meaning in some way (e.g. [100], p. 109 defines co-speech gestures as 'actions that are treated by coparticipants in interaction as part of what a person meant to say'; see overview in [101] for what constitutes a gesture in gesturing research). Other kinds of movements have also been examined in relation to speech, such as finger tapping, head movements and torso movements.

Previous research has demonstrated the relevance of gesturing in communication [101–106]. A number of studies have examined the role of gesturing in speech planning, arguing that it facilitates lexical retrieval [107–109] and conceptual planning of an utterance [110]. Evidence that gesturing is related to speech planning is also found in studies examining speech errors. For example, in disfluent speech, gesturing is suspended before speech is interrupted and speakers hold their co-speech gestures during disfluencies [111]. Tiede *et al.* [112] found that head movement amplitude increased near speech errors.

Recently, gesturing has been examined in relation to prosodic structure. An early study by Kendon [102] shows that there is a correspondence between hand, arm and head movements and the prosodic and the discourse hierarchical organization (see also [113] for a relationship between discourse structure and gesturing). It has also been shown that manual gesturing and head movement is timed with prominent syllables [16,17,114–117]). Interestingly, co-speech gestures at pitch accented words are affected by prosodic boundaries in a manner parallel to the way pitch accents are affected by boundaries: in a study of Catalan, Esteve-Gibert & Prieto [118] found that, like intonation peaks in pitch accents, gesturing targets were produced earlier (retracted) when a prosodic boundary followed the prominent word. Furthermore, and unlike pitch accents, this retraction was also affected by a preceding prosodic boundary.

The above studies provide strong evidence for a link between prosodic structure and gesturing, but a number of questions remain. For example, it is not clear which part of the speech signal and which part of the co-speech gesture is targeted in alignment, although for the gesturing the apex seems to be emerging as the most relevant point (see [118,119]; the apex is the peak of a gesturing movement, e.g. the furthest point a finger reaches in a pointing movement). It is also unclear how systematic the co-occurrence of gesturing with prominence is in conversational speech (see also [116]). Thus, Ferré [120] in a study of conversational French found that only 171 of 1289 hand movements and 104 of 2520 head movements aligned with a prosodically prominent element. Similarly, for Dutch, Swerts & Kraemer [116] found that out of 228 head nods only 60 (26.3%) nods co-occurred with a strongly accented (i.e. highly prominent) word, whereas the remaining 168 nods occurred on words with a weak accent (124 nods) or no accent (44 nods). It is worth pointing out, however, that the most prominent words co-occurred with a head nod in 89.6% of the cases (60 of 67 words), whereas words without an accent co-occurred with a head nod in only 7.2% of the cases (44 of 609 words). For English, on the other hand, Loehr [115,121] found that most hand and head apexes occurred within approximately one word of the prominent word. Similarly,

Mendoza-Denton & Jannedy [117] found a tight integration of gesturing and prominence in that almost all gesture apices (in hand, arm and torso movements) occurred together with pitch accented words. It is worth keeping in mind that co-occurrences are difficult to interpret, because prosodically prominent elements might have other co-speech gestures associated with them (torso movements, eyebrow movements; on eyebrow movements see [116]) and on the other hand, gesturing might be marking prominence other than prosodic prominence (syntactic prominence for example, see discussion in [120]). However, the co-occurrence question will need to be addressed for a full understanding of the relationship between prosodic structure and gesturing.

Finally, work on finger tapping at prominent syllables has shown that the movement of both fingers and oral gestures lengthen under prominence [122,123]. In Dutch, Krahmer & Swerts [124] found that words exhibit acoustic properties of prominence (e.g. lengthening) when occurring with various types of gesturing (e.g. head nods and manual gestures). This work indicates that the influence of prosodic structure extends beyond vocal tract gestures and includes body movements.

(d) Gesturing at prosodic boundaries

The relationship between prosodic boundaries and gesturing has received little attention to date. A study by Barkhuysen *et al.* [125] found that Dutch speakers were better at distinguishing phrase-medial words from words at prosodic phrase boundaries when they could both see and hear speakers (rather than just seeing or just hearing the speakers). Barkhuysen *et al.* [125] further conducted an analysis of half of the visual stimuli they used in their study and found that facial gesturing differed between phrase-final and phrase-medial words (e.g. there were differences in the direction of the gaze and the head, in the position of the eyebrows and in the amount of blinking and nodding). Similarly in Japanese, Ishi *et al.* [126] found that head nods occurred frequently towards ends of strong prosodic phrases. Thus, there is evidence of boundary-related gesturing. There is also evidence of alignment of gesturing and boundaries. A study of one speaker by Shattuck-Hufnagel *et al.* [127] has found indication of alignment of intonation phrases and torso leans. Loehr [115,121] examined alignment of gesturing with intermediate phrases and stated that in the majority of the data, gesturing aligned closely with intermediate phrases. These studies lend strong support to the idea that body movements are in some way coordinated with prosodic phrasing. However, research in this area is only beginning. In the remainder of this section, we explore the question of whether properties of body movements at prosodic boundaries can inform us about the nature of linguistic boundaries. Specifically, the goal is to evaluate whether the questions raised in §4a about the distinction between grammatical and ungrammatical boundaries, and about different cognitive processes at boundaries, can be further illuminated by examining body movement. To begin addressing this issue, we present a small study of gesturing during pauses. The goal of this study, which uses data from Myers [128], is to examine whether body movements display lengthening at boundaries.

Myers [128] examined gesturing at pauses. She analysed short video clips of six speakers engaging in a spontaneous monologue (the speakers were answering a question). One

of the six speakers was excluded, however, because she had only four gestures. For different speakers, different movements were analysed, depending on where an initial screening of the participants revealed the participants to gesture most. For two subjects, manual gesturing was analysed, and for three subjects, head movement was analysed. The movements were analysed irrespective of a possible contribution to meaning. The only movements that were not analysed were those that were clearly not related to speech (e.g. head scratching). Movements were analysed in terms of 'onset' and 'target' for each distinct motion. The onset was defined as the movement of the head or the hand as it leaves one position or changes direction from a previous movement, and the target was defined as the point when the hand or head stops moving or changes direction (in that case it would be also the onset of the next movement).⁶ In total, 82 pauses were analysed (19 pauses for speaker 1, 16 for speaker 2, 19 for speaker 3, 15 for speaker 4 and 13 for speaker 5). Myers divides movement at pauses as follows: *bridging in* movement, which is movement that starts during speech and ends in a pause, *bridging out* movement, which is movement that starts during a pause and ends in the following speech utterance, *bridging* movement (referred to as pre-to-post movement in [128]), which is movement that starts in the speech preceding the pause and ends in the speech following the pause, and *in pause* movement, which is movement that starts and ends in a pause. Myers finds that most of the movement is of the bridging out type, that is, starting in the pause and ending during the following speech part. She suggests that this might indicate that speakers use movement to help initiate speech, thereby functioning 'as a cognitive bridge between pauses and fluent speech' [128]. For additional analyses not reported here, she also classified the pauses as fluent—indicating a prosodic boundary—or disfluent, based on the perception of two listeners (Myers and one naive listener). The labellers agreed in 78.3% of the cases. In the case of disagreement, we have used Myers' scoring, as, based on a check of a sample of the data, her judgement seemed more accurate. Julia Myers has kindly provided her data to us for further analysis.

Dividing the movement data as they occurred between fluent and disfluent pauses shows that in disfluent pauses the most common gesturing type was again bridging out (26 gestures), followed by bridging in (18 gestures) and in pause (15 gestures), and the least common was bridging (five gestures). In fluent pauses, the most common type of gesturing was also bridging out (18), followed by bridging (14) and bridging in (13) and then in pause gesturing (6). The hand and head movements were further analysed for lengthening. If body movements behave like articulatory constriction gestures, in the sense that they are under the control of the π -gesture, then they should exhibit lengthening at boundaries, and the lengthening should increase as boundary strength increases. Thus, a positive correlation is expected between co-speech gesturing duration and pause duration: because more lengthening is expected at stronger boundaries, with an increase in boundary strength (evidenced by an increase in pause duration) co-speech gesturing should also increase in duration. In the case of bridging gestures, this correlation would be trivial, because they encompass the pause. Note, however, that the prediction is that lengthening will occur only at *fluent* pauses, because only a fluent pause can be considered to be a part of a prosodic boundary, whereas a

disfluent pause cannot. While this is clearly too small a dataset to draw firm conclusions, we conducted a preliminary analysis on the relationship of these manual/head gesturing types with prosodic boundaries (except for bridging disfluent and in pause fluent gesturing, where there were fewer than 10 data points, five for bridging disfluent and six for in pause fluent).

To pool the speakers, pause durations were *z*-scored for each subject separately, as were gesturing durations. A regression analysis was conducted on each type of gesturing, examining whether pause duration (as an indicator of prosodic boundary strength) and gesturing duration are correlated. There is a nearly significant positive correlation only for the bridging in gesturing in fluent pauses ($r^2 = 0.242$, $p = 0.0504$). No other correlations were significant or even approached significance. Thus, there is suggestive evidence of lengthening of one kind of gesturing, namely the bridging in gesturing in fluent pauses. This suggests that, like oral constriction gestures, the movements at the end of prosodic phrases at large boundaries (which have fluent pauses) show lengthening. This in turn indicates—within the limitations that this is a very small set of data—that body movements are under the control of the π -gesture. The lack of a correlation with the bridging gesturing is surprising, given that this type of gesturing encompasses the pause. However, the pauses where this type of gesturing occurred were particularly short (−1.35 to 0.2 standard deviations), whereas the gesturing was not (0.35 to 2.5 standard deviations). This indicates that the gesturing was probably too long, and the boundaries too weak, for the lengthening effect of boundaries to become evident, even if lengthening of the manual/head movement took place. The lack of a correlation between bridging out gesturing and boundary strength suggests that the bridging out gesture is not controlled by prosodic structure. One explanation could be that bridging out gesturing is related to planning. It has been suggested that gesturing facilitates lexical retrieval [107,108,129] or is involved in the conceptual planning of an utterance [110]. Both events can be understood as non-structural and therefore might not be under control of the π -gesture.

One question that arises for future research is whether pause duration and gesturing could reflect different *types* of planning. For example, pause duration might reflect the planning process of the rough representation of the whole utterance, together with the full planning of an initial part of the utterance (as would be expected in Keating and Shattuck-Hufnagel's model of speech planning [93]). This could be considered linguistic planning. Hand and head movement could reflect conceptual planning or lexical retrieval, which are not structural in nature (see [85] on the related idea that pauses and F0 reflects different planning levels).

Work discussed in this section provides clear evidence that gesturing is related to linguistic structure but that this relationship needs to be further examined. Combining co-speech gesturing research with the work of Ramanarayanan *et al.* [95,96,99], which shows that articulation during pauses differs depending on the type of pause, we arrive at the following questions: (i) at which types of pauses does co-speech gesturing occur (e.g. is there a difference in co-speech gesturing at prosodic boundaries in comparison with disfluent pauses)? (ii) what kind of co-speech gesturing occurs with which kind of pauses (e.g. is there a difference in the occurrence of semantically emptier in comparison with semantically more

specified co-speech gesturing at different kinds of pauses; is the use of different body parts in co-speech gesturing related to a specific type of pause)? (iii) how is co-speech gesturing coordinated with oral constriction gestures and prosodic gestures (i.e. where in the boundary or disfluent pause do co-speech gestures occur and with which specific linguistic gestures are they coordinated)? (iv) can we gain information about speech planning from the answers to the first three questions, in the sense that different types of planning (e.g. conceptual or linguistic) might be reflected in different types of co-speech gesturing or different coordinations? and (v) how are boundaries, prominence and gesturing integrated? The answers to these questions will inform us about what constitutes a prosodic boundary, how different linguistic gestures and co-speech gesturing combine to form a prosodic boundary, and what the role of co-speech gesturing is in speech planning. This information can in turn contribute significantly to understanding the distinction between prosodic boundaries and planning-driven behaviour.

5. Conclusion

Research on prosodic structure has made significant advances over the past decades and new directions of research are continually emerging. From the large number of fascinating findings and exciting research questions, we have selected a small set of questions that are centred on the coordination of events at prosodic boundaries. The emerging picture suggests that prosodic boundaries can be understood as arising from a set of coordinated prosodic gestures, namely π -gestures, boundary tone gestures, and μ -gestures [7,13,14,61]. Studies examining the relationship between prosodic structure and body movement provide evidence of their tight integration, and there is indication that some body movement is under the control of prosodic structure (e.g. [118,119,125] and the small analysis presented here). We suggest that further examination of the vocal tract gestures and body movements at prosodic boundaries promises to yield a fuller understanding of processes at prosodic boundaries and a distinction between structurally controlled and planning-driven behaviour. On the whole, research on gestural coordination at prosodic boundaries allows us to refine our understanding of prosodic structure and how it is reflected in and shaped by the nature of speech production processes.

Data accessibility. The conclusions in this paper are based on video data of speakers. As these videos contain identifying features, we cannot provide these videos as additional data. For more information, please contact the corresponding author.

Acknowledgements. We are very grateful to Julia Myers, Dani Byrd, Mark Tiede, Mandana Seyfeddinipur and Stefanie Shattuck-Hufnagel for discussions about various aspects of prosody and of gesturing and to Dani Byrd, Rajka Smiljanić, George D. Allen, two anonymous reviewers and Rachel Smith for their helpful comments on earlier versions of this manuscript.

Funding statement. Research supported by NIH grant DC002717 to Doug Whalen and NIH grant P01 HD 001994 to Jay Rueckl.

Endnotes

¹The role of prominence is debated in the literature. In one view, it is only the last prominent element (the nuclear pitch accent) that carries semantic meaning, whereas preceding (i.e., pre-nuclear) pitch accents contribute very little to the meaning of an utterance (e.g.

[9,10]). The other view argues that pre-nuclear accents do contribute to the meaning of an utterance [11,12].

²Note that languages differ in how prosodic structure fulfills this function. For example, in edge-prominence languages, prominence is marked by means of phrasing (see [6,9,22]).

³Note that, unless otherwise indicated, the work cited in this article refers to properties of English.

⁴Under this assumption, the occurrence and strength of a prosodic boundary would be determined by a number of factors (such as syntactic structure, phrase length and rhythm) and the exact onset of the π -gesture would be determined by the position of the lexically stressed syllable (see [30], p. 175 for further discussion).

⁵It is unclear at this point whether a separate coordination of the boundary tone gesture with the phrase-final vowel gesture is needed to account for the temporal relationship between the boundary tone and the phrase-final vowel (see [14] for discussion).

⁶Gesturing movements are often divided into further, semantically defined parts. Thus for example [130], following largely the division by [102], divides a co-speech gesture into a preparation, pre-stroke hold, stroke, post-stroke hold and retraction part, where the stroke is the only obligatory part of the co-speech gesture. The stroke carries whatever meaning the co-speech gesture has, the preparation brings the gesturing body part (e.g., the hand) to the starting point of the stroke, the retraction brings the hand back to its rest position and the two holds are the parts of the co-speech gesture when the hand is held in the stroke onset/end position [130]. Myers [128] did not analyse gesturing in this way as her study was primarily concerned with the occurrence of *any* body movements at prosodic boundaries and disfluencies in comparison with fluent speech. To this end, she analysed gesturing as onset-to-target movements which allowed her to capture all the movements a speaker made, regardless of their semantic interpretation.

References

- Shattuck-Hufnagel S, Turk A. 1996 A prosody tutorial for investigators of auditory sentence processing. *J. Psycholinguist. Res.* **25**, 193–247. (doi:10.1007/BF01708572)
- Wagner M, Watson D. 2010 Experimental and theoretical advances in prosody: a review. *Lang. Cogn. Process.* **25**, 905–945. (doi:10.1080/01690961003589492)
- Fletcher J. 2010 The prosody of speech: timing and rhythm. In *The handbook of phonetic sciences* (eds W Hardcastle, WJ Laver, F Gibbon), pp. 523–602. London, UK: Wiley-Blackwell.
- Pierrehumbert J. 1999 Prosody and intonation. In *The MIT encyclopedia of cognitive sciences* (eds RA Wilson, F Keil), pp. 479–482. Cambridge, MA: MIT Press.
- Ladd DR. 2001 Intonation. In *Language typology and language universals: an international handbook*, vol. 2 (eds M Haspelmath, E König, W Oesterreicher, W Raible), pp. 1380–1390. Berlin, Germany: Mouton de Gruyter.
- Jun S-A. 2005 Prosodic typology. In *Prosodic typology: the phonology of intonation and phrasing* (ed. S-A Jun), pp. 430–458. Oxford, UK: Oxford University Press.
- Byrd D, Saltzman E. 2003 The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *J. Phon.* **31**, 149–180. (doi:10.1016/S0095-4470(02)00085-2)
- Beckman ME. 1986 *Stress and non-stress accent*. Dordrecht, The Netherlands: Foris.
- Ladd DR. 1996/2008 *Intonational phonology*. Cambridge, UK: Cambridge University Press.
- Calhoun S. 2010 The centrality of metrical structure in signaling information structure: a probabilistic perspective. *Language* **86**, 1–42. (doi:10.1353/lan.0.0197)
- Bishop J. 2013 Prenuclear accentuation in English: phonetics, phonology, information structure. PhD thesis, University of California, Los Angeles, CA, USA.
- Petrone C, Niebuhr O. 2014 On the intonation in German intonation questions: the role of the prenuclear region. *Lang. Speech* **57**, 108–146. (doi:10.1177/0023830913495651)
- Gao M. 2008 Mandarin tones: an articulatory phonology account. PhD thesis, Yale University, New Haven, CT, USA.
- Katsika A. 2012 Coordination of prosodic gestures at boundaries in Greek. PhD thesis, Yale University, New Haven, CT, USA.
- Byrd D, Riggs D. 2008 Locality interactions with prominence in determining the scope of phrasal lengthening. *J. Int. Phon. Assoc.* **38**, 187–202. (doi:10.1017/S0025100308003460)
- Yasinnik Y, Renwick M, Shattuck-Hufnagel S. 2004 The timing of speech-accompanying gestures with respect to prosody. In *Proc. From Sound to Sense*, pp. 97–102. See <http://www.rle.mit.edu/soundtosense/conference/starthere.htm>.
- Rochet-Capellan A, Laboissière R, Galván A, Schwartz JL. 2008 The speech focus position effect on jaw-finger coordination in a pointing task. *J. Speech, Lang. Hear. Res.* **51**, 1507–1521. (doi:10.1044/1092-4388(2008/07-0173)
- Selkirk E. 1984 *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.
- Nespor M, Vogel I. 1986 *Prosodic phonology*. Dordrecht, The Netherlands: Foris Publications.
- Beckman ME, Pierrehumbert JB. 1986 Intonational structure in Japanese and English. *Phonol. Yearbook* **3**, 255–309. (doi:10.1017/S095267570000066X)
- Pierrehumbert JB. 1980 The phonology and phonetics of English intonation. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Jun S-A. 1993 The phonetics and phonology of Korean prosody. PhD thesis, The Ohio State University, Columbus, OH, USA.
- Jun S-A. 1998 The accentual phrase in the Korean prosodic hierarchy. *Phonology* **15**, 189–226. (doi:10.1017/S0952675798003571)
- Goldman Eisler F. 1968 *Psycholinguistics. Experiments in spontaneous speech*. London, UK: Academic Press.
- Fletcher J. 1987 Some micro and macro effects of tempo change on timing in French. *Linguistics* **25**, 951–967. (doi:10.1515/ling.1987.25.5.951)
- Trouvain J, Grice M. 1999 The effect of tempo on prosodic structure. In *Proc. of the XIVth Int. Congress of Phonetic Sciences. San Francisco, CA, 1–7 August 1999*, vol. 2, pp. 1067–1070. Martinez: East Bay Institute for Research and Education Inc.
- Grosjean F, Grosjean L, Lane H. 1979 The patterns of silence: performance structures in sentence production. *Cogn. Psychol.* **11**, 58–81. (doi:10.1016/0010-0285(79)90004-5)
- Zvonik E, Cummins F. 2003 The effect of surrounding phrase lengths on pause duration. In *Proc. 8th Europ. Conf. Speech Communication and Technology, 1–4 September 2003, Geneva, Switzerland*, pp. 777–780. International Speech Communication Association.
- Krivokapić J. 2007 Prosodic planning: effects of phrasal length and complexity on pause duration. *J. Phon.* **35**, 162–179. (doi:10.1016/j.wocn.2006.04.001)
- Krivokapić J. 2007 The planning, production, and perception of prosodic structure. PhD thesis, University of Southern California, Los Angeles, CA, USA.
- Carlson K, Clifton Jr C, Frazier L. 2001 Prosodic boundaries in adjunct attachment. *J. Mem. Lang.* **45**, 58–81. (doi:10.1006/jmla.2000.2762)
- Clifton Jr C, Carlson K, Frazier L. 2002 Informative prosodic boundaries. *Lang. Speech* **45**, 87–114. (doi:10.1177/00238309020450020101)
- Jun S-A. 2003 Prosodic phrasing and attachment preferences. *J. Psycholinguist. Res.* **32**, 219–249. (doi:10.1023/A:1022452408944)
- Frazier L, Clifton Jr C, Carlson K. 2004 Don't break, or do: prosodic boundary preferences. *Lingua* **114**, 3–27. (doi:10.1016/S0024-3841(03)00044-5)
- Krivokapić J, Byrd D. 2012 Prosodic boundary strength: an articulatory and perceptual study. *J. Phon.* **40**, 430–442. (doi:10.1016/j.wocn.2012.02.011)
- Wagner M. 2005 Prosody and recursion. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Oller DK. 1973 The effect of position in utterance on speech segment duration in English. *J. Acoust. Soc. Am.* **54**, 1235–1247. (doi:10.1121/1.1914393)

38. Wightman CW, Shattuck-Hufnagel S, Ostendorf M, Price JP. 1992 Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* **91**, 1707–1717. (doi:10.1121/1.402450)
39. Shattuck-Hufnagel S, Turk A. 1998 The domain of phrase-final lengthening in English. In *The sound of the future: a global view of acoustics in the 21st Century, Proc. 16th Int. Congress on Acoustics and 135th Meeting Acoustical Society of America*, pp. 1235–1236.
40. Tabain M. 2003 Effects of prosodic boundary on /aC/ sequences: acoustic results. *J. Acoust. Soc. Am.* **113**, 516–531. (doi:10.1121/1.1523390)
41. Fougeron C, Keating P. 1997 Articulatory strengthening at edges of prosodic domains. *J. Acoust. Soc. Am.* **101**, 3728–3740. (doi:10.1121/1.418332)
42. Byrd D, Saltzman E. 1998 Intragestural dynamics of multiple phrasal boundaries. *J. Phon.* **26**, 173–199. (doi:10.1006/jpho.1998.0071)
43. Byrd D. 2000 Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* **57**, 3–16. (doi:10.1159/000028456)
44. Cho T. 2005 Prosodic strengthening and featural enhancement: evidence from acoustic and articulatory realizations of /a,i/ in English. *J. Acoust. Soc. Am.* **117**, 3867–3878. (doi:10.1121/1.1861893)
45. Tabain M. 2003 Effects of prosodic boundary on /aC/ sequences: articulatory results. *J. Acoust. Soc. Am.* **113**, 2834–2849. (doi:10.1121/1.1564013)
46. Bombien L, Mooshammer C, Hoole P. 2013 Articulatory coordination in word-initial clusters of German. *J. Phon.* **41**, 546–561. (doi:10.1016/j.wocn.2013.07.006)
47. Cho T, Keating P. 2009 Effects of initial position versus prominence in English. *J. Phon.* **37**, 466–485. (doi:10.1016/j.wocn.2009.08.001)
48. Fougeron C. 2001 Articulatory properties of initial segments in several prosodic constituents in French. *J. Phon.* **29**, 109–135. (doi:10.1006/jpho.2000.0114)
49. Keating P, Cho T, Fougeron C, Hsu C. 2004 Domain-initial articulatory strengthening in four languages. In *Phonetic interpretation (papers in laboratory phonology VI)* (eds J Local, R Ogden, R Temple), pp. 143–161. Cambridge, UK: Cambridge University Press.
50. Cho T, Keating P. 2001 Articulatory and acoustic studies on domain-initial strengthening in Korean. *J. Phon.* **29**, 155–190. (doi:10.1006/jpho.2001.0131)
51. Ladd DR. 1986 Intonational phrasing: the case of recursive prosodic structure. *Phonol. Yearbook* **3**, 311–340.
52. Byrd D. 2006 Relating prosody and dynamic events: commentary on the papers by Cho, Navas, and Smiljanić. In *Laboratory phonology 8: varieties of phonological competence* (ed. L Goldstein), pp. 549–561. Berlin, Germany: Mouton De Gruyter.
53. Browman CP, Goldstein LM. 1992 Articulatory phonology: an overview. *Phonetica* **49**, 155–180. (doi:10.1159/000261913)
54. Browman CP, Goldstein L. 1995 Dynamics and articulatory phonology. In *Mind as motion: explorations in the dynamics of cognition* (eds RF Port, T Van Gelder), pp. 175–193. Cambridge, MA: The MIT Press.
55. Goldstein L, Byrd D, Saltzman E. 2006 The role of vocal tract gestural action units in understanding the evolution of phonology. In *Action to language via the mirror neuron system* (ed. M Arbib), pp. 215–249. New York, NY: Cambridge University Press.
56. Byrd D, Krivokapić J, Lee S. 2006 How far, how long: on the temporal scope of phrase boundary effects. *J. Acoust. Soc. Am.* **120**, 1589–1599. (doi:10.1121/1.2217135)
57. Berkovits R. 1993 Progressive utterance-final lengthening in syllables with final fricatives. *Lang. Speech* **36**, 89–98.
58. Berkovits R. 1993 Utterance-final lengthening and the duration of final-stop closures. *J. Phon.* **21**, 479–489.
59. Berkovits R. 1994 Durational effects in final lengthening, gapping, and contrastive stress. *Lang. Speech* **37**, 237–250.
60. Cambier-Langeveld T. 1997 The domain of final lengthening in the production of Dutch. In *Linguistics in the Netherlands 1997*. (eds J Coerts, H de Hoop), pp. 13–24. Amsterdam, The Netherlands: John Benjamins.
61. Saltzman E, Nam H, Krivokapić J, Goldstein L. 2008 A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proc. Speech Prosody 2008 Conf.* pp. 175–184.
62. Saltzman E. 1986 Task dynamic co-ordination of the speech articulators: a preliminary model. In *Generation and modulation of action patterns* (eds H Heuer, C Fromm), pp. 129–144. Berlin, Germany: Springer.
63. Saltzman E, Kelso JAS. 1987 Skilled actions: a task dynamic approach. *Psychol. Rev.* **94**, 84–106. (doi:10.1037/0033-295X.94.1.84)
64. Saltzman E, Munhall K. 1989 A dynamical approach to gestural patterning in speech production. *Ecol. Psychol.* **1**, 333–382. (doi:10.1207/s15326969eco0104_2)
65. Mücke D, Nam H, Hermes A, Goldstein L. 2012 Coupling of tone and constriction gestures in pitch accents. In *Consonant clusters and structural complexity* (eds P Hoole, M Pouplier, L Bombien, Ch Mooshammer, B Kühnert), pp. 205–230. Berlin, Germany: Mouton de Gruyter.
66. Katsika A, Krivokapić J, Mooshammer C, Tiede M, Goldstein L. 2014 The coordination of boundary tones and their interaction with prominence. *J. Phon.* **44**, 62–82. (doi:10.1016/j.wocn.2014.03.003)
67. Saltzman E, Byrd D. 2000 Task-dynamics of gestural timing: phase windows and multifrequency rhythms. *Hum. Mov. Sci.* **19**, 499–526. (doi:10.1016/S0167-9457(00)00030-0)
68. Nam H, Saltzman E. 2003 A competitive, coupled oscillator model of syllable structure. In *Proc. XIIIth Int. Cong. Phonetic Sciences*, Barcelona, Spain, pp. 2253–2256.
69. Nam H. 2007 Syllable-level intergestural timing model: split-gesture dynamics focusing on positional asymmetry and mora structure. In *Laboratory phonology 9* (eds J Cole, JI Hualde), pp. 483–506. Berlin, Germany: Walter de Gruyter.
70. Jun S-A. 2014 Prosodic typology: by prominence type, word prosody, and macro-rhythm. In *Prosodic typology II: the phonology of intonation and phrasing* (ed. S-A Jun), pp. 520–539. Oxford, UK: Oxford University Press.
71. Cho T, Kim J, Kim S. 2013 Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in English. *J. Acoust. Soc. Am.* **133**, EL384–EL390. (doi:10.1121/1.4800179)
72. Turk AE, Shattuck-Hufnagel S. 2007 Multiple targets of phrase-final lengthening in American English words. *J. Phon.* **35**, 445–472. (doi:10.1016/j.wocn.2006.12.001)
73. Rusaw E. 2013 Modeling temporal coordination in speech pattern generator neural network. PhD thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA.
74. White L. 2002 English speech timing: a domain and locus approach. PhD thesis, The University of Edinburgh, Edinburgh, UK.
75. Lehiste I, Olive JP, Streeter LA. 1976 Role of duration in disambiguating syntactically ambiguous sentences. *J. Acoust. Soc. Am.* **60**, 1199–1202. (doi:10.1121/1.381180)
76. Price PJ, Ostendorf M, Shattuck-Hufnagel S, Fong C. 1991 The use of prosody in syntactic disambiguation. *J. Acoust. Soc. Am.* **90**, 2956–2970. (doi:10.1121/1.401770)
77. Speer SR, Kjelgaard MM, Dobroth KM. 1996 The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *J. Psycholinguist. Res.* **25**, 247–268. (doi:10.1007/BF01708573)
78. Schafer AJ, Speer SR, Warren P, White SD. 2000 Intonational disambiguation in sentence production and comprehension. *J. Psycholinguist. Res.* **29**, 169–182. (doi:10.1023/A:1005192911512)
79. Snedeker J, Trueswell J. 2003 Using prosody to avoid ambiguity: effects of speaker awareness and referential context. *J. Mem. Lang.* **48**, 103–130. (doi:10.1016/S0749-596X(02)00519-3)
80. Kraljic T, Brennan SE. 2005 Prosodic disambiguation of syntactic structure: for the speaker or for the addressee? *Cogn. Psychol.* **50**, 194–231. (doi:10.1016/j.cogpsych.2004.08.002)
81. Carlson K. 2009 How prosody influences sentence comprehension. *Lang. Linguist. Compass* **3**, 1188–1200. (doi:10.1111/j.1749-818X.2009.00150.x)
82. Ferreira F. 1991 Effects of length and syntactic complexity on initiation times for prepared utterances. *J. Mem. Lang.* **30**, 210–233. (doi:10.1016/0749-596X(91)90004-4)
83. Krivokapić J. 2012 Prosodic planning in speech production. In *Speech planning and dynamics* (eds S Fuchs, M Wehrich, D Pape, P Perrier), pp. 157–190. München, Germany: Peter Lang.
84. Wheeldon L, Lahiri A. 1997 Prosodic units in speech production. *J. Mem. Lang.* **37**, 356–381. (doi:10.1006/jmla.1997.2517)
85. Fuchs S, Petrone C, Krivokapić J, Hoole P. 2013 Acoustic and respiratory evidence for utterance

- planning in German. *J. Phon.* **41**, 29–47. (doi:10.1016/j.wocn.2012.08.007)
86. Cooper WE, Paccia-Cooper J. 1980 *Syntax and speech*. Cambridge, MA: Harvard University Press.
 87. Ferreira F. 1993 Creation of prosody during sentence production. *Psychol. Rev.* **100**, 233–253. (doi:10.1037/0033-295X.100.2.233)
 88. Ferreira F. 2007 Prosody and performance in language production. *Lang. Cogn. Process.* **22**, 1151–1177. (doi:10.1080/01690960701461293)
 89. Strangert E. 1997 Relating prosody to syntax: boundary signaling in Swedish. In *Proc. the 5th European Conf. on Speech Communication and Technology, Rhodes, Greece*, vol. 1 (eds G Kokkinakis, N Fakotakis, E Dermatas), pp. 239–242.
 90. Levelt WJM. 1989 *Speaking*. Cambridge, MA: MIT Press.
 91. Ferreira F. 1988 Planning and timing in sentence production: the syntax-to-phonology conversation. PhD thesis, University of Massachusetts, Amhurst, MA, USA.
 92. Ferreira F, Swets B. 2002 How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. Lang.* **46**, 57–84. (doi:10.1006/jmla.2001.2797)
 93. Keating P, Shattuck-Hufnagel S. 2002 A prosodic view of word form encoding for speech production. *UCLA Working Papers Phon.* **101**, 112–156.
 94. Narayanan S, Nayak K, Lee S, Sethy A, Byrd D. 2004 An approach to real-time magnetic resonance imaging for speech production. *J. Acoust. Soc. Am.* **115**, 1771–1776. (doi:10.1121/1.1652588)
 95. Ramanarayanan V, Byrd D, Goldstein L, Narayanan S. 2010 Investigating articulatory setting—pauses, ready position and rest - using real-time MRI. In *Proc. Interspeech 2010, Makuhari, Japan, September 2010*.
 96. Ramanarayanan V, Goldstein L, Byrd D, Narayanan S. 2013 A real-time MRI investigation of articulatory setting across different speaking styles. *J. Acoust. Soc. Am.* **134**, 510–519. (doi:10.1121/1.4807639)
 97. Perkell J. 1969 Physiology of speech production: results and implications of quantitative cineradiographic study. Research monograph no. 53. Cambridge, MA: MIT Press.
 98. Gick B, Wilson I, Koch K, Cook C. 2004 Language-specific articulatory settings: evidence from inter-utterance rest position. *Phonetica* **61**, 220–233. (doi:10.1159/000084159)
 99. Ramanarayanan V, Bresch E, Byrd D, Goldstein L, Narayanan S. 2009 Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation. *J. Acoust. Soc. Am. Express. Lett.* **126**, EL160–EL165.
 100. Kendon A. 1997 Gesture. *Ann. Rev. Anthropol.* **26**, 109–128. (doi:10.1146/annurev.anthro.26.1.109)
 101. Kendon A. 2004 *Gesture: visible action as utterance*. Cambridge, UK: Cambridge University Press.
 102. Kendon A. 1972 Some relationships between body motion and speech. In *Studies in dyadic communication* (eds E Seigman, B Pope), pp. 177–210. Oxford, UK: Pergamon Press.
 103. McNeill D. 1985 So you think gestures are non-verbal? *Psychol. Rev.* **92**, 350–371. (doi:10.1037/0033-295X.92.3.350)
 104. McNeill D. 2005 *Gesture and thought*. Chicago, IL: University of Chicago Press.
 105. Özyürek A, Willems RM, Kita S, Hagoort P. 2007 On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cogn. Neurosci.* **19**, 605–616. (doi:10.1162/jocn.2007.19.4.605)
 106. Willems RM, Özyürek A, Hagoort P. 2007 When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* **17**, 2322–2333. (doi:10.1093/cercor/bhl141)
 107. Butterworth B, Hadar U. 1989 Gesture, speech and computational stages: a reply to McNeill. *Psychol. Rev.* **96**, 168–174. (doi:10.1037/0033-295X.96.1.168)
 108. Rauscher FH, Krauss RM, Chen Y. 1996 Gesture, speech, and lexical access: the role of lexical movements in speech production. *Psychol. Sci.* **7**, 226–231. (doi:10.1111/j.1467-9280.1996.tb00364.x)
 109. Morsella E, Krauss RM. 2004 The role of gestures in spatial working memory and speech. *Am. J. Psychol.* **117**, 411–424. (doi:10.2307/4149008)
 110. Alibali MW, Kita S, Young AJ. 2000 Gesture and the process of speech production: we think, therefore we gesture. *Lang. Cogn. Process.* **15**, 593–613. (doi:10.1080/016909600750040571)
 111. Seyfeddinipur M. 2006 Disfluency: interrupting speech and gesture. PhD thesis, Radboud University, Nijmegen, The Netherlands.
 112. Tiede M, Goldstein L, Mooshammer C, Nam H, Saltzman S, Shattuck-Hufnagel S. 2011 Head movement is correlated with increased difficulty in an accelerating speech production task. In *Proc. ninth Int. Seminar on Speech Production, Montreal, Canada, June 2011*.
 113. McNeill D, Quek F, McCullough K-E, Duncan S, Furuyama N, Bryll R, Ma X-F, Ansari R. 2001 Catchments, prosody, and discourse. *Gesture* **1**, 9–33. (doi:10.1075/gest.1.1.03mcn)
 114. De Ruiter JPA. 1998 Gesture and speech production. PhD thesis, Radboud University, Nijmegen, The Netherlands.
 115. Loeher D. 2004 Gesture and intonation. PhD thesis, Georgetown University, Washington DC, USA.
 116. Swerts M, Krahmer E. 2010 Visual prosody of newsreaders: effects of information structure, emotional content and intended audience on facial expressions. *J. Phon.* **38**, 197–206. (doi:10.1016/j.wocn.2009.10.002)
 117. Mendoza-Denton N, Jannedy S. 2011 Semiotic layering through gesture and intonation: a case study of complementary and supplementary multimodality in political speech. *J. Eng. Linguist.* **39**, 265–299. (doi:10.1177/0075424211405941)
 118. Esteve-Gibert N, Prieto P. 2013 Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *J. Speech, Lang. Hear. Res.* **56**, 850–864. (doi:10.1044/1092-4388(2012/12-0049))
 119. Leonard T, Cummins F. 2011 The temporal relation between beat gestures and speech. *Lang. Cogn. Process.* **26**, 1457–1471. (doi:10.1080/01690965.2010.500218)
 120. Ferré G. 2014 A multimodal approach to markedness in spoken French. *Speech Commun.* **57**, 268–282. (doi:10.1016/j.specom.2013.06.002)
 121. Loehr D. 2012 Temporal, structural, and pragmatic synchrony between intonation and gesture. *J. Lab. Phonol.* **3**, 71–89.
 122. Parrell B, Goldstein L, Lee S, Byrd D. 2011 Temporal coupling between speech and manual motor actions. In *Proc. ninth Int. Seminar on Speech Production, Montreal, Canada, June 2011*.
 123. Parrell B, Goldstein L, Lee S, Byrd D. 2014 Spatiotemporal coupling between speech and manual motor actions. *J. Phon.* **41**, 1–11. (doi:10.1016/j.wocn.2013.11.002)
 124. Krahmer E, Swerts M. 2007 The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* **57**, 396–414. (doi:10.1016/j.jml.2007.06.005)
 125. Barkhuysen P, Krahmer E, Swerts M. 2008 The interplay between the auditory and visual modality for end-of-utterance detection. *J. Acoust. Soc. Am.* **123**, 354–365. (doi:10.1121/1.2816561)
 126. Ishi CT, Ishiguro H, Hagita N. 2014 Analysis of relationship between head motion events and speech in dialogure conversations. *Speech Comm.* **57**, 233–243.
 127. Shattuck-Hufnagel S, Ren P, Tauscher E. 2010 Are torso movements during speech timed with intonational phrases? In *Proc. Speech Prosody 2010, Chicago, IL*. See <http://www.speechprosody2010.illinois.edu/papers/100974.pdf>.
 128. Myers J. 2012 The behavior of hand and facial gestures at pauses in speech. BA thesis, Yale University, New Haven, CT, USA.
 129. Krauss RM. 1998 Why do we gesture when we speak? *Curr. Dir. Psychol. Sci.* **7**, 54–60. (doi:10.1111/1467-8721.ep13175642)
 130. Kita S. 1993 Language and thought interface: a study of spontaneous gestures and Japanese mimetics. PhD thesis, University of Chicago, Chicago, IL, USA.