Communicative function and prosodic form in speech timing

Laurence White

Plymouth University

Corresponding author:

Laurence White

School of Psychology

Plymouth University

Plymouth PL4 8AA

UK

Email: laurence.white@plymouth.ac.uk

Phone: +44(0)1752584876

Fax: +44(0)1752233176

Abstract

Listeners can use variation in speech segment duration to interpret the structure of spoken utterances, but there is no systematic description of how speakers manipulate timing for communicative ends. Here I propose a functional approach to prosodic speech timing, with particular reference to English. The disparate findings regarding the production of timing effects are evaluated against the functional requirement that communicative durational variation should be perceivable and interpretable by the listener. In the resulting framework, prosodic structure is held to influence speech timing directly only at the heads and edges of prosodic domains, through large, consistent lengthening effects. As each such effect has a characteristic locus within its domain, speech timing cues are potentially disambiguated for the listener, even in the absence of other information. Diffuse timing effects – in particular, quasi-rhythmical compensatory processes implying a relationship between structure and timing throughout the utterance – are found to be weak and inconsistently observed. Furthermore, it is argued that articulatory and perceptual constraints make shortening processes less useful as structural cues, and they must be regarded as peripheral, at best, in a parsimonious and functionally-informed account.

Keywords

Speech timing, prosody, rhythm, prosodic structure, final lengthening, initial lengthening, lexical stress, phrasal accent, speech perception

1. Introduction

Speech timing appears to offer an ambiguous guide to speech structure. There are multiple potential influences on the duration of speech sounds, but the resultant variation is essentially one-dimensional: sounds can be either longer or shorter than expected. Despite this, numerous studies have shown that listeners are able to exploit durational variation in judgements of lexical and phrasal structure (e.g., Beach 1991; Gow & Gordon, 1995; Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991; Quené, 1992).

Classification of factors affecting speech timing is problematic, given the apparently paralinguistic nature of much durational variation. In Ladd's formulation, paralinguistic "aspects of vocal communication are clearly meaningful but not apparently organised along linguistic lines" (Ladd, 1996, p. 33); such aspects include the indication of interpersonal attitude, emotional state and formality of speech register. Ladd, however, made a constructive working distinction between *categorical* linguistic form and *gradient* paralinguistic form. As the following review aims to demonstrate, some consistent aspects of speech timing can be related to linguistic entities with categorical settings. Such linguistic influences on speech segment duration may be broadly stratified as segmental, syllabic and prosodic. At the segmental level, Klatt (1976) identified several intrinsic articulatory properties of broad classes: low vowels longer than high vowels; voiceless fricatives longer than voiced fricatives; bilabial stops longer than alveolar and velar stops. At the syllabic level, vowels are longer preceding voiced coda consonants, and consonants tend to be shorter when they occur in clusters (Klatt, 1976). Prosodic timing factors, the focus of this discussion, are the durational consequences of the organisation of syllables into words and higher-level constituents.

A range of timing effects have been proposed to be conditioned by prosodic organization. Considering just word-level prosody, Turk and Shattuck-Hufnagel (2000) assessed the evidence for five durational mechanisms: word-initial lengthening, polysyllabic shortening, accentual lengthening, "syllable ratio equalization" and word-final lengthening, finding support for all but the last of these. Additional processes have been associated with higher levels of prosodic structure, such as phrase-final lengthening (Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992) and greater degrees of initial lengthening at higher phrase boundaries (Fougeron & Keating, 1997). There are also durational adjustments which are hypothesised to be conditioned by the composition of prominence-delimited constituents, such as the lengthening of a lexically or phrasally stressed syllable when immediately followed by another (e.g., Bolinger, 1965; Van Lancker, Kreiman & Bolinger, 1988), as well as trends or underlying tendencies towards isochrony of stress-delimited feet (e.g., Lehiste, 1977).

1.1. A functional approach to prosodic speech timing

A functional approach is here proposed to cut through the thicket of putative prosodic timing effects in English. The intention is to identify those effects that have a linguistic communicative function and as such are integrated into the speech planning process, rather than arising from physiological constraints or from transient performance factors.

To this end, speech must be considered in communicative context – the interactive exchange of information between two or more interlocutors. Speakers use variation in the signal to encode information; furthermore, information is only transmitted if this variation is not predicted by the listeners. In theory, an auditorily-based communication system could encode information through

spectral variation alone; however, over the development of human languages, temporal variation has been adapted or exapted to communicative ends.

The glossogenetic origins of the temporal encoding of information may sometimes relate to durational consequences of articulation. Thus, the greater length of vowels before voiced consonants than before voiceless consonants is a natural corollary of the cessation of voicing during the transition to unvoiced sounds (Klatt, 1976). In languages such as English, this seems to have been exaggerated and systematised so that vowel duration is a robust cue to consonant voicing (Klatt, 1976; Raphael, 1972). At the prosodic level, the lengthening of segments at the end of phrases and utterances has often been described as arising from a universal, non-linguistic tendency, for example: "related to the general deceleration of motor activity" (Klatt, 1976, p. 1212) or reflecting "the braking that inertial systems show generally as they stop gently" (Fowler, 1990, p. 205). This parallels Gussenhoven's (2002) identification of the "Production Code" in pitch variation, with greater subglottal pressure at the start of the utterance associated with higher pitch, a tendency which has become integrated into intonational phonology so that low final pitch is associated with topic closure whilst high final pitch implies continuation (and *vice versa* for initial pitch). Similarly, as the review below indicates, if final lengthening does have a non-linguistic, motoric origin, it appears to have become systematised so that it is characterised by a phonologically-defined locus, whilst lengthening within other distinct loci may serve as cues to the onset of a word or phrase or to identify the strongest element therein. Thus, one pressure for systematisation of natural speech timing tendencies is disambiguation: given the multiple influences on segment duration, it is argued here that differential distributions of distinct prosodic timing effects serve to indicate their function. (Arguing somewhat against the non-linguistic final lengthening hypothesis, it may be noted that Snow, 1994, found that phrase-final intonation

patterns are acquired by infants at a younger age than consistent final lengthening, which he argues is therefore an acquired skill rather than a product of articulatory constraints.)

Temporal coding of information in speech is limited, however, by the number of discrete durational distinctions that speakers can produce and listeners can interpret. Many phonological distinctions are binary, including phonemic vowel length in almost all studied languages (e.g., Bye, 1997; Chomsky & Halle, 1968). The evidence reviewed below suggests that temporal coding of prosodic distinctions may also be binary, at least to the extent that deviations from expected timing are only exploited as cues to structure when segments are *lengthened* rather than shortened. Whether listeners distinguish multiple levels of lengthening is considered below.

There are several pressures conspiring towards the use of lengthening rather than shortening as a prosodic cue. Firstly, there is a temporal asymmetry intrinsic in speech production: the maximum duration of continuant sounds is primarily limited by the respiratory cycle and vowels lasting several seconds are easily achievable. However, there are obvious articulatory limits to the shortening of speech sounds at typical speech rates. At extremes of shortening, sounds are no longer realised as recognisable exemplars of underlying phonemes. For example, in Dinka, a Nilo-Saharan language providing a rare demonstration of a three-level phonemic length contrast, the shortest vowel is more centralised than the two longer ones, as there is insufficient time for full articulation; thus, it is argued that ternary quantity contrasts are unstable and tend to revert to quality contrasts (Remijsen & Gilley, 2008). Clearly, in languages with phonemic vowel length contrasts, the scope for suprasegmental lengthening *and* shortening is particularly restricted.

Secondly, there may be perceptual constraints on the processing of lengthening and shortening cues by listeners. The Effort Code interpretation of intonational patterns states that there is a universal tendency for high pitch and high pitch range to be associated with high information load (Gussenhoven, 2002). Lengthening of segments likewise implies articulatory care and should similarly be associated by the listener with informational significance (see also Lindblom, 1990, 1996, regarding "H&H" theory). Additionally, non-speech auditory events that occur earlier than expected are less well attended than those that occur later than expected (Jones & Boltz, 1989; Kim & McAuley, 2013). The principle that attention is greater to events that are overdue than those that are premature implies that lengthening of speech sounds, effectively delaying the perceptual centre of a subsequent syllable, should be a more salient structural cue for listeners than shortening.

This paper presents a synthesis of previous work on the relationship between suprasyllabic speech structure and speech timing, proposing a parsimonious and functionally-informed account. A minimum set of principles are invoked to account for previous findings, identifying timing effects that are reliably associated – through consistency, audibility and mutual exclusivity – with communicative functions (see Xu 2006, 2010, for further discussion). Thus, even in the absence of other sources of information, such cues can be interpreted by listeners as cues to speech structure.

As reviewed, the temporal processes which meet these functional criteria and offer the most parsimonious account of previous findings are lengthening effects marking important points in prosodic structure: domain-edge effects – word-initial lengthening, phrase-final lengthening; domain-head effects – lengthening of lexically-stressed syllables; lengthening of phrasally-stressed words. As the experimental evidence indicates, such effects are focused on phonologically-defined

loci. Thus, the durational events which have an abstract representation in speech planning are strictly localised within the utterance. On the basis of the evidence, I furthermore contend that there is no direct, systematic and consistent relationship between prosodic structure and speech segment duration outside the loci of such lengthening effects.

1.2. Speech timing and speech rhythm

The functional framework embodies a perspective on speech timing which contrasts with "rhythmical" approaches. The term "rhythm" has been applied in various senses to speech, and for precision I identify two related but separable components: *periodic rhythm*, defined by Couper-Kuhlen (1986, p. 51) as "the recurrence of an event at regular periods or intervals"; and *contrastive rhythm*, an alternation of strong and weak elements. Although the term "speech rhythm" is sometimes used to embrace additional, or indeed all aspects of speech timing, similar distinctions – between periodic and contrastive components of rhythm – have been made by other researchers, although the terminology varies (e.g. Brown, 1911; Port, 2003).

In English and many widely-researched languages, speech is undeniably rhythmical in the contrastive sense. Contrast between stronger and weaker elements is used to convey information, for example, in minimal stress pairs such as the nominal and verbal senses of English *impact*, *permit*, etc., or in many Spanish contrasts, including *saco* ("sack", "I take out") and *sacó* ("she took out"). As discussed below, there are durational consequences of such contrasts at both lexical and phrasal levels. However, Nolan and Jeon (under revision) discuss languages, such as Korean, which may lack even contrastive rhythm.

Periodic rhythm implies the organization of sounds into groups which impose temporal constraints on their subconstituents, analogous to bars in musical composition. Typically, the structure of these groups is defined with respect to contrastive relations between elements (e.g., stress-delimited feet or other metrical units), thus periodicity usually implies contrastive rhythm (whereas the converse is not the case). As outlined below, the phonetic evidence for regular periodic rhythm is elusive. There are nonetheless rhythmical approaches to speech timing which share the common proposition that segment duration is related to metrical structure throughout the utterance (Couper-Kuhlen, 1993; Kim & Cole, 2005; O'Dell & Nieminen, 1999; O'Dell & Nieminen, 2009; Port, 2003). Such conceptions of rhythmicity in speech have in common "the hierarchical organisation of temporally coordinated prosodic units" (Cummins & Port, 1998, p. 145). These approaches do not generally assume isochrony of any given unit, but rather predicate an interaction of temporal constraints between two or more levels of structure. The interacting constraints are typically modelled as systems of coupled oscillators: for example, O'Dell and Nieminen's (1999) model postulates syllable-level and stress-group-level oscillators. The relative coupling strength of the oscillators determines which unit tends to dominate and therefore to impose the strongest temporal constraints on the system; however, each oscillator has a characteristic eigenfrequency, determining the underlying period of each unit. Thus each individual oscillator would give rise to isochronous units were it not constrained within the system by coupling with oscillators associated with other levels of structure.

This paper is not intended as an exhaustive critique of rhythmical approaches to timing. For one thing, syllable-level timing constraints are not considered within the functional framework presented here, which relates to the influence of word and higher-level structures on timing. However, rhythmical approaches entail mediation of speech timing by suprasyllabic units

throughout the utterance. Such mediation necessarily implies that compensatory durational effects (or "rhythmic gradation") should be ubiquitous in speech: "The [coupled oscillator model] shows that rhythmic gradation is a consequence of synchronizing rhythms hierarchically. We therefore expect the phenomenon to be very widespread […]" (O'Dell & Nieminen, 2009, p. 182).

Simply put, the more segments there are within a particular superordinate timing unit, the greater should be the compression on the individual segments, and so the shorter should be their duration. For example, at the word level, [i] would be longer in *sleep* than in *sleepiness;* at the foot level, [i] would be longer in *sleep soundly* than in *sleep on it soundly*. This is the Procrustean bed of speech timing, proposed by White and Turk (2010) to have a more parsimonious interpretation in terms of localised lengthening effects (see below).

Finally, the lack of periodicity in speech – except in certain constrained cases examined later – is often attributed to non-temporal factors in linguistic production, for example, "the vagaries of syntax and lexical choice" (Cummins, 2011, p. 3). Alternatively, aperiodicity may be seen as a design feature – relating to the communicative function of spoken interaction – rather than as an artefact. As discussed above, the exchange of information between interlocutors requires that the speaker produce a signal that is not predictable for the listener. Regular timing is, by definition, predictable and thus conveys no information in itself (see also Nolan & Asu, 2009, p. 76, for discussion of language's "antirhythmic predilection").

The empirical evidence regarding timing effects in English is reviewed in Section 2, in which compensatory effects are seen to be weak and inconsistently observed, and thus not for inclusion

within a functional account of speech timing. The functional framework itself is presented in Section 3. Some implications of the framework are considered in Section 4.

2. Structural influences on speech timing

In this brief review of suprasyllabic speech timing factors, I assess the extent to which each type of effect meets the functional criteria adduced above. Broadly, I consider whether an effect is reliably observed given the appropriate structural configuration, whether it is restricted to a distinct and definable locus, and whether it is of sufficient magnitude to be readily interpretable by listeners (e.g., Xu, 2006, 2010). Whilst the audibility criterion is informed by studies that have shown just noticeable differences in speech sound duration of around 25ms (e.g., Klatt & Cooper, 1975), interpretation of prosodic timing effects must be more complex than the task of determining that one sound in a sequence is longer than another. The listener presumably evaluates the perceived length of sounds with expectations based on foregoing rate, and thereby determines whether timing has been manipulated for linguistic ends. This interpretive process must necessarily be modulated by multiple extrinsic factors – e.g., conversational context, intra-interlocutor familiarity, listening conditions – as well as the availability of other cues to speech structure (segmental, prosodic, lexical; see Mattys, White & Melhorn, 2005).

Evaluation of timing effects on the basis of their magnitude alone is additionally problematic because, given the manifold influences on segment duration, direct comparison between studies is near impossible. Thus, rather than simple assessment of the magnitude of durational variation, the functional framework requires evidence that listeners exploit specific effects to make judgements regarding speech structure. Furthermore, no more relations between prosodic structure and speech

timing should be proposed than are minimally required to account for speaker and listener behaviour.

It is useful at this point to acknowledge the distinction between speech segment duration and underlying speech timing (Kohler, 2003). Theoretical interpretations of durational data have the ultimate goal of revealing the means by which physical, articulatory, grammatical and perceptual influences on the realisation of speech gestures result in the manifest temporal structure of speech, and conversely, how this temporal structure is interpreted by listeners. Given the coordination and coarticulation of speech gestures underpinning the realisation of abstract phonemes, speech segment duration – typically measured through visual inspection of the speech waveform and a spectrographic representation – is only an approximate guide to speech timing. Particularly at the structural level, however, it serves as the best available index of the distribution and extent of underlying temporal processes.

For the sake of coherence and brevity, and towards an internally consistent synthesis, this review primarily considers studies of speech timing in English, which has been extensively researched. Work on other languages is consulted where gaps exist in the literature on English, particularly regarding the exploitation by listeners of durational cues to structure. Cross-linguistic applicability of the functional framework is considered in Section 4, where potential modifications to take account of cross-linguistic variation in prominence systems are discussed. For fuller reviews of research on prosodic speech timing, see Klatt (1976), White (2002, chapter 2) and Fletcher (2010).

2.1. Domain and locus description of speech timing

To fully characterise each suprasyllabic influence on speech timing, two stretches of speech need to be identified, the domain and the locus (White, 2002). The domain, a familiar concept in prosodic phonology (e.g., Nespor & Vogel, 1986), is the constituent that conditions the occurrence of a durational effect. The locus is the stretch of speech within which the effect is manifest. For example, consonants are longer word-initially than word-medially or word-finally (e.g., Oller, 1973); thus, the domain of word-initial lengthening is the word, and the locus is the onset of the word-initial syllable.

Within this framework, structural timing effects may be classified according to the nature of the relationship between the domain and locus. Lengthening effects are observed at domain heads and domain edges for at least two levels of prosodic structure, and may serve as cues to prominence and to constituent boundaries. Compensatory adjustments are proposed to span domains, potentially stretching or compressing all subconstituents so that successive domains of a particular type (words, feet, etc.) are of more similar duration than would arise given the intrinsic duration of their constituent segments. As the evidence reviewed below demonstrates, compensatory effects are, at best, slight and inconsistently observed.

2.2. Domain-edge effects: Durational cues to prosodic boundaries

2.2.1. Initial lengthening

Syllable onset consonants are substantially longer word-initially than word-medially or word-finally (Oller, 1973). In stressed syllables, the word-initial onset duration may be 20%-30% greater than that observed word-medially (Oller, 1973; White & Turk, 2010), and the durational difference between initial and medial position may be even greater for onsets in unstressed syllables (Oller,

1973). Word-initial lengthening affects multiple segments within the onset (Oller, 1973), but does not appear to extend to the vocalic nucleus (Turk & Shattuck-Hufnagel, 2000; White, 2002).

Listeners have been shown to use word-initial lengthening as a cue to lexical segmentation (Gow & Gordon, 1995; White, Mattys, Stefansdottir & Jones, 2014, for English; Quené, 1992, for Dutch; Tagliapietra & McQueen, 2010, for Italian). Furthermore, several studies have found that speakers increase the degree of initial lengthening following higher-level prosodic boundaries (Fougeron & Keating, 1997; Byrd, Lee, Riggs & Adams, 2005), and there is evidence that phrase-initial lengthening affects listeners' interpretation of the structure of ambiguous phrases (Cho, McQueen & Cox, 2007). In articulatory data, Byrd and Riggs (2008) found small lengthening effects for some speakers on stressed onsets at more than one syllable remove from the boundary, but given the temporal separation from the preceding boundary, it seems unlikely that any such effect could be regarded as a phrase boundary cue, in the absence of corroborative production and perception evidence.

Consonants tend to be relatively short in utterance-initial position, in some cases comparable to their duration word-medially (Fourakis & Monahan, 1988; Fougeron & Keating, 1997; White, 2002), although articulatory data would be required to determine the duration of utterance-initial stop or fricative closure. In addition to the various articulatory constraints which may apply at the (re-)initiation of speech, White (2002) suggested a functional interpretation, whereby utterance-initial boundaries – at least in the single-sentence tokens examined in the studies cited above – are unambiguously cued by the cessation of the preceding silence, rendering word-initial lengthening perceptually redundant.

2.2.2. Final lengthening

It is well established that segments are lengthened at the end of syntactic/prosodic phrases. The locus of lengthening appears to be determined by the metrical structure of the word, typically affecting the final stressed vowel and subsequent segments; thus, where the phrase-final syllable is unstressed, the nucleus and/or coda of a preceding stressed syllable may be lengthened, as well as the unstressed syllable immediately before the boundary (Cambier-Langeveld, 2000; Klatt, 1975; Nakatani, Connor & Aston, 1981; Turk & Shattuck-Hufnagel, 2007; Wightman et al., 1992). Lengthening is progressive, so that segments closer to the boundary generally receive greater lengthening, but not all segments following the final stressed vowel are necessarily affected (Turk & Shattuck-Hufnagel, 2007; White, 2002). Thus the locus of phrase/utterance-final lengthening appears to be a constituent beginning with the final stressed vowel and continuing to the boundary (termed the "word-rhyme" by White, 2002), with the distribution of lengthening within the locus dependent on its size and segmental composition.

Numerous studies have shown that listeners use lengthening at the ends of phrases and utterances as a cue to syntactic structure (e.g., Beach, 1991; Marslen-Wilson, Tyler, Warren, Grenier & Lee, 1992; Price et al., 1991; Scott, 1982). Indeed, Fletcher (2010) suggested that listeners' patterns of exploitation of suprasegmental cues to phrase boundaries provide the strongest evidence for a dissociation between syntactic structure and prosodic structure. However, an important unresolved issue concerns the number of levels of prosodic structure. There is evidence that the degree of lengthening may increase with the strength of the prosodic boundary, at least up to the intonational phrase (Wightman et al., 1992), but it remains controversial whether multiple levels of phrasing between the word and the intonational phrase are consistently marked in natural speech, as opposed to constrained laboratory materials.

The evidence is also mixed regarding the existence of word-final lengthening in the absence of higher-level boundaries (see Fowler, 1990, and Cutler, 1990, for contrasting views). This is partly because the effect is difficult to disentangle unambiguously from other prosodic effects, such as phrase-final lengthening and the interaction of word length and phrasal accent (see below). Beckman and Edwards (1990), using tightly-constrained materials, reported lengthening of vowels in word-final syllables. White and Turk (2010) further suggested that the locus of the word-final effect may be distinct from that of phrase/utterance-final lengthening: thus, in phrase-medial words, the final stressed vowel may be lengthened whilst subsequent pre-boundary consonants are unaffected. With regard to perception, Klatt (1976) speculated that word-final lengthening, in the absence of a following phrase boundary, is of insufficient magnitude to serve as a segmentation cue, but some studies have indicated that duration is a key predictor of whether a word-initial stressed syllable is interpreted as a monosyllabic word rather than as the first syllable of a disyllable (Davis, Marslen-Wilson & Gaskell, 2002; Salverda, Dahan & McQueen, 2003). Whilst this potentially indicates the perceptual validity of word-final lengthening, it may be that listeners interpret the durational pattern as evidence of a phrase boundary, or of the greater duration of phrasally-stressed syllables in shorter words. There is also evidence from artificial language learning that lengthening of vowels at the end of statistically-defined words promotes segmentation (e.g., Saffran, Newport & Aslin, 1996).

2.3. Domain-head effects: Durational cues to prominence

Prominent units within speech are lengthened and this lengthening is salient for listeners (e.g., Fry, 1955; Klatt, 1976). However, the number of distinct levels of prominence remains at issue (e.g., Shattuck-Hufnagel & Turk, 1996), and no complete model of the role of timing in prominence

perception can ignore the impact of covariation in vowel quality, loudness and, in particular, fundamental frequency. For example, vowels accompanied by an F0 movement are perceived as longer than when F0 is static (e.g., Cumming, 2011; Pisoni, 1976). Additionally, phonetic variation associated with prominence may extend beyond the domain head – thus, for example, Xu and Xu (2005) found that the F0 range of post-focus constituents was reduced – although the impact of such effects on listeners' judgements remains to be determined.

In light of the above caveats, the minimal claim arising from this brief review is that durational cues serve to distinguish at least three levels of prominence distinction, in English and other languages with similar stress systems: no stress, lexical stress and phrasal accent (as discussed further below, some other languages may lack durational marking of prominence altogether, whilst still manifesting domain-edge timing effects). For both lexical stress and phrasal accent, the perceptual salience associated with durational and other cues to prominence has a clear communicative function. At the lexical level, this is demonstrated by the existence of stress-based minimal pairs (e.g., '*insight* vs *in*'*cite)* in English. At the phrasal level, the pragmatic interpretation of an utterance is well established to be affected by the placement of phrasal accent (e.g., Ladd, 1996).

For lexical stress, quantification of the magnitude of lengthening is confounded because, in English at least, syllables that lack stress usually contain reduced vowels. Consonants in unstressed syllables – particularly coda consonants – are also subject to reduction processes in naturalistic discourse. Studies using reiterant speech have shown, however, that both vowels and consonants in unstressed syllables are shorter than those in stressed syllables (e.g., Oller, 1973), although note that target words in such studies are typically in focus, as the new information in the utterance, so

effects of lexical stress and phrasal accent are frequently confounded. With regard to perception, it is well established that longer syllables are more likely to be perceived as prominent, with F0 variation and relativeness loudness also contributing (e.g., Fry, 1955; Kochanski, Grabe, Coleman & Rosner, 2005).

Klatt (1975) reports a small (~5%) durational difference between primary and secondary stressed syllables, although the caveat regarding the confounding of lexical stress and phrasal accent in experimental tokens also applies here. With regard to perception, Mattys (2000) reported that listeners can perceive a distinction between primary and secondary stress, relying on F0 and loudness as well as duration, the relative contribution of the different cues being indeterminate. Thus, lengthening distinguishes stressed from unstressed syllables, and possibly primary from secondary lexical stress.

Lexical stress and phrasal accent are distinguished, however, in both the magnitude and the locus of lengthening. The greatest degree of accentual lengthening is seen on the primary stressed syllable of the word (Sluijter, 1995; Turk & Sawusch, 1997), but other syllables are also lengthened in polysyllables, with concomitant attenuation of lengthening of the primary stress according to the number of additional syllables (Turk & White, 1999; White & Turk, 2010). Segments at word edges, in addition to the primary lexical stress, appear to attract the greatest degree of accentual lengthening (White, 2002).

Data are not available on whether different levels of phrasal prominence (e.g., prenuclear vs nuclear vs contrastive) have durational consequences over distinct loci, and the difficulty of comparison between studies also leaves open the possibility of variation in the magnitude of

lengthening. Thus, work remains to be done on the number of acoustically-distinct levels of phrasal prominence and whether listeners are capable of interpreting such distinctions for communicative ends.

2.4. Compensatory effects

2.4.1. Stress-delimited feet and the Procrustean bed

The isochrony hypothesis (e.g., Abercrombie, 1967) is the most notable instance of the Procrustean bed in speech timing, proposing that segments in languages like English or Dutch are stretched or compressed to preserve uniformity of duration in stress-delimited feet. By contrast, in French and Spanish, it is the syllables themselves that are hypothesised to be subject to durational equalisation. Notoriously, however, the evidence for foot-level isochrony in English is simply absent (Classe, 1939, and many subsequent studies). For example, Lehiste (1973) and Dauer (1983) found that inter-stress interval duration increases almost linearly with the number of intervening unstressed syllables. Furthermore, Roach (1982) and Dauer (1983) found no less variability in inter-stress intervals in "stress-timed" languages than "syllable-timed" languages. Other levels of prominence have been considered: Bolinger (1965) failed to find isochrony of intervals between phrasally-stressed syllables, and Shen and Peterson (1962) found wide variation in intervals between syllables carrying nuclear accent. With regard to syllable isochrony, Roach (1982) found that the variability in syllable duration in "syllable-timed" French, Telugu and Yoruba was comparable to that in "stress-timed" Arabic, English and Russian. Likewise, Pointon (1980), re-analysing previous studies, found that the duration of Spanish syllables depends largely on their segmental composition, with little evidence of syllable-based isochrony.

Lehiste (1977), in the influential paper "Isochrony Reconsidered", claimed that, nevertheless, foot-level isochrony exists as a timing principle in English, because listeners do not perceive most deviations from isochrony. She cites Lehiste (1975) as showing that listeners had considerable difficulty identifying the longest or shortest of four stress-delimited feet in spoken utterances, whereas the equivalent task using clicks separated by noise was much easier. However, the result could equally be interpreted as showing that listeners do not attend to inter-stress intervals because they do not provide any information required for linguistic interpretation. Perceptual isochrony was indeed questioned by Scott, Isard and Boysson-Bardies (1985), who found that French and English listeners, when imitating the rhythm of auditory stimuli, regularised inter-stress intervals in speech more than the beats of a simple non-speech stimulus. There was no specific bias for English listeners to perceive inter-stress isochrony in speech, and they concluded that participants simply regularised intervals when the task became difficult due to complexity of the stimulus, regularisation being also found with non-speech stimuli which had the acoustic complexity of speech.

Inspired by coupled oscillator models of timing, Kim and Cole (2005) looked again for evidence that the stress-delimited foot is a "unit of planned timing" in English. As expected, the isochrony hypothesis was not supported: foot durations increased as a linear function of the number of syllables therein (in this case they measured foot duration from stressed vowel onset, as an approximation to perceptual centre location, see Morton, Marcus & Frankish, 1976). They found that rhymes of stressed syllables, but not those of unstressed syllables, were shortened as a function of the number of syllables in the foot; however, this shortening was only found within the "intermediate phrase" in prosodic structure. Although Kim and Cole interpreted the results as evidence for foot-based timing, the localised nature of the effect, together with the interaction with

prosodic boundary location, permits a more parsimonious interpretation in terms of established effects. Thus, stressed syllables are longer when they are closer to upcoming word and phrase boundaries, i.e., when fewer unstressed syllables intervene (see earlier discussion of domain-final lengthening). Additionally, as found in other studies discussed below, by far the greatest "shortening" effect was between feet of one and two syllables: it is unclear why, under a temporal coordination hypothesis, commensurate shortening was not observed in feet of three or more syllables, whilst this is easily explained in terms of established localised lengthening effects. Thus, not only are isochronous feet not observed in English, but there is little justification for claiming isochrony as an underlying tendency from which large deviations are interpreted as linguistically significant. The evidence simply does not support the stress-delimited foot as a dominant element in English speech timing under unconstrained speaking conditions (see Cummins & Port, 1998, discussed below, regarding the foot as a coordinating unit in speech cycling tasks).

In accordance with the findings of Kim and Cole (2005), there do appear to be grounds for the claim that a stressed syllable is lengthened when followed immediately by another stressed syllable (Bolinger, 1965), a localised effect that may be termed "stress-adjacent lengthening" (White, 2002). Rakerd, Sennett and Fowler (1987) and Van Lancker, Kreiman and Bolinger (1988) found lengthening of, for example, *peach* in *peach light* compared with *peach delight*. The effect was very small, however, and there are several experimental factors which make interpretation difficult, particularly the uncertain phrasal accent status of the target syllables. If such an effect exists, independent of other established lengthening effects, then it appears, as Rakerd et al. found, to occur regardless of word and higher-level prosodic constituent boundaries. Indeed, Fant, Kruckenberg and Nord (1991), reporting evidence from Swedish, proposed stress-adjacent lengthening as the sole timing consequence of the "rhythmical" organisation of speech: "The main

effect appears to be in the step from none to one following unstressed syllables in the foot. However […] these effects are marginal and not sufficient as a basis for a theory of 'stress timing'" (Fant et al., 1991, p. 84).

2.4.2. Periodic rhythm in speech performance

Several strands of experimental work have examined the ability of people to speak with periodic rhythm, particularly when given a regular external stimulus. The results of such experiments are sometimes overgeneralised. For example, Rakerd et al. (1987) cited Fowler (1981) as finding that stressed syllable duration was inversely related to the number of following unstressed syllables. However, subjects produced the syllables within a fixed frame sentence in time with a metronome (i.e., they placed stressed syllables on regularly-occurring beats). In the original paper, Fowler admitted that metronome pacing may have strongly increased the shortening effect of subsequent unstressed syllables (necessary for speakers to align the beats). Thus, Rakerd et al.'s extrapolation to ordinary unconstrained speech does not appear justified.

Similar caveats apply to "speech cycling" experiments which look for coordination of temporal units. For example, Cummins and Port (1998) required speakers to repeat short phrases (e.g., *big for a duck*), containing two stressed syllables to be aligned with the low and high tones of a metronome. Temporal intervals between the tones were varied. They found that English speakers tended to place the onsets of stressed syllables more regularly in time than the occurrence of the tones: specifically, stresses were aligned with points within the overall phrase repetition cycle which divided the cycle into regular intervals. They suggested that the patterns can be understood as the nesting of one unit (the stress foot) within a larger unit (the phrase repetition cycle). This hierarchical coordination of speech units has been modelled in subsequent work as a product of the

attraction of beats to certain harmonic fractions of the periodic cycles of phase-locked coupled oscillators (e.g. Port, 2003).

Demonstrating temporal coordination of constituents under highly constrained conditions is clearly not the same as showing that such dependencies play a role in ordinary speech. Indeed, Cummins and Port (1998) repeatedly stated that the coordination between stress placement and the higher-level cycle is task-specific, indicative of the emergence of stable point attractors in the constrained system. When rhetorical impact is desired, however, natural speech can tend towards similar highly coordinated rhythmicity. For example, Knight (2013) proposed that persuasive oratory may be manifest by a periodic speech style compared to everyday interaction, possibly serving to enhance entrainment between speaker and listeners. In support of this, she found that listeners required to tap along to the "beat" of recorded speech samples showed less inter-tap interval variability with rhetorical than conversational speech, and even less variability to recitations of poetry with a regular metre.

Figure 1 shows the waveform of former US President Bill Clinton's notable declaration: "I did not have sexual relations with that woman, Miss Lewinsky". There is striking regularity of inter-stress intervals, particularly for the first four stressed syllables. Clinton's engagement with the utterance's contrastive rhythm is evident: he can be heard and seen banging on his lectern in time with the stressed syllables (the footage is easily found on video-sharing websites). Indeed, the coordination of gesture and speech in such an emphatic and emotive context may well serve to induce stable attractors in the system and hence relative periodicity, similar to that modelled for speech cycling tasks.

However, in common with many notable public speakers, Clinton achieves this quasi-regular periodicity not by compression or expansion of speech segments, but by the insertion of silent intervals between words. His greatest deviation from isochrony comes where several weak syllables intervene between stresses – "relations with that woman" – where he speeds up his speech excessively and undershoots the desired interval duration. The insertion of non-hesitation silent intervals within phrases is relatively uncommon in ordinary speech, so to achieve the same degree of temporal coordination would require the operation of compensatory processes, for which, as reviewed here, there is little consistent evidence.

2.4.3. Compensatory effects within prosodic constituents

Rhythmical approaches to speech timing, as discussed above, necessarily entail the ubiquity of compensatory effects. If a superordinate unit imposes temporal constraints on its constituents (e.g., O'Dell & Nieminen, 1999, 2009), there must be inverse relationships between the number of constituents in that unit and the duration of those constituents. As discussed above, the evidence for compensatory processes within metrical constituents is weak and can be more parsimoniously explained with reference to other well-established lengthening effects. As this brief review indicates, this is also true of compensatory processes that have been hypothesised to operate over prosodic constituents, i.e., constituents of spoken language that relate to syntactic rather than metrical structure and are generally delimited by syntactically defined boundaries (no particular theoretical account of prosodic structure is implied here).

Various compensatory processes have been held to operate within prosodic constituents, most notably at the word level, for which the polysyllabic shortening hypothesis suggests that the number of syllables in the word is inversely related to the duration of the segments therein (e.g.,

Lehiste, 1972; Port 1981). However, most studies have used words or nonsense words in fixed-frame sentences (e.g., *I say [dip|dipper|dipperly] again every Monday*, Port, 1981), entailing that target words inevitably carried phrasal accent, and were therefore subject to lengthening of the primary stressed syllable and other syllables within the word (e.g., Turk & Sawusch, 1997; Turk & White, 1999).

Subsequent studies have examined the evidence for polysyllabic shortening and other word-level timing effects – in particular, word-initial lengthening and word-final lengthening – using targets with and without phrasal accent (Turk & Shattuck-Hufnagel, 2000; White, 2002; White & Turk, 2010). For example, White and Turk (2010) used monosyllables, disyllables and trisyllables, both left-headed (e.g., *mace*, *mason*, *masonry*) and right-headed (e.g., *mend*, *commend*, *recommend*). Strong polysyllabic shortening effects were observed in the primary stressed syllables of accented words, but in unaccented words, the durational effects of word length related largely to the alignment of segments with boundaries. Thus, for example, [m] was longer in *mend* than in *commend*, attributable to word-initial lengthening (Oller, 1973), but the vowel in the monosyllable was no longer than in the disyllable, and there was no shortening of either vowel or consonant in *recommend* compared with *commend*. There were very minor residual effects of word length, which may be subject to various interpretations, such as word-final lengthening, but these did not affect all constituents of the stressed syllable nor apply analogously to left-headed and right-headed words, as would be expected under the polysyllabic shortening hypothesis.

The evidence from this and earlier studies, in which all target words carried phrasal accent, are clearly consistent with the view that the effect of word length on stressed syllable duration arises primarily from the distribution of accentual lengthening among the subconstituents of the word

(White, 2002; White & Turk, 2010). This modulation of accentual lengthening according to word length is consistently observed (e.g., Turk & Shattuck-Hufnagel, 2000; Turk & White, 1999; White & Turk, 2010). Thus, in monosyllables, the stressed syllable receives all of the prosodic lengthening, but in disyllables and trisyllables, some of the lengthening spreads to the unstressed syllables and lengthening of the stressed syllable is concomitantly attenuated. This is not, therefore, a compensatory effect, but rather the attenuation of a prosodically-determined lengthening effect.

The interpretation of polysyllabic shortening as the attenuation of a lengthening effect is parsimonious, invoking only one well-established process – salient elements (domain heads) are lengthened in speech – and requiring one consistently-observed principle of implementation: the magnitude of lengthening on any element within the locus is proportional to the number of additional elements therein. Other compensatory effects that have been hypothesised to operate over prosodic domains are likewise interpretable as the attenuation of localised prosodic lengthening effects due to the addition of segmental material within the locus. For example, an apparent compression of segments due to the addition of syllables at the utterance-level (e.g., Gaitenby, 1965) can be more parsimoniously interpreted as an attenuation of phrase-final lengthening: thus, a word that is lengthened in final position loses that prosodic lengthening when additional syllables follow it within the phrase, not because of an overall compression due to increasingly phrase length, but simply because of the removal of the word from a localised source of lengthening. Furthermore, such compensatory effects as may remain debatable (e.g., a possible polysyllabic shortening effect in the nucleus of right-headed disyllables lacking phrasal accent, Turk & Shattuck-Hufnagel, 2000), are without exception a matter of a few milliseconds between

longer and shorter words, and manifestly do not provide listeners with reliable information about structure.

3. The functional framework: Localised lengthening effects in a domain-and-locus schema

Studies of speech timing have robustly demonstrated that the duration of segments increases at certain important locations within utterances. These localised lengthening effects are reliable, mutually exclusive in their loci of effect, and have been shown to influence listeners' interpretation of linguistic materials. For domain-edge effects, segmental lengthening guides listeners' judgement of the location of word and phrase boundaries. For domain-head effects, lengthening is a key determiner in the perception of prominence, specifically, lexical stress or phrasal accent (in languages with such prominence systems).

By contrast, compensatory shortening effects are small and – at best – inconsistently observed, diffuse rather than associated with a particular domain or locus, and have not been reliably demonstrated to affect listeners' linguistic judgements. Indeed, it is difficult to see how listeners could interpret compensatory processes in parallel with the well-attested localised lengthening effects, particularly compensatory effects over the multiple nested domains characteristic of rhythmical approaches to speech timing. Thus, compensatory effects are excluded from the functional framework, the central principles of which are listed in Table 1 and explicated below.

3.1. No privileged timing unit

There is no unit into which an utterance may be exhaustively parsed that consistently imposes timing constraints upon its subconstituents. In this regard, this approach accords with the view expressed by van Santen (1997, p. 237): "[There is] rarely if ever […] any type of constancy of

larger units. So, even though the larger units play critical roles in speech production, they are not a happy choice as temporal units. Speakers do not carefully control timing over long stretches of speech." By contrast, the rhythmical approaches to speech timing discussed above imply that metrical structure imposes temporal constraints throughout the utterance, modulated by the coupling strength between oscillators corresponding to different levels of that structure. Of course, speech rate may be relatively consistent over long stretches: how this consistency may arise is discussed further below.

3.2. Prosodic timing domains

Structural influences on speech timing relate to the organisation of syllables into words and higher-level constituents. Unlike strict prominence-delimited constituents, prosodic domains respect word boundaries and have a relationship to syntactic structure, mediated by non-linguistic factors such as speech rate and constituent size (e.g. Nespor & Vogel, 1986; Selkirk, 1996; Wightman et al., 1992). The linguistic communicative utility of timing effects conditioned by such domains is more transparent than that of compensatory processes within metrically-defined constituents. (Although the locus of final lengthening – the word rhyme, see below – is defined with respect to stress placement, it is terminated by a word boundary.)

.

3.3. Localised lengthening effects

Speech segment duration is determined with respect to prosodic structure only at particular points in speech. At these loci (listed in 3.4), segments are lengthened; there are no shortening processes in the functional framework. As argued above, the lengthening-not-shortening principle offers articulatory and perceptual plausibility.

Domain-edge effects are word-initial lengthening (e.g., Oller, 1973) and phrase-final lengthening (e.g., Klatt, 1975). Domain-head effects are lengthening in lexically-stressed syllables (e.g., Oller, 1973) and in phrasally-stressed words (e.g., Turk & Sawusch, 1997). These effects have been shown to serve as cues for listeners to syntactic boundaries (e.g., Price et al., 1991) and prominences (e.g., Fry, 1955). The status of word-final lengthening remains uncertain, but some studies have suggested that stressed vowel duration increases with proximity to the end of the word even in the absence of higher-level boundaries (e.g., Beckman & Edwards, 1990; White & Turk, 2010). There is insufficient evidence to determine whether word-final lengthening in the absence of a following phrase boundary is a reliable cue for listeners, although Saffran et al. (1996) demonstrated the utility, in an artificial language task, of lengthening of word-final vowels.

3.4. Structurally-defined loci

At each domain edge or domain head, segments are lengthened within a structurally-defined locus. The loci of lengthening are distinct from each other, thus facilitating listeners' interpretation of these effects as boundary or prominence cues (*cf* the mutual-exclusivity principle, Xu, 2006, 2010). Detection of lengthening and recognition of the locus serve to indicate the communicative function.

Word-initial lengthening. The locus is the onset of the word-initial syllable (e.g., Oller, 1973).

Phrase-final lengthening. The locus is the word-rhyme, which extends from the nucleus of the final stressed syllable to the end of the phrase (e.g., Turk & Shattuck-Hufnagel, 2007; White, 2002).

Lexical stress lengthening: The locus is the stressed syllable, with greatest lengthening on the vowel (e.g., Klatt, 1974; Oller, 1973).

Phrasal accent lengthening: The locus is the accented word. Within the word, the distribution of lengthening depends on word structure, but the greatest lengthening occurs on the primary stressed syllable; word edges also show significant lengthening (e.g., Turk & Sawusch, 1997, Turk & White, 1999).

The locus of word-final lengthening appears distinct from that of phrase-final lengthening. Specifically, the primary stressed vowel is lengthened with increasing proximity to end of the word. Segments following the primary stressed vowel do not appear to be subject to lengthening in the absence of a phrase-boundary (White, 2002; White & Turk, 2010).

The magnitude of lengthening of any given segment within the locus is dependent on the number and nature of the subconstituents therein. Thus, the fewer phonetic elements the locus contains, the greater the expansion of any particular segment. At domain edges, lengthening is typically progressive, i.e., increases with proximity to the boundary (Turk & Shattuck-Hufnagel, 2007). Thus, durational cues for listeners are directionally oriented with respect to prosodic structure, in contrast with typical rhythmical timing approaches (see below for discussion of the alignment of lengthening effects in terms of the π-gesture account).

Both phrase-final lengthening and phrasal accent lengthening have sometimes been observed to be discontinuous, in polysyllabic loci in particular: i.e., not all segments within the locus undergo significant lengthening (e.g., Dimitrova and Turk, 2012; Turk & Shattuck-Hufnagel, 2007). The distribution of lengthening in the loci seems, in part, a matter of phonological specification (e.g., phrasally-stressed words are typically most lengthened at their edges and on the primary stress; final lengthening is progressive towards the boundary) and partly determined by the articulatory

mechanics associated with the segments within the locus. The concept of segmental elasticity is useful here: thus, certain segments are more resistant to being substantially extended beyond their intrinsic duration than others (Campbell & Isard, 1991). This factor seems more relevant to consonants than to vowels: whilst van Santen (1992) found vowels to be relatively uniform in terms of their responses to timing effects, Klatt (1976) reported that the magnitude of lengthening of final consonants depends on their manner of articulation, sonorants and fricatives being more expandable than plosives.

It should be noted that elasticity has also been expressed in terms of "incompressibility" (Klatt, 1976). However, the notion of compression of segments is not useful within the functional framework, in which all structural influences are realised as lengthening effects. What have previously been observed as compensatory effects can be seen to reflect the sharing out of lengthening amongst multiple subconstituents of the locus. Most notably, polysyllabic shortening is parsimoniously interpreted as attenuation on the primary stress of a polysyllabic word of phrasal accent lengthening, by comparison with a monosyllable in which all the additional length is concentrated (White & Turk, 2010).

Finally, because the loci of lengthening effects are structurally determined, this framework contrasts with the interpretation of speech timing as a product of purely biomechanical constraints on articulation, for example, the view that final lengthening reflects a generalised and diffuse deceleration of the articulatory system (Berkovits, 1994; Cummins, 1999; Fowler, 1990; Tabain, 2003). Naturally, biomechanical constraints are relevant to the realisation of localised effects, as noted above.

3.5. No temporal mediation outside loci

There is no direct relationship between prosodic structure and segment duration apart from at domain edges and domain heads. Excepting lengthening effects within these loci, observed patterns of speech timing arise from the articulatory requirements associated with the phonological specification of the segmental string, as mediated by global speech rate. This echoes the description by Pointon (1980) of "segment timing" (rather than "syllable-timing") in Spanish, which he described as "antirhythmic". Likewise, van Santen (1997, page 237) argued: "For temporal units in speech production, the smaller, the better."

Detailed exposition of the articulatory influences on "intrinsic" segmental duration is beyond the scope of this paper. Some well-established patterns are outlined in the introduction, where durational consequences of syllabic composition are also considered (e.g., lengthening of a nucleus when followed by a voiced coda; shortening of consonants in clusters). The functional framework does, however, afford a perspective on interactions between syllabic composition and prosodic timing. For example, Klatt (1976) observed that large coda voicing effects were only seen in phrase-final position. The interpretation is that, within the locus of final lengthening, additional duration is shared out among the subconstituents. The amount of lengthening on a given segment depends, as stated above, on the number and nature of the other segments in the locus. Because voiced consonants are less expandable than voiceless consonants, the nucleus receives more final lengthening than when followed by a greatly lengthened voiceless coda. Likewise, utterance-final stressed vowels manifest greater duration in open than closed syllables (Campbell & Isard, 1991) because there are no other constituents within the locus to share the lengthening.

The functional framework also leads to the prediction that the coda voicing effect should be amplified in phrasally-stressed words, due again to the sharing out of prosodic lengthening amongst segments of variable elasticity (although the effect may not be so great in the absence of the directional influence that holds phrase-finally). Likewise, shortening of consonants within clusters compared to singletons may be amplified within domain edges (such as word-initially) and domain heads.

As well as being specific testable predictions arising from the functional framework, these potential interactions between segmental/syllabic and structural timing processes indicate the need for careful control of materials in studies of segmental influences on timing. At minimum, the phrasal accent status and orientation with respect to boundaries of target words should be considered; at best, target segments should be recorded in contrasting phrasal positions (initial, medial, final), and in words with and without phrasal accent. This practise is more prevalent in modern studies, but some earlier studies – e.g., of polysyllabic shortening – have been frequently cited without reference to the sentence position and phrasal accent of target words, factors which crucially interact with other potential timing effects.

3.6. Speech rate

Given that prosodic structure does not influence timing throughout the speech string, an understanding of the factors underlying the emergent speech rate is important. As described by Kohler (2003, p8), rate "sets the frame for timing vocal tract trajectories and for pitch control over long stretches of speech." The factors that contribute to manifest rate are multiple and interacting, including individual anatomy and physiology, emotion and arousal, age and dialect (see Fletcher, 2010, for a review). In addition, performance factors such as external time pressure and rhetorical

intent influence overall speech rate, and may, under certain circumstances, result in the transient emergence of periodic rhythmicity, as described above. In general, however, it seems unlikely that speech rate is explicitly modulated with reference to some central internal timekeeper: studies of synchronised speech clearly demonstrate the two speakers can maintain a mutual and consistent rate whilst reading the same text in parallel (e.g., Cummins, 2009), a task which has been argued to demonstrate the task-specific sensorimotor coordination possible in skilled action even in the absence of a framing temporal regularity (Cummins, 2011). Similarly, the emergence and maintenance of a consistent speech rate in an individual speaker need not require the rate to be explicitly calibrated at some level of articulatory planning.

The role for speech rate within the functional framework is as a framing background within which structural lengthening effects can be interpreted by listeners. It has been demonstrated that variation in foregoing speech rate can affect subsequent perception, both in terms of whether segments are perceived or not (Dilley & Pitt, 2010), and whether localised lengthening is interpreted as a cue to word juncture (Reinisch, Jesse & McQueen, 2011). Clearly, as outlined above, listeners must generate an expectation about segment duration based on overall speech rate against which structural lengthening can be judged. In the functional framework, the lack of need for adjustment of durational expectations which would be required given ubiquitous compensatory effects renders the listener's detection of domain-head and domain-edge lengthening cues less onerous.

4. Implications of the functional framework

The functional framework makes a number of strong claims about the relationship between speech timing and structure, on the basis of the evidence from research reviewed above. These claims lead

to experimentally testable predictions which are distinct from those of alternative approaches to timing. Thus, in normal spoken interactions:

1. Compensatory timing processes related to metrical or prosodic structure should be minimal outside the loci of domain-edge and domain-head effects, except under conditions of external timing constraint.
2. Because localised effects are manifest across distinct loci, listeners should be able to distinguish domain-initial, domain-final and domain-head lengthening.
3. Listeners should be more sensitive, as cues to immediate structure:
    - To lengthening than to shortening effects.
    - To localised lengthening than to diffuse timing effects.

There is already wealth of evidence supporting prediction 1 from the foregoing literature, as discussed above. Localised lengthening effects are large, consistent and clearly interpretable, whilst compensatory effects are elusive, small in magnitude when observed and permit multiple interpretations. Further refinement of the functional framework would require experimental studies to align production and perceptual data. In conclusion, potential implications of this work are briefly discussed in the following sections.

4.1. Speech production

4.1.1. The gestural implementation of structural timing effects

Two standpoints on speech timing may be usefully contrasted. One view proposes that speech is intrinsically rhythmical, in the periodic sense, entailing a direct relationship between structure and timing throughout the utterance (e.g., O'Dell & Nieminen, 1999, 2009). Alternatively, observed timing patterns may be viewed as a by-product of the speech production process, arising from the interaction of neural control mechanisms with the physical constraints of the articulators, such as

the view of final lengthening as due to a gradual deceleration in supralaryngeal articulation towards the end of an utterance (e.g., Fowler, 1990). The functional framework implies a middle ground between these two, wherein intrinsic speech production constraints are responsible for manifest timing throughout the utterance and structural influences impinge, to communicative ends, only at domain heads and domain edges.

One possible implementation of this framework as a formal model could be a modified version of π-gesture approach to prosodic timing effects. As originally formulated (Byrd & Saltzman, 2003), initial and final lengthening were modelled as slowing of articulatory control around the boundary: the prosodic gesture (π-gesture) warped the local clock-rate symmetrically, slowing as the boundary approaches and speeding up again as it is passed. However, as discussed above, final lengthening effects in English are strongly conditioned by the location of the final stressed vowel (e.g., Oller, 1973; Turk and Shattuck-Hufnagel, 2007; White, 2002), extending earlier than the final syllable when that is unstressed. To take into account the dependence of boundary effects on stress location, Byrd and Riggs (2008) proposed that the π-gesture approach could be modified, by allowing the gesture either to shift or to stretch towards the boundary-adjacent stress. In contrast, large and consistent post-boundary effects are only observed on the onset of the word- or phrase-initial syllable (Oller, 2002; White, 2002).

The lengthening effects within the functional framework could potentially be instantiated by a set of π-gestures, each aligned with its distinct locus. Implementation of these gestures may differ between domain-heads and domain-edges, however. Final lengthening is associated with a change in gestural stiffness, analogous to a slowing of local speech rate, whilst phrasal accent (but not final lengthening) is realised through substantial increase in the displacement of articulators

(Beckman & Edwards, 1992; Edwards, Beckman & Fletcher, 1991). The notion that the two structural functions of lengthening are realised through distinct articulatory mechanisms is a compelling one, but implementation is complicated by the degree of prosodic interaction between domain-heads and domain-edges. For one thing, the loci of final lengthening effects are determined with respect to the location of the preboundary stressed syllable; furthermore, in English at least, many words subject to phrase-final lengthening are also lengthened as a result of phrasal accent, given that the default location of primary phrasal accent is the final content word in the phrase. (For fuller discussions of the interaction between domain-edge and domain-head lengthening effects, see Cambier-Langeveld, 2000, Fletcher, 2010, and references therein.)

4.1.2. The functional framework applied to languages other than English

Within the functional framework, the critical role for prosodic structure with regard to timing is to condition the occurrence of domain-edge and domain-head lengthening effects. In English, the locus of final lengthening and the distribution of phrasal accent lengthening within the word are determined with respect to the location of lexical stress. This requires some modification for languages in which the concepts of lexical stress is problematic. For example, in French, prominence is only defined with respect to the phrase and not at the lexical level; indeed, the marking of prominence and phrase finality are confounded (see discussion in Fletcher, 2010). (It may be noted that this is critically problematic for the rhythm class typology that defines French as rhythmically analogous to Spanish, a language with many minimal pairs contrasting only in the location of lexical stress.) Likewise, Korean prominence appears undefined at the lexical level (Nolan & Jeon, under revision), but substantial phrase-final lengthening effects have been reported (Lee & Seong, 1996), as well as phrase-initial lengthening localised to the syllable onset (Cho & Keating, 2001).

Anticipating such findings, Beckman (1992) suggested that domain-edge lengthening effects are evident in all languages. Indeed, many languages have been shown to manifest phrase-final lengthening (see references in Fletcher, 2010), and word/phrase-initial lengthening has been consistently observed where studied (e.g., Keating, Cho, Fougeron and Hsu, 2003). However, Beckman raised the possibility that the marking of domain-heads through lengthening is restricted to certain languages. As noted above, French and Korean would appear to be candidates for exclusion from this category. In Beckman's formulation, she specifically identified "stress-timed" languages as those which manifest domain-head lengthening effects. However, many researchers question the concept of a categorical rhythm typology (e.g., Arvaniti, 2009; White & Mattys, 2007a; White, Wiget & Mattys, 2012), and evidence from both production (White & Mattys, 2007a,b) and perception (Arvaniti & Rodriquez, 2013; White et al., 2012) indicates that differences between languages in terms of contrastive rhythm are far from categorical. Any typological statements about the relationship between contrastive rhythm and prosodic timing would therefore be expected to show similar gradience.

An alternative hypothesis (White & Mattys, 2007b; White, Payne & Mattys, 2009) is that languages show covariance in the magnitude of durational marking of all aspects of prosodic structure: lexical stress, phrasal accent and domain-edges (clearly, if languages like French and Korean do not mark domain-heads durationally, they must be excluded from this generalisation). With specific regard to lexical stress and phrasal accent, it has similarly been suggested that "the preference of a language for a steep or shallow prominence gradient may extend to prosodic phenomena beyond the syllable" (Nolan & Asu, 2009, p. 66). There is some evidence for such correlations within well-studied European languages. English and Dutch have both a high degree

of temporal stress contrast and strong prosodic timing effects: i.e., both phrasal accent and phrase finality are associated with substantial lengthening. White and Mattys (2007b) noted preliminary evidence that durational marking of both domain edges and domain heads is attenuated in Spanish (Frota, D'Imperio, Elordieta, Prieto & Vigário, 2007; Ortega-Llebaria & Prieto, 2007). More recently, a study exploring this hypothesis found support for covariance in the magnitude of head and edge durational effects (Prieto, del Mar Vanrell, Astruc, Payne & Post, 2012). However, only Catalan, English and Spanish were considered, and more extensive cross-linguistic work is required.

As discussed above, languages may differ not only in the magnitude of localised lengthening effects, but also in the loci over which they are distributed. Consistent with the functional framework, Suomi, Meister, Ylitalo and Meister (2013) found that the magnitude of phrasal accent lengthening in Northern Finnish and in Northern Estonian is similar between words of different lengths and structures, whilst the locus of the lengthening effect is consistent language-internally but differs between the two. Furthermore, the central timing principle expounded here, that structure is cued by localised lengthening but not shortening, has been shown to be valid in Northern Finnish, distinct from English in having quantity distinctions in both vowels and consonants. Thus Suomi (2007, 2009) found that Finnish words, regardless of syllable number, showed similar overall magnitude of lengthening due to contrastive phrasal accent, that this lengthening was distributed within a consistent locus and that there was little evidence of polysyllabic shortening in words that did not carry contrastive stress.

4.2. Speech perception

Given that domain-head and domain-edge effects are consistently produced by speakers and interpreted by listeners, they must be transmitted in early language acquisition. Although Snow (1994) found that infants' development of utterance-final durational patterns lags behind consistent final intonation, lengthening of final syllables is exaggerated in adult speech directed to infants in their first year of life (Albin & Echols, 1996). Furthermore, infants as young as five months old appear to pay attention to final lengthening in discrimination tasks, better distinguishing utterances that differ in the magnitude of this localised timing effect over differences in more global timing properties (White, Floccia, Goslin & Butler, in press). Such early sensitivity to domain-edge cues is unsurprising given the importance of speech segmentation as a precursor to development of the lexicon (Christophe & Dupoux, 1996). Additionally, once a recognition vocabulary has begun to develop, the interpretation of domain-head effects – specifically, phrasal accent cues – would facilitate the associations of words with real-world referents in infants of six months or more.

Given the preponderance of strong contextual cues to juncture, it is likely that adult listeners, by contrast with infants, do not habitually rely on durational effects for word segmentation. Experiments demonstrating such effects often rely on stimuli from which other potential juncture cues have been eliminated (e.g., near-homophones such as American English *two lips* vs *tulips*, Gow & Gordon, 1995). Mattys et al. (2005) have shown that acoustic-phonetic cues to word boundaries tend to be overlooked when higher-level information – lexical, syntactic, semantic – is sufficient to for listeners to infer the locations of word boundaries. In the absence of higher-level cues, such as when speech is decontextualized or ambiguous, acoustic-phonetic variation is recruited to this purpose. Where intelligibility is further compromised by impoverished listening conditions, such as background noise, then those acoustic-phonetic cues that remain perceptible come to the fore. Lexical stress is salient in moderate noise levels and thus represents an important

fall-back cue for segmentation (Mattys et al.). Durational effects may also be relatively robust in noise, to the extent that they change the overall amplitude envelope of speech and can be interpreted without full segmental information. Further work is required in this regard, but it may be that infants, in particular, initially perceive durational cues simply as patterns in the amplitude envelope rather than with respect to expected timing of particular segments.

4.3. Structural timing cues in conversational interaction

For lengthening effects to be perceived and interpreted as cues to structure, the working hypothesis is that listeners compare observation with expectations, and interpret deviations therefrom as communicatively significant. Studies discussed above indicate that foregoing speech rate variation can influence listeners' interpretation of local timing (Dilley & Pitt, 2010; Reinisch, Jesse & McQueen, 2011), but the mechanism by which this is achieved remains unclear. Consideration of the prediction of turn-taking in conversational interaction may be useful in this regard, not least because lengthening is well established as cue to an upcoming boundary (Price et al., 1991).

Conversational turn-taking tends to adhere to a "minimal gap, minimal overlap" principle across languages (Stivers et al., 2009); thus, smooth turn-taking relies on listeners' anticipation of the termination of a speaker's contribution. At first sight, this suggests a tension: interaction is facilitated by predictability, whilst speech is only informative insofar as the signal is not predictable for listeners. A promising approach to resolving this paradox, both for turn-taking in general and for the interpretation of timing effects in particular, may lie in theories that consider the mutual entrainment of listener and speaker in conversation. Wilson and Wilson (2005) proposed specifically that entrainment, mediated by the rate of syllable production of the current speaker, arises in endogenous neural oscillators in the brains of the speaker and hearer and that

these oscillators modulate the readiness of interlocutors to initiate speech acts. Scott, McGettigan, and Eisner (2009) elaborated on this proposal to suggest a specific role for sensorimotor circuits in entraining to speech rate and "rhythm". Of course, as Cummins (2012) argued, speech is rarely periodic, and the relationship between signal and syllable flow is complex; thus, to hypothesise mutual entrainment of neural oscillators mediated by speech rate raises as many problems as it purports to solve. To this end, cross-linguistic differences in the coordination of turn-taking (Stivers et al., 2009) may be informative about the types of prosodic information utilised in entrainment. Although such research avenues remain largely unmapped at present, they may lead to a better understanding of the predictive mechanisms through which prosodic timing effects can be interpreted.

5. Summary

The functional framework represents a parsimonious synthesis of foregoing research into speech timing, in which prosodic rather than metrical constituents are the structural determinants of speech timing. Within these domains, timing is only related to structure at domain heads and domain edges, the loci of specific lengthening effects. Outside these loci, duration is determined by the interaction of speech rate and intrinsic articulatory factors arising from the nature of the segmental string and the organisation of segments into syllables.

Much work remains to be done to determine the combination of factors which contribute to the construction of prosodic structure within speech. The evidence presented here strongly suggests, however, that the durational consequences of prosodic structure are not distributed throughout the speech string, but localised at significant points, with characteristic loci, and expressed through lengthening, but not shortening. In this way, the extraction of useful information from the diverse

influences on the duration of speech sounds appears a more tractable problem for the listener.

Acknowledgements

To follow.

References

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Albin, D.D., & Echols, C.H. (1996). Stressed and word-final syllables in infant-directed speech. *Infant Behavior and Development*, 19, 401-418.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*, 46-63.

Arvaniti, A. & Rodriquez, T. (2013). The role of rhythm class, speaking rate and F0 in language discrimination. *Laboratory Phonology*, *4*, 7-38.

Beach, C.M. (1991). The interpretation of prosodic patterns at points of syntactic ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, *30*, 644-663.

Beckman, M. E. (1992). Evidence for speech rhythms across languages. In Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (eds.), *Speech Perception, Production and Linguistic Structure* (pp. 457-463). Oxford: IOS Press.

Beckman, M.E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston & M.E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. (pp. 152-178). Cambridge: Cambridge University Press.

Beckman, M.E., & Edwards, J. (1992). Intonational categories and the articulatory control of duration. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sapisaka (eds.), *Speech Perception, Production and Linguistic Structure* (pp. 356-375). Tokyo: Ohmsha.

Berkovits, R. (1994). Durational effects in final lengthening, gapping, and contrastive stress. *Language and Speech*, *37*, 237–50.

Bolinger, D. (1965). *Forms of English: Accent, Morpheme, Order*. Cambridge, Massachusetts: Harvard University Press.

Brown, W. (1911). Studies from the psychological laboratory of the University of California. XVI. Temporal and accentual rhythm. *Psychological Review, 18,* 336-346.

Bye, P. (1997). A generative perspective on 'overlength' in Estonian and Saami. In I. Lehiste & J. Ross (Eds.), *Estonian Prosody: Papers from a Symposium*. (pp. 36–70). Tallinn: Institute of Estonian Language.

Byrd, D., Lee, S., Riggs, D., & Adams, J. (2005). Interacting effects of syllable and phrase position on consonant articulation. *Journal of the Acoustical Society of America*, *118,* 3860–3873.

Byrd, D., & Riggs, D. (2008). Locality interactions with prominence in determining the scope of phrasal lengthening. *Journal of the International Phonetic Association, 38,* 187-202.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31,* 149-180.

Cambier-Langeveld, T. (2000). *Temporal marking of accents and boundaries*. PhD dissertation, University of Amsterdam, Holland Institute of Generative Linguistics, Netherlands Graduate School of Linguistics.

Campbell, W.N., & Isard, S.D. (1991). Segment durations in a syllable frame. *Journal of Phonetics*, *19,* 37-47.

Cho, T., McQueen, J., & Cox, E. (2007). Prosodically driven detail in speech processing: the case of domain-initial strengthening in English. *Journal of Phonetics*, *35*, 210-243.

Cho, T. & Keating, P.A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29,* 155-190.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Christophe, A., & Dupoux, E. (1996). Bootstrapping lexical acquisition: the role of prosodic structure. *The Linguistic Review, 13,* 383-412.

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

Couper-Kuhlen, E. (1986). *An Introduction to English Prosody*. London: Edward Arnold.

Couper-Kuhlen, E. (1993). *English Speech Rhythm*. Amsterdam: John Benjamins.

Cumming, R. (2011). The effect of dynamic fundamental frequency on the perception of duration. *Journal of Phonetics, 39,* 375-387.

Cummins, F. (1999). Some lengthening factors in English speech combine additively at most rates. *Journal of the Acoustical Society of America, 105,* 476-480.

Cummins, F. (2009). Rhythm as entrainment: the case of synchronous speech. *Journal of Phonetics, 37,* 16-28.

Cummins, F. (2011). Periodic and aperiodic synchronization in skilled action. *Frontiers in Human Neuroscience, 5*: 170.

Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Frontiers in Psychology, 3*: 364.

Cummins, F. & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26,* 145-171.

Cutler, A. (1990). From performance to phonology: Comments on Beckman and Edwards's paper. In J. Kingston & M.E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. (pp. 208-214). Cambridge: Cambridge University Press.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11,* 51-62.

Davis, M.H., Marslen-Wilson, W.D., & Gaskell, M.G. (2002). Leading up the lexical garden-path: segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 218-244.

Dilley, L.C. & Pitt, M.A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*, 1664-1670.

Dimitrova, S., & Turk, A. (2012). Patterns of accentual lengthening in English four-syllable words. *Journal of Phonetics, 40*, 403-418.

Edwards, J., Beckman, M.E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, *89*, 369-382.

Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. *Attention, Perception, & Psychophysics*, 72, 2274-2288.

Fant, G., Kruckenberg, A. & Nord, L. (1991). Durational correlates of stress in Swedish, French and English. *Journal of Phonetics, 19*, 351-365.

Fletcher J. (2010). The prosody of speech: timing and rhythm. In Hardcastle, W., Laver J., & Gibbon F. (Eds.), *The Handbook of Phonetic Sciences* (pp. 523-602). Chichester: Wiley-Blackwell.

Fougeron, C. & Keating, P.A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101*, 3728-3740.

Fourakis, M., & Monahan, C.B. (1988). Effects of metrical foot structure on syllable timing. *Language and Speech*, *31*, 283-306.

Fowler, C. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica, 38*, 35-50.

Fowler, C.A. (1990). Lengthenings and the nature of prosodic constituency: Comments on Beckman and Edwards's paper. In J. Kingston & M.E. Beckman (Eds.), *Papers in Laboratory*

Phonology I: Between the Grammar and the Physics of Speech. (pp. 201-207). Cambridge: Cambridge University Press.

Frota, S., D'Imperio, M., Elordieta, G., Prieto, P. and Vigário, M. (2007). The phonetics and phonology of intonational phrasing in Romance. In P. Prieto, J. Mascaró & M.-J.Solé (eds.), *Segmental and Prosodic issues in Romance Phonology* (pp. 131-153). Amsterdam: John Benjamins.

Fry, D.B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27,* 765-768.

Gaitenby, J.H. (1965). The elastic word. *Technical Report SR-2, Haskins Laboratories, New York.*

Gow, D.W. & Gordon, P.C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 344-359.

Gussenhoven, C. (2002). Intonation and interpretation: phonetics and phonology. In *Proceedings of Speech Prosody, Aix-en-Provence.*

Jones, M.R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96,* 459-491.

Keating, P. A., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.). *Papers in Laboratory Phonology 6* (pp. 145-163). Cambridge: Cambridge University Press.

Kim, E., & McAuley, J.D. (2013). Effects of pitch distance and likelihood on the perceived duration of deviant auditory events. *Attention, Perception & Psychophysics, 75,* 1547-1558.

Kim, H., & Cole, J. (2005). The stress foot as a unit of planned timing: evidence from shortening in the prosodic phrase. *Interspeech 2005: Proceedings of the 9th European Conference on Speech Communication and Technology, Lisbon* (pp. 2365-2368).

Klatt, D.H. (1974). The duration of [s] in English words. *Journal of Speech and Hearing Research, 17*, 51-63.

Klatt, D.H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics, 3*. 129-140.

Klatt, D.H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59*, 1208-1220.

Klatt, D.H. & Cooper, W.E. (1975). Perception of segment duration in sentence contexts. In A. Cohen & S. Nooteboom (Eds.), *Structure and Process in Speech Perception* (pp. 69-89). Heidelberg: Springer Verlag.

Knight, S. (2013). An investigation of passive entrainment, prosociality and their potential roles in persuasive oratory, University of Cambridge PhD dissertation.

Kochanski, G., Grabe, E., Coleman, J. & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America, 118*, 1038-1054.

Kohler, K.J. (2003). Domains of temporal control in speech and language: From utterance to segment. In M.J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences, Barcelona* (pp. 7-10).

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Lee, H. & Seong, C. (1996). Experimental phonetic study of the syllable duration of Korean with respect to the positional effect. *Fourth International Conference on Spoken Language Processing*, (pp. 1193-1196).

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America, 51*, 2018-2024.

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America, 54*, 1228-1234.

Lehiste, I. (1975). The role of temporal factors in the establishment of linguistic units and boundaries. In W.U. Dressler & F.V. Mares (Eds.), *Phonologica 1972* (pp. 115-122). Munich-Salzburg: Verlag.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5*, 253-263.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W.J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403-439). Amsterdam: Kluwer.

Lindblom, B. (1996). Role of articulation in speech perception: clues from production. *Journal of the Acoustical Society of America, 99*, 1683-1692.

Marslen-Wilson, W.D., Tyler, L.K., Warren, P., Grenier, P., & Lee, C.S. (1992). Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology*, 45, 73-87.

Mattys, S.L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics*, 62, 253-265.

Mattys, S.L., White, L., & Melhorn, J.F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134, 477-500.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review, 83*, 405.

Nakatani, L.H., O'Connor, K.D., & Aston, C.H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica, 38*, 84-106.

Nespor, M. & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.

Nolan, F., & Asu, E.L. (2009). The pairwise variability index and coexisting rhythms in language. *Phonetica, 66*, 64-77.

Nolan, F. & Jeon, H. (under revision). Speech rhythm: A metaphor? *Proceedings of the Royal Society B.*

O'Dell, M. L., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco* (pp. 1075–1078).

O'Dell, M. L., & Nieminen, T. (2009). Coupled oscillator model for speech timing: Overview and examples. In M. Vainio, R. Aulanko & O. Aaltonen (Eds.), *Nordic Prosody: Proceedings of the 10th Conference, Helsinki 2008* (pp. 179-190).

Oller, D.K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America, 54*, 1235-1247.

Ortega-Llebaria, M. & Prieto, P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró & M.-J.Solé (Eds.), *Segmental and Prosodic issues in Romance Phonology* (pp. 155-176). Amsterdam: John Benjamins.

Pisoni, D.B. (1976). Fundamental frequency and perceived vowel duration. *Journal of the Acoustical Society of America, 59*, S39.

Pointon, G.E. (1980). Is Spanish really syllable-timed? *Journal of Phonetics, 8*, 293-304.

Port, R.F. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America, 69*, 262-274.

Port, R.F. (2003). Meter and speech. *Journal of Phonetics, 31*, 599-611.

Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*, 2956-2970.

Prieto, P., Vanrell, M. D. M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication*, *54*, 681-702.

Quené, H. (1992). Durational cues for word segmentation in Dutch, *Journal of Phonetics*, *20*, 331-350.

Rakerd, B., Sennett, W., & Fowler, C. A. (1987). Domain-final lengthening and foot-level shortening in spoken English. *Phonetica*, *44*, 147-155.

Raphael, L.J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in English. *Journal of the Acoustical Society of America*, *51*, 1296-1303.

Reinisch, E., Jesse, A., & McQueen, J.M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 978-996.

Remijsen, B. & Gilley, L. (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics*, *36*, 318-344.

Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (ed.) *Linguistic controversies* (pp. 73-79), London: Edward Arnold.

Saffran, J.R., Newport, E.L., & Aslin, R.N. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language*, *35*, 606-621.

Salverda, A.P., Dahan, D., & McQueen, J.M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51-89.

Scott, D.R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, *71*, 996-1007.

Scott, D.R. Isard, S.D., & Boysson-Bardies, B. (1985). Perceptual isochrony in English and French. *Journal of Phonetics, 13*, 155-162.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action – candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience, 10*, 295-302.

Selkirk, E.O. (1996). The prosodic structure of function words. In J.L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 187-213). Mahwah, New Jersey: Lawrence Erlbaum.

Shattuck-Hufnagel, S. & Turk, A.E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*, 193-247.

Shen, Y., & Peterson, G.G. (1962). *Isochronism in English*. Studies in Linguistics, Occasional Papers 9, University of Buffalo.

Sluijter, A.M.C. (1995). *Phonetic Correlates of Stress and Accent*. University of Leiden PhD dissertation.

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech, Language and Hearing Research, 37*, 831.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymanna, G., Rossanoa, F., de Ruitera, J.P., Yoon, K., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106*, 10587-10592.

Suomi, K. (2007). On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics, 35*, 40–55.

Suomi, K. (2009). Durational elasticity for accentual purposes in Northern Finnish. *Journal of Phonetics, 37*, 397-416.

Suomi, K., Meister, E., Ylitalo, R., & Meister, L. (2013). Durational patterns in Northern Estonian and Northern Finnish. *Journal of Phonetics, 41*, 1-16.

Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Articulatory results. *Journal of the Acoustical Society of America*, 113, 2834-2849.

Tagliapietra, L., & McQueen, J. M. (2010). What and where in speech recognition: Geminates and singletons in spoken Italian. *Journal of Memory and Language, 63*, 306-323.

Turk, A.E, & Sawusch, J.R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics, 25*, 25-41.

Turk, A.E, & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics, 28*, 397-440.

Turk, A.E, & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics, 35*, 445-472.

Turk, A.E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics, 27*, 171-206.

Van Lancker, D., Kreiman, J. & Bolinger, D. (1988). Anticipatory lengthening. *Journal of Phonetics, 16*, 339-347.

van Santen, J.P.H. (1992). Contextual effects on vowel duration. *Speech Communication, 11*, 513-546.

van Santen, J.P.H. (1997). Segmental duration and speech timing. In Y. Sagisaka, N. Campbell & N. Higuchi (Eds.), *Computing Prosody: Computational Models for Processing Spontaneous Speech* (pp. 225-249). New York: Springer-Verlag.

White, L. (2002). *English speech timing: A domain and locus approach*. University of Edinburgh PhD dissertation (http://www.cstr.ed.ac.uk/projects/eustace/dissertation.html).

White, L., Floccia, C., Goslin, J., & Butler, J. (in press). Utterance-final lengthening is predictive of infants' discrimination of English accents. *Language Learning*.

White, L., & Mattys, S.L. (2007a). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*, 501-522.

White, L., & Mattys, S.L. (2007b). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró & M.-J.Solé (Eds.), *Segmental and Prosodic issues in Romance Phonology. Current Issues in Linguistic Theory series* (pp. 237-257). Amsterdam: John Benjamins.

White, L., Mattys, S.L., Stefansdottir, L., & Jones, V. (2014, to appear). Lengthened consonants are interpreted as word-initial. In *Proceedings of Speech Prosody, Dublin*.

White, L., Mattys, S.L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language, 66*, 665-679.

White, L., Payne, E., & Mattys, S.L. (2009). Rhythmic and prosodic contrast in Venetan and Sicilian Italian [pdf file]. In M. Vigario, S. Frota & M.J. Freitas (Eds.), *Phonetics and Phonology: Interactions and Interrelations* (pp. 137-158). Amsterdam: John Benjamins.

White, L., & Turk, A.E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics, 38*, 459-471.

Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America, 91*, 1707-1717.

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review, 12*, 957-968.

Xu, Y. (2006).Speech prosody as articulated communicative functions. In *Proceedings of Speech Prosody, Dresden* (p. SPS5-4-218).

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics, 38*, 329-336.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*, 159-197.

Tables

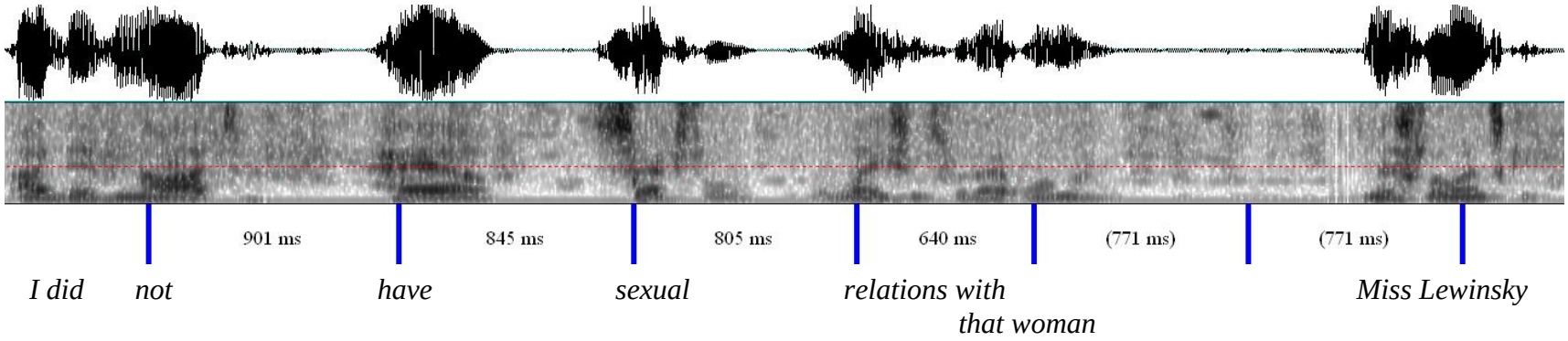
Table 1. Principles of the functional framework of prosodic timing effects.

No privileged timing unit	There is no unit into which an utterance may be exhaustively parsed that consistently imposes timing constraints upon its subconstituents.
Prosodic timing domains	Structural influences on speech timing relate to the organisation of syllables into words and higher-level constituents.
Localised lengthening effects	Prosodic structure influences speech timing through localised lengthening effects at domain edges and domain heads.
Structurally-defined loci	At each domain edge or domain head, segments are lengthened within a structurally-defined locus. These loci are distinct from one another.
No temporal mediation outside loci	There is no direct, consistent relationship between prosodic structure and segment duration apart from at domain edges and domain heads.

Figure captions

Figure 1. Bill Clinton waveform and spectrogram. The onset points of stressed vowels are indicated with vertical bars, with durations of the intervals between onsets shown between the bars. The penultimate beat is a notional "silent beat" bisecting the inter-stress interval in ***wo****man*… *Miss Le****win****sky*: the two intervals thus created are 771 ms long, close to the mean duration, 798 ms, of the earlier intervals.

Figure 1.


901 ms
845 ms
805 ms
640 ms
(771 ms)
(771 ms)
I did not have sexual relations with that woman Miss Lewinsky