

# Chapter 1

## The domain and locus in speech timing

### 1.1 Introduction

Speech timing is a potentially ambiguous guide to speech structure. There are a large number of influences on the duration of segments within a speech string, some of which are a result of factors which are not fundamentally durational, such as biomechanical constraints on speech production. Evidence presented in this dissertation suggests, however, that certain suprasyllabic timing processes are directly related to linguistic structure within particular domains; furthermore, each process has a characteristic locus within its domain. Because the loci are distinct, the apparent ambiguity in speech timing is reduced.

The descriptive framework incorporating the concepts of the domain and the locus, representing cause and effect in speech timing, is described below, where a characterisation of domain-edge and domain-span processes is abstracted from a description of lengthening and shortening effects. The experimental work reported here focuses on domain-edge and domain-span processes at the word level and the utterance level. The results suggest a model of suprasyllabic speech timing based upon localised lengthening effects at domain edges and in pitch-accented words, each process having a particular phonologically-defined locus. There is little empirical support for domain-span processes, in contrast with approaches to speech timing that suggest that suprasyllabic linguistic structure has durational consequences throughout the speech string.

### 1.1.1 Lengthening and shortening

Much of the systematic variation in English speech timing can be described either in terms of constituent structure or in terms of the distribution of prominences. Within both frames of reference, relatively large increases in duration tend to be associated with intonational events: constituents that have an intonational marking at their final edge—a boundary tone—also tend to show lengthening of segments near the boundary (for example: Price *et al.* 1991); likewise, the phrase-level prominence indicated by a pitch accent is associated with significant lengthening of the accented syllable and neighbouring syllables within the word (Sluijter 1995; Turk & Sawusch 1997; Turk & White 1999).

The timing consequences of constituent structure and the distribution of prominences are not restricted to the locations of intonational events, however: at lower levels, both prominence and constituent boundaries may be marked primarily by duration. With regard to prominence, segments are longer in lexically-stressed syllables than in unstressed syllables; furthermore, a stressed syllable when followed immediately by an unstressed syllable may be shorter than when followed by another stressed syllable (for example: Van Lancker *et al.* 1988). With regard to constituents, both initial and final boundaries may be indicated by lengthening: syllable onset consonants are longer word-initially than word-medially (for example: Oller 1973; Cooper 1991); the final syllable rhyme of words and phrases may be lengthened even in the absence of an intonational phrase boundary (for example, Beckman & Edwards 1990; Wightman *et al.* 1992). Additionally, the duration of the constituents of a lexical word may be influenced by its length: “polysyllabic shortening” refers to the hypothesised inverse relationship between primary stressed syllable duration and the number of syllables in the word (for example: Lehiste 1972; Port 1981).

The timing consequences of such effects when investigated experimentally are described by comparing segmental durations at two or more levels of some condition, with the many other potential sources of durational variation kept constant as far as possible. As there is no procedure for fixing an objective reference duration for segments of any given type, any comparison between segments of different durations may be described either as “lengthening” or “shortening”, but the conventional usage of these terms reflects an apparent distinction of processes. Typically, processes associated with constituent boundaries are described as lengthening, relative to the duration of segments placed constituent-medially, for example:

**Word-initial lengthening:** /p/ is longer in *porter* than in *report*.

**Phrase-final lengthening:** /æk/ is longer in  
*The big cat was black, but the small one was grey.*

than in

*The black cat was big, but the grey one was small.*

The effects of prominence, particularly phrasal stress, are also described in terms of lengthening relative to the absence of prominence. For example:

**Accentual lengthening:** /sæmən/ is longer in

*I said salmon sandwich, not tuna sandwich.*

than in

*I said salmon sandwich, not salmon salad.*

In contrast, durational variation resulting from the phonological length (typically, the number of syllables) of the constituent in which measured segments are placed is generally described as shortening: thus, the duration in the phonologically-shortest constituent is the reference point<sup>1</sup>. A stressed syllable is held to have greatest duration in a monosyllabic word or utterance, and to become shorter as more syllables are added within the constituent. For example,

**Polysyllabic shortening:** /stɪk/ is longest in *stick*, shorter in *sticky*

and shorter still in *stickiness*.

Similarly, variation in stressed syllable duration due to the number of following unstressed syllables is described as foot-level shortening (for example: Rakerd *et al.* 1987): the duration of the first of a pair of adjacent monosyllables is greater than when one or more unstressed syllables intervene.

**Foot-level shortening:** /pɪtʃ/ is longer in *peach light* than in *peach delight*.

Both constituents and prominences are thus associated with “lengthening” and with “shortening”.

### 1.1.2 Cause and effect in speech timing

There is a large body of experimental work examining durational processes, reviewed in detail in Chapter 2; however, due to the apparently hierarchical organisation of speech, precise qualitative characterisation of the cause and the effect of timing processes can be experimentally elusive. For example, in order to ascertain that word

---

<sup>1</sup>Here and subsequently, the term “phonological length” is used where clarification is required to mean the size of some constituent measured in terms of some phonological unit, generally the syllable. This is in distinction to what might be termed “phonetic length”, that is, acoustic duration. As far as possible the terms “length” and “constituent size” are used in this “phonological” sense, and the term “duration” refers to the acoustically-observed quantity; the use of terms such as “lengthening” and “shortening” to refer to contrasts in acoustic duration is unavoidable, however. The additional use of the term “phonological length” to refer to the distinction between tense and lax vowels is acknowledged, but not intended in the following discussion, unless explicitly indicated.

length influences segment duration, it is necessary to keep constant the length of other constituents which include that word. If the lengths of a number of nested constituents vary at the same time, it cannot be said for certain where the influence lies. Furthermore, the problem of qualitative characterisation applies not only to the constituent which causes the durational variation, but also to the section of speech which manifests the effect: for instance, if the effect of phrase-finality is reported in terms of lengthening of a syllable or word or other arbitrary unit, information about the distribution of the effect within or beyond that unit will be missed.

This imprecision in the characterisation of timing processes, particularly in the description of the location of the effect, would matter less if such processes were the result of diffuse changes in speech rate: for example, Klatt (1976) and Cummins (1999) suggest that utterance-final lengthening is a gradient process, reflecting “global deceleration and a reduction in articulatory effort over a number of syllables” Cummins (1999:476). Similarly, lengthening associated with the presence of a pitch accent could be interpreted as a by-product of the fundamental frequency movement, with segments lengthened in order to provide sufficient “ground” for the realisation of a perceptible pitch excursion (for example: Beckman & Edwards 1992). In this type of analysis, the distribution of lengthening amongst segments in the vicinity of an utterance boundary or pitch accent depends upon factors such as the expandability of individual segments and the relative sharpness of deceleration or fundamental frequency movement.

In contrast, the working hypothesis adopted here is that many speech timing effects above the level of syllabic organisation are directly related to linguistic structure rather than indirect effects of other processes. A systematic description of durational variation in speech is proposed here in terms of two structures defined in phonological terms: the “domain” and the “locus”, the former characterising the cause and the latter describing the effect. For example, lengthening is observed at the end of intonational phrases, relative to utterance-medial position: thus, the domain of lengthening is the intonational phrase and the locus—that is, the stretch of speech which undergoes lengthening—is the rhyme of the phrase-final syllable.

The domain-and-locus approach is applied in the present research to durational variation arising from structural relationships within speech above the level of syllabic organisation. The suprasyllabic level of description is distinguished from the segmental level and the syllabic level in Section 1.2, where the treatment of speech rate, and the distinction between global and local rate variation is also discussed. The rationale for the domain-and-locus description of durational variation in speech is discussed further in Section 1.3, where a classification of durational effects in domain-and-locus terms is presented. The specific research questions addressed in this dissertation and

an overview of the theoretical and empirical work undertaken are presented in Section 1.4.

## 1.2 Sources of durational variation

This dissertation examines the durational consequences for segments of the organisation of syllables into words and larger constituents: that is, suprasyllabic factors rather than segmental and syllabic factors. These distinctions are illustrated by classifying Klatt's (1976) list of durational factors in English speech—which he organises according to their possible origin: syntax, semantics, physiology, etc.—into segmental, syllabic and suprasyllabic factors<sup>2</sup>. The other factor mentioned by Klatt, speech rate, is also discussed.

### 1.2.1 Segmental factors

Segmental durational factors are those which are intrinsic to particular phonemes or classes of phonemes. Klatt identifies a number of such factors: the difference between long and short vowels; the greater duration of voiceless fricatives than voiced fricatives; the greater duration of bilabial stops than alveolar and velar stops. There are likely to be articulatory explanations for the origin of these distinctions: for example, short vowels are more centralised and thus their articulation takes less time; likewise, a bilabial stop creates a larger supralaryngeal cavity than alveolar or velar stops, thus more time is required for sufficient pressure to build-up behind the point of constriction to create an audible release. Although influenced by such articulatory considerations, these distinctions may in many cases be part of the phonology of the language.

Klatt also suggests another aspect of segmental duration which must be considered. He observes that, when durational factors are expressed as shortening rules, the co-application of two shortening rules is less than the sum of the two effects separately. To account for this, he proposes that a certain minimum duration is “required to execute a satisfactory articulatory gesture”, thus the segment is incompressible beyond a certain amount. The mathematical details of models of timing incorporating the concept of incompressibility are not relevant to the present discussion, but there are perceptual and articulatory reasons for believing that segments will indeed exhibit some degree of incompressibility; furthermore, this is likely to vary between phonemes of different types. As discussed in Chapter 2, not all phonemes may manifest a particular durational effect to the same degree.

---

<sup>2</sup>This classification is intended to illustrate the distinctions between segmental, syllabic and subsyllabic factors rather than being an exhaustive survey of the types of processes at each level of description.

### 1.2.2 Syllabic factors

Syllabic durational factors are those which result from the organisation of segments into a string of syllables. One of the most perceptually salient of the syllabic factors identified by Klatt is the lengthening of vowels in stressed syllables compared with unstressed syllables; onset consonants are also longer in stressed syllables. Klatt also suggests that there is a small durational difference between vowels in syllables carrying primary and secondary lexical stress, although, as discussed below, this may only be present in phrasally-stressed words. The durational difference between stressed and unstressed syllables is clearly a linguistically-motivated distinction: greater duration is conferred as a result of the stress value of the syllable as determined by morphology rather than by articulatory considerations. Klatt also notes that the durational difference between stressed and unstressed syllables is greatest in phrase-final position, indicating an interaction between syllabic and suprasyllabic levels.

Within and between syllables, there are durational effects arising from the number and type of adjacent segments. Consonants are shorter in clusters than when they occur singly, this difference being more marked in stressed syllables: whether consonants must be tautosyllabic to manifest shortening in clusters is uncertain. Vowels are longer when followed by voiceless coda consonants than when followed by voiced coda consonants<sup>3</sup>, a difference that Klatt claims is, like the stressed vs unstressed syllable durational difference, much greater in phrase-final syllables than in phrase-medial syllables. Indeed, he says that the duration of the preceding vowel can serve in phrase-final position as a primary perceptual cue to the distinction between voiced and voiceless consonants in the syllable coda.

### 1.2.3 Suprasyllabic factors

Suprasyllabic durational factors are those that arise from the linguistic structure of a syllabified string. Klatt identifies a number of effects which may be classified as suprasyllabic, and which can be placed into three categories: boundary-related lengthening, lengthening due to prominence, and shortening due to the phonological size of the constituent.

There are two effects that Klatt identifies as primary perceptual cues, on the basis of their perceptual salience and linguistic significance: phrase-final lengthening, and lengthening due to emphatic or contrastive stress. The type of phrase which occasions final lengthening is discussed in detail in Chapter 2; as Klatt observes there is a lack of consistent correspondence between syntactic phrasing and phrase-final lengthening. Additional boundary effects proposed include word-initial and word-final lengthening.

---

<sup>3</sup>This difference is less marked in some dialects, particular in Scottish English, where it may be obscured by larger dialect-specific allophonic durational effects.

ing, utterance-final lengthening extending over several syllables, and paragraph-final lengthening, whereby the final sentence of a read paragraph tends to be longer than if it were non-final. As well as the effects of emphatic or contrastive stress, which may be related to the presence of a pitch accent within the phrasally-stressed word, Klatt also suggests that entire phrases may be emphasised by lengthening: for example, at a dramatic point in the relation of a story, a speaker may reduce speech rate across a whole phrase or sentence.

There are two constituents for which Klatt proposes an inverse relation between constituent size and subconstituent duration, the word and the utterance. Initial stressed syllables are shorter in polysyllabic words: for example, /spid/ is shorter in *speedy* than in *speed*, and shorter still in *speediness*. This can either be interpreted as word-final lengthening or as a result of what he calls: “a desire on the part of a talker to maintain a nearly constant duration for each word, independent of its phonetic composition” Klatt (1976:1214). Klatt appears not to favour the latter explanation, continuing: “it is hard to believe that constant word duration is a very important constraint in English, since duration has already been implicated as functioning to differentiate inherently long from short vowels, voiced from voiceless fricatives, phrase-final from non-final syllables, etc.” With regard to the effect of utterance length, Klatt claims that there is a general trend for speech rate to become faster in longer sentences, thus shortening individual segments within the sentence.

Previous experimental studies of suprasyllabic durational processes related to boundaries, constituent length and prominence are reviewed in Chapter 2, including the durational effects associated with variation in the size of hypothesised prominence-delimited constituents. Experimental investigations of such processes at the word-level and the utterance-level are presented in Chapter 3 and Chapter 4.

#### 1.2.4 Speech rate

In addition to the durational factors classified here as segmental, syllabic and suprasyllabic, Klatt also observes that changes in overall speech rate, which may be affected by a number of non-linguistic factors, “exert a complex influence on the durational patterns of a sentence. When speakers slow down, a good fraction of the extra duration goes into pauses (Goldman-Eisler 1968). On the other hand, increases in speaking rate are accompanied by phonological and phonetic simplifications as well as differential shortening of vowels and consonants” Klatt (1976:1210).

It is worth considering the concept of speech rate further, as it is an unstated factor in much of the experimental work reviewed in Chapter 2. The term “speech rate” may be used in two senses. Most commonly, it is used to mean *global* speech rate: the rate of production of speech segments over a stretch of speech of arbitrary length as

influenced by factors such as physiology, dialect and emotional arousal. Occasionally it refers to *local* speech rate: the influence of linguistic structure, such as the distribution of constituents and prominences, on speech timing (it is used in this sense by, for example, Dankovičová 1997). Clearly, the latter usage relates to the linguistically-motivated durational patterns that are the present objects of study, so to prevent confusion, the term “speech rate” is only used here in the global sense.

In the discussion of previous research and in the experiments presented here, speech rate is taken to be normal except where otherwise stated, where “normal” implies a self-selected rate by the speaker, possibly under explicit instruction to speak normally. In a small study, Gaitenby (1965) compares segmental duration between subjects talking at their own “normal” rate and finds that most segments are expanded or compressed in proportion to the length of the whole utterance; voiced consonants, however, are very close in absolute duration between talkers. She also finds that the widest variations in length are observed on stressed syllables.

It is not certain whether patterns observed between speakers at their individual natural speaking rates also apply to the different rates of a single speaker. In particular, it is possible that the results obtained at normal speech rate with respect to suprasegmental effects may not generalise across the range of possible rates.

Global speech rate variation does have two well-attested consequences. Firstly, as described by Gee & Grosjean (1983), at faster speech rates utterances tend to be broken up into fewer and larger phrases, a process discussed further in Chapter 2. Secondly, researchers who look at the effect of global speech rate variation find that some of the durational effects evident at normal speech rates may be more marked at slower rates and may be attenuated or disappear at faster rates. Rakerd *et al.* (1987), for example, find significant interactions between speech rate and syntactic structure, and between speech rate and foot structure: preboundary lengthening is attenuated in fast speech, as is the shortening of a stressed syllable followed by an unstressed syllable compared with when followed by another stressed syllable. Inspection of their graphical data suggests that phrase-final lengthening is preserved to a reasonable extent at higher speech rate, whereas the difference in stressed syllable duration between monosyllabic and disyllabic feet is greatly reduced in fast speech. The evidence that Beckman & Edwards (1990) find for a phrase-medial word-final lengthening effect is stronger at slower speech rates. In contrast, Port (1981) finds that the rhyme of a pitch-accented syllable shows less shortening in response to an increase in global speech rate than does the overall sentence: at a faster rate the accented syllable rhyme is 76% of its normal duration whereas the sentence as a whole shortens to 72% of its normal duration. Port reports that this result is in agreement with an earlier study by Peterson & Lehiste (1960).

It appears that some durational effects are more robust under increases in speech rate than others. It is possible that rephrasing as observed by Gee & Grosjean (1983) could underlie some of these observations: results that appear to reflect a diminution of final lengthening may actually be a result of some boundaries not being realised at all at faster rates. This explanation could not account for the attenuation of shortening in disyllabic feet at fast rate, however, unless the second of a pair of adjacent stressed syllables became unstressed—and thus refooted—at fast rate. A speculative interpretation is that the durational effects associated with intonational events are more robust in fast speech than other timing processes: thus, intonational-phrase-final lengthening and accentual lengthening are preserved whereas final lengthening of minor constituents and the effects of stress adjacency are attenuated.

## 1.3 A descriptive framework

### 1.3.1 The concept of the domain

The concept of the domain is a familiar one in prosodic phonology—the study of connected speech processes—as presented, for example, by Selkirk (1986) and Nespor & Vogel (1986). Within the hierarchies of prosodic constituents proposed by these and other researchers, each prosodic constituent is proposed to have a certain relation to syntax, which is in general more straightforward for lower-level constituents: at higher levels, a number of arrangements of prosodic constituents are available for the same syntactic structure. The evidence for constituents is provided by segmental processes occurring within and between words, such as flapping in American English and r-insertion in non-rhotic dialects of British English. There are three types of rules proposed by Nespor & Vogel to explain the occurrence of segmental variations in connected speech: domain-span rules, domain-juncture rules and domain-limit rules. Domain-span rules describe processes which may occur anywhere within a particular constituent but not at its edge; domain-limit and domain-juncture rules occur only at the edges of constituents, with the latter further requiring particular segmental conditions on the other side of the constituent boundary.

These prosodic constituents and segmental processes are described in more detail in Chapter 2; for the purposes of the present discussion there are two important points to be made about the distinction between prosodic constituents and the concept of the domain as used here.

Firstly, the constituents described by researchers such as Selkirk and Nespor & Vogel are not necessarily co-extensive with the domains of durational effect, although they may prove to be so. Nespor & Vogel, for example, claim that some of the constituents that they identify as domains of segmental processes may also be the do-

mains of durational variations, but do not present empirical evidence to support these claims. Furthermore, although it seems clear that syntax is not wholly reliable in predicting the occurrence of durational effects such as final lengthening, as discussed in Chapter 2, it remains to be seen whether the prosodic constituents evinced by segmental processes are in all cases isomorphic with the domains of durational processes. There is indeed, no intrinsic reason why a domain need be a constituent at all, although the fact that a constituent serves as the domain of some effect in speech may be used as evidence of its existence. An illustration of a domain which is not co-extensive with a constituent might be taken from intonational phonology: Faure *et al.* (1980), for example, suggest that the phonetic realisation of stress contrasts is different following the nuclear accent in an intonational phrase. One might thus assert that the stretch of speech between the nuclear accent and the intonational phrase boundary is the “domain of stress neutralisation”. Such a statement may be theoretically sufficient without the requirement that this domain be identified as a constituent of some kind.

The second important distinction to be made between the concept of the domain as exemplified in the prosodic phonology of Nespor & Vogel (1986) and its use here is in the type of processes manifested within the domain. As described above, durational variation due to the position of measured segments within some constituent tend to be described as lengthening, whereas the effect constituent length tends to be described as shortening. An alternative to the subjective standpoint inherent in a lengthening and shortening description is to describe the relationship between the domain and the measured segments as either a domain-edge process or a domain-span process. These terms and their relation to the domain-limit, domain-juncture and domain-span processes of prosodic phonology are described further in Section 1.3.3.

### 1.3.2 The concept of the locus

Many early studies of durational effects such as final lengthening report the results in terms of the amount of lengthening of some unit, typically the word or syllable, in the experimental condition compared with some control condition, for example, constituent-final syllables as compared with constituent-medial syllables. As discussed above, this could in some cases reflect an implicit assumption that durational variation is not related to linguistic structure, but is instead a consequence of the articulatory process: for example, final lengthening may be seen as a process of deceleration at the end of a speech unit and thus not targeted at any particular segments of speech. More recent evidence suggests that the distribution of durational variation is, in some cases at least, determined by linguistic structure: thus, for example, phrase-final lengthening is manifest on the final syllable rhyme (for example: Wightman *et al.* 1992) and accentual lengthening is distributed throughout the word (Sluijter

1995; Turk & White 1999) and attenuated by word boundaries.

If durational variation is related to linguistic structure, then it becomes necessary to know what the constituents of this structure are. The concept of the domain represents one part of the process: segments may be lengthened when they stand in a particular relation to a particular domain, such as the constituents of syntactic or prosodic structure. An account of linguistically-derived timing also requires a description of where the effect associated with some domain is manifest: this is the locus of durational variation.

One reason for the proposal here of the term locus is the resolution of an ambiguity present in some recent investigations of the distribution of durational variation associated with certain processes. For example, studies of final lengthening, such as Cambier-Langeveld (2000) for Dutch, report that it primarily affects the rhyme of the final syllable before the constituent boundary, citing this as evidence that the “domain” of final lengthening is the syllable rhyme. In customary recent usage, however, the term “domain” is applied to the constituent causing the lengthening effect (for example, Fougeron & Keating 1997). As the domain of final lengthening cannot be at the same time the intonational phrase and the rhyme of the final syllable of that phrase, the additional term “locus” is proposed to describe the phonologically-defined section of speech which is directly subject to a durational process, whether characterised as lengthening or shortening.

A further reason for emphasising the importance of the locus in the description of durational variation is that its identification may serve to clarify the nature of the effect. There are many factors which may influence the duration of a speech segment, outlined in Section 1.2, and as Port (1981:262) observes: “this large and impressive literature [on the durational realisation of linguistic features] has not resulted in an integrated view of the role of linguistically motivated timing in speech perception or speech production [...] Not only may several different phonological features affect the duration of a single phonetic interval, such as a vowel duration, but many other factors that are marginally phonological or nonlinguistic may also influence its duration [such as word length, syntactic boundary location, speech rate ...] How can timing be an effective source of phonological information when it is subject to such a variety of overlapping distortions?” The question of whether morphosyntactic facts such as word length and boundary location are integrated into phonological structure may be left aside for the moment; the important point here is that, given the diversity of influences, speech segment duration may appear an intrinsically ambiguous cue to linguistic structure.

Port observes—like Klatt mentioned above—that limits on the compressibility of speech segments mean that a combination of “shortening factors” have less effect than

the factors individually, he then concludes that: “despite massive evidence that relatively fine-grained articulatory timing comprises an ubiquitous aspect of the acoustic specification of English words, there is, nevertheless, also evidence that in normally complex speech situations very little of this potential information may actually be available for a listener.” Of course, listeners do not experience segmental duration in isolation. As well as the variations in acoustic energy across time and frequency which allow determination of segmental identity, there are variations in the fundamental frequency and amplitude of the speech signal which are known to correlate with durational variation. Thus, greater phrasal prominence is conferred by a combination of duration, fundamental frequency and amplitude variation. Similarly, major boundaries are marked by both duration and fundamental frequency.

Experimental studies show, however, that durational variation alone may serve to indicate the presence of a boundary to the listener in some cases (for example: Price *et al.* 1991; White 1993). Furthermore, there must be sufficient information about linguistically-motivated patterns of speech timing accessible to the listener or the systematic variations which do not arise as a direct consequence of articulation could not be learned. Indeed, it is not clear why such systematic durational variation would be present in speech if it were not accessible to listeners.

Klatt (1976:1220) expresses the apparently intrinsic ambiguity of speech timing thus:

A paradox seems to exist which might prevent listeners from making effective use of durational cues in forming hypotheses about sentence structure. In order, e.g., to perceive the lengthening at a phrase boundary as a cue to the end of a constituent, it seems necessary to know the identity of the lengthened segments. Durations are lengthened relative to the inherent durations for the segments in question. But a listener can't know the inherent duration until he has identified the vowel, so we have a version of “the chicken or the egg” paradox<sup>4</sup>.

At the segmental level, however, the intrinsic durational properties of different phonemes are in most cases not a primary cue to segmental identity in English: like microprosodic variations in fundamental frequency, they may be regarded as a predictable perturbation which can be factored out by listeners when attending to suprasyllabic variation. For example, although different vowels in English have different durations in citation form, differences in vowel quality are more important for identification. Similarly, voiced obstruents tend to be shorter than voiceless obstruents, but duration itself is not a significant cue to voicing in most cases. Likewise, at the syllabic level, much of the variation is likely to be predictable if segmental identities are known.

---

<sup>4</sup>Klatt's paradox has been resolved in modern approaches to speech recognition where multiple hypotheses are processed in parallel. The quotation serves to illustrate potential problems for listeners, however these may be solved in practice.

In some contexts, certain features may be neutralised and duration may play a more important role: for example, Klatt suggests that where voiced fricatives are devoiced syllable-finally, duration may be the primary cue to voicing. This is not, however, simply a matter of judging absolute duration, but of comparing the relative durations of adjacent segments. For example, the interpretation of the word /meɪs/ may depend on both vowel duration and consonant duration: in “mace”, in many dialects of English, the vowel will be relatively short and the coda relatively long; in “maize” the vowel will be relatively long and the coda relatively short. Thus, the distribution of durational variation may sometimes play a role in identifying of segments in connected speech.

At the suprasyllabic level, it is argued in this dissertation that the distribution of variation is of primary importance in the interpretation of speech timing as a cue to structure. Klatt identifies duration as a primary cue in two structural distinctions: phrase-finality vs non-phrase-finality, and the presence or absence of emphatic stress: both phrase-final and emphatically-stressed syllables show substantial lengthening. As observed by Shattuck-Hufnagel & Turk (1996:238): “Apparently, listeners are able to distinguish prominence-related from boundary-related duration lengthening, perhaps because the two factors affect different parts of the syllable (Edwards *et al.* 1991).” As discussed above, the lengthening associated with phrase boundaries appears to affect primarily the syllable rhyme (for example: Wightman *et al.* 1992); as shown by Turk & Sawusch (1997) and Turk & White (1999), the lengthening associated with emphatic accent also affects the accented syllable onset and adjacent syllables within the same word, and may extend further.

Clearly, fundamental frequency variation will also be associated with final lengthening, at least at the end of intonational phrases, and with emphatic stress. Fundamental frequency variation is not always unambiguous, however: as discussed in Chapter 2, pitch accents are not always realised with a large fundamental frequency movement, particularly towards the end of a phrase. Furthermore, some structurally-related durational variation, such as word-initial lengthening and polysyllabic shortening, is apparently not associated with fundamental frequency variation. It is hypothesised here that the ambiguity in speech timing as a cue to linguistic structure is greatly reduced if the locus of durational variation is considered in each case.

### 1.3.3 A domain-and-locus description

The concepts of the domain and the locus provide a descriptive framework for speech timing processes which allows established effects to be treated more systematically. In particular, structurally-determined lengthening may, in many cases, be seen as a domain-edge process and shortening may often be seen as a domain-span process.

As described now, this classification implies processes with different characteristics, particularly with regard to the locus.

### Domain-span processes

The durational effect of constituent length is generally held to be manifest in an inverse relationship between the phonological length of some constituent and the duration of some subconstituent: in other words, there is an inverse relationship between the length of the domain and the duration of the locus. As indicated in Section 1.1, a number of units have been held to be domains of such processes, both syntactic/prosodic constituents such as the word and the sentence/utterance<sup>5</sup>, and prominence-delimited constituents, such as the cross-word foot: evidence for such processes is described in detail in Chapters 2 and 3.

As outlined above, this effect is generally described as “shortening”; the alternative adopted here is to adapt the terminology of prosodic phonology (for example: (Nespor & Vogel 1986)), and describe it as a “domain-span” process. This expresses the fact that the duration of the locus is related to the domain’s length and not to the position of the locus within the domain. This usage of “domain-span” is, however, distinct from its usage by Nespor & Vogel: in prosodic phonology, a domain-span applies within a domain, but not at its edge. As used here with regard to durational processes, the effect may in principle apply to the locus *anywhere* within a domain; the crucial factor is not the position of the locus but the length of the domain.

The primary characteristic of a domain-span process is the inverse relationship between domain length and subconstituent duration. This may be depicted as a compression of the measured subconstituent proportional to the amount of additional material within the domain. Thus, the term “domain-span compression” is sometimes used to clarify the nature of the effect: the term is pleonastic, however, as all domain-span processes are hypothesised to manifest such compression.

There are a number of candidates for the subconstituents that may comprise the locus of a domain-span process. One possibility is that the locus is some privileged unit, such as the head of the constituent that comprises the domain. In a polysyllabic word, for example, the head would be the syllable carrying the primary lexical stress, or the nucleus of that syllable. In a larger prosodic constituent, for example, the intonational phrase, the head would be the syllable or word carrying the main stress accent (nuclear accent). In prominence-delimited constituents, the heads are part of the definition; for example, a foot of almost any type comprises a strong syllable followed by a number of weak syllables.

---

<sup>5</sup>Differences in syntactic and prosodic definitions of the word are discussed in Chapter 2, as is the prosodic concept of the utterance.

A head-based account of domain-span shortening is suggested by studies which report the effect of constituent length on the duration of the constituent head, but this is frequently the *only* unit whose duration is reported, sometimes for practical reasons: for example, in a comparison of monosyllabic, disyllabic and trisyllabic words, the primary stressed syllable is the only unit measurable in all conditions. In such cases, it is impossible to know if the locus extends beyond the constituent head.

The alternative hypothesis in domain-span shortening is that the locus is co-extensive with the domain: thus, all segments within a domain are shortened if the length of the domain increases. This account is suggestive of accounts of speech timing, such as the theories of isochrony discussed in Chapter 2, which suggest a unit with a global timing function.

The distinction between these two characterisations of the locus of domain-span processes may in practice be a matter of degree. Within the word, for example, the greatest polysyllabic shortening effect may be seen within the primary stressed syllable, either because this is the phonological head or because it is likely to be longer and thus more compressible than other syllables. Likewise, the syllable nucleus may show the greatest effect, either as the head of the syllable, or, because of the intrinsic durational properties of vowels and their relative expandability compared to other phone classes such as stops, as the longest part of the syllable.

The experimental investigation of domain-span effects is complicated by the fact that an increase in the size of some lower-level constituent such as the word also increases the size of higher-level constituents which dominate the word, such as syntactic or prosodic phrases. This potential ambiguity of interpretation may be resolvable if the locus of the effect is different between two domains, as discussed further in relation to domain-edge processes.

### **Domain-edge processes**

Durational processes associated with constituent boundaries are typically characterised as “lengthening”, where segments at the boundary have greater duration than similar segments placed constituent-medially: for example, word-initial lengthening or phrase-final lengthening. In the domain-and-locus description, such effects are called “domain-edge” processes.

Nespor & Vogel (1986) discuss two types of segmental process at constituent boundaries in prosodic phonology: domain-limit processes and domain-juncture processes. The former occur at the initial or final edges of constituents, regardless of context; the latter additionally require a certain segmental context on the other side of the boundary. Durational processes at constituent boundaries appear to have more in common with domain-limit segmental processes, as position in the constituent is apparently the

determining factor rather than any preceding or following context. The term domain-edge is preferred to domain-limit here, however, and subsumes both domain-initial and domain-final processes.

The locus of domain-edge processes is, of course, adjacent to the domain edge itself. Studies discussed in Chapter 2 suggest that the locus may be structurally determined rather than simply being a matter of proximity. In particular, the position of the final stressed syllable in a phrase or utterance may influence the extent of the locus of final lengthening: where the stressed syllable is in absolute-final position, the stressed syllable rhyme appears to manifest the lengthening effect; where there is an unstressed syllable between the final stressed syllable and the boundary, the unstressed syllable shows the greater lengthening, but the stressed syllable may also show some lengthening (for example: Nakatani *et al.* 1981; Turk 1999, Cambier-Langeveld 2000). This suggests that domain-edge lengthening is progressive within the locus: that is, the subconstituents of the locus nearer to the boundary show a greater effect. As discussed in Chapter 2, within the rhyme of an absolute-phrase-final syllable, there may be more lengthening on the coda than on the nucleus.

The terms “lengthening” and “shortening” suggest that domain-edge and domain-span processes are in conflict, but this is not necessarily the case, as shown by considering a lexical word in citation form: that is, spoken in isolation. For example, the syllable /dɛt/ is predicted to shorten in the sequence

*debt ... cadet ... cadetship*

by both domain-edge and domain-span processes. The domain-span hypothesis predicts that stressed syllable will have the longest duration in the monosyllable because the domain is phonologically shortest; in the phonologically-longest domain, the trisyllable, the stressed syllable will have the shortest duration. The domain-edge hypothesis predicts that stressed syllable have the longest duration in the monosyllable because it is both domain-initial and domain-final; it will have the shortest duration in the trisyllable, because it is neither domain-initial nor domain-final. Thus, determining which process is responsible for observed effects is not always straightforward. Furthermore, in this case, the domain of either type of process could be the word, or a phrase of some kind, or the utterance, as these constituents are co-extensive in citation form.

A similar dichotomy of interpretation exists for durational processes within prominence-delimited constituents, as pointed out by Fowler (1990:201): “Stressed syllables in monosyllabic stress feet that are identified as unshortened in a stress-timed account can just as well be identified as lengthened; so the absence of stress-timed shortening can instead be described as a lengthening of a stressed syllable at the right edge of a foot. As such it may count as a sort of preboundary lengthening.” The use of

the terms “lengthening” and “shortening” may be simply a matter of the viewpoint from which the effects are considered; however, although domain-edge and domain-span processes may make the same predictions about durational variation at the syllable level, they are theoretically distinguishable if patterns of durational variation in subsyllabic constituents are taken into account. This is the approach taken in the experimental work reported in Chapter 4 and in other recent work (Turk & Shattuck-Hufnagel 2000).

A domain-edge process is suggested if the locus is aligned with the boundary and the magnitude of the effect within the locus diminishes with distance from the boundary; in contrast, in a domain-span process, the variation is not expected to be progressive within the locus and may be greatest on the phonological head of the constituent.

As noted above, there is another potential ambiguity of interpretation of observed domain-edge effects, particularly effects in constituents within nested hierarchies, such as syntactic or prosodic constituents. The association of a domain-edge effect with a particular constituent is made more difficult by the fact that the edge of that constituent also corresponds with the edge of subordinate constituents: in Nespor & Vogel (1986)’s prosodic phonology, for example, the final boundary of an intonational phrase is necessarily the final boundary of a phonological phrase and all other subordinate constituents. This is a consequence of the “strict layer rule”, described in Chapter 2. Thus, unless appropriate experimental controls are exercised, it cannot be guaranteed that the effect is associated with the intonational phrase rather than with a subordinate constituent. Similarly, the presence of higher-level boundaries should be taken into account when examining domain-edge processes in lower-level constituents. As when distinguishing domain-edge and domain-span processes, consideration of the locus may be helpful here: it is possible that higher-level constituents may have a locus which extends further from the domain-edge, as well as manifesting a greater magnitude of lengthening than lower-level constituents. Evidence relating to possible patterns of hierarchical lengthening is considered in later chapters.

### **Prominence**

The arrangement of prominences, such as stressed syllables, may have durational consequences, because as domain-span processes are associated with prominence-delimited units, such as cross-word feet, as well as with syntactic or prosodic constituents. In addition, the direct lengthening effects of prominence—stressed syllables are longer than unstressed syllables; accented syllables are longer still—should be accommodated within a complete descriptive framework of speech timing. These observations do not appear to be classifiable as domain-edge or domain-span effects, however. Indeed, although the locus of lengthening may be observable, it is less cer-

tain how the domain should be characterised.

The lengthening associated with lexical stress appears to have a syllable locus: all parts of the syllable are likely to be longer in a stressed syllable than an unstressed syllable, although as noted above, Klatt (1976) only alludes to lengthening in the stressed syllable onset and coda. If lengthening were only associated with primary stress, then it could be said that the lexical word is the domain of stress-related lengthening; within polysyllabic words, however, there may be more than one syllable which has a lexical stress, with the difference between the two only becoming apparent when the word carries a pitch accent. This suggests a domain with more than one locus, which whilst not inconceivable, is less theoretically satisfactory. As discussed in Chapter 2, the within-word foot may be defined as beginning with a stressed syllable and continuing to the next stressed syllable or to a word boundary, whichever is the sooner. Such a unit might be said to be the domain of stress-related lengthening, which could then be viewed as a domain-edge effect, specifically a domain-initial effect with a syllable locus. There is, however, a circularity in the definition of such a unit, unless independent evidence for its existence can be found. Furthermore, stress-related lengthening seen as a domain-edge effect would lack the progressive quality of lengthening associated with other domain-edge effects.

With regard to the phrasal level of prominence, studies such as Turk & Sawusch (1997) and Turk & White (1999) refer to the domain of accentual lengthening to mean the stretch of speech which undergoes lengthening. It is clearly more consistent within the current framework to use the term locus for this purpose, although, as for lexical stress, it is not clear how the domain of the process ought to be characterised. Both the experimental evidence regarding the locus of accentual lengthening and the question of its domain are discussed in more detail in Chapter 2. Also raised in Chapter 2 is the possibility that many previous studies of durational effects in speech—including domain-edge and domain-span effects—use materials in which the words being measured are likely to contain a pitch accent. Thus, the findings of such studies do not necessarily apply to words which do not carry an accent.

The fact that both lexical and phrasal stress are associated with greater duration may be seen as a reason for treating degrees of prominence as variations within a single dimension. This is the approach often taken in phonological accounts of prominence: Hayes (1983) and Selkirk (1984), for example, represent different levels of stress as columns of different heights on a metrical grid, in an attempt to account for the distribution of lexical and phrasal stresses. The approach taken here, however, is to treat lexical and phrasal stress as largely distinct from the point of view of speech timing. This separation reflects the identification by Bolinger (1981) of two different types of prominence relations: full vs reduced vowels (in effect, the presence or absence of lex-

ical stress) and accented full vowels vs unaccented full vowels (in effect, the presence or absence of phrasal stress). There are phonetic and theoretical reasons to favour this approach in the analysis of durational effects.

In the context of a discussion about the distribution of stress and the phenomenon of stress shift, Fourakis & Monahan (1988:284) say that metrical phonology is “a theoretical construct based on impressionistic transcriptions and very little acoustic research has been carried out to test its predictions.” Certainly, the acoustic correlates of the different levels of prominence represented in the grid do not support a unified approach: differences in the acoustic marking of different degrees of stress are qualitative as much as quantitative. In particular, phrasal stress is indicated primarily by pitch accent and its durational consequences may extend beyond the stressed syllable; in contrast, the primary cues to lexical stress are vowel quality and duration, with the most significant durational effects on the vowel nucleus of the stressed syllable. Beckman & Edwards (1994) examine the articulatory differences underlying the accented vs unaccented distinction and the stressed vs unstressed distinction, finding that the two prominence contrasts are associated with variation in different articulatory parameters. In the terms of the task-dynamic model they apply to their data, accented syllables are articulated with greater gestural magnitude than unaccented syllables, whereas stressed syllables are articulated with less gestural stiffness than unstressed syllables. This effectively means that increased duration associated with phrasal stress is, at least in part, a reflection of the more extreme displacement of articulators compared with unaccented stressed syllables, whereas the increased duration associated with lexical stress reflects slower movement of articulators towards their target<sup>6</sup>.

There are also levels of prominence distinguished in grid-based approaches to metrical phonology which do not have specific durational consequences. Sluijter & van Heuven (1995) suggest that the distinction between primary and secondary lexical stress is not realised durationally, but by variation in high frequency energy levels. Indeed, Bolinger’s two-tier approach to prominence suggests that “primary and secondary word stress differ not in the degree or type of articulatory or acoustic prominence, but in the instructions they provide for the placement of pitch accent” (Shattuck-Hufnagel & Turk 1996:22); likewise Ladefoged (1993) suggests that the distinction is only phonetically realised when one of the stresses carries a pitch accent. Nuclear pitch accents and non-nuclear pitch accents are also placed on different levels in metrical grid representations: Silverman & Pierrehumbert (1990), while not directly addressing the durational consequences of this distinction, make the claim based on their results regarding peak alignment that “the distinction exists in prosodic organisation, but [we] do not find evidence for it in the inventory of English pitch accents or the

---

<sup>6</sup>The vocalic sections of unstressed syllables will also have a less extreme articulation than stressed syllables in most cases, by the nature of the centralised reduced vowel.

phonetic rules for pronouncing them” (Silverman & Pierrehumbert 1990:105).

The theoretical reasons for treating lexical stress and phrasal stress as distinct relate to the observation of Beckman & Edwards (1990:165) that “accent patterns belong to the intonation and thus are part of the post-lexical phrasal phonology, whereas stress patterns are largely specified in the lexicon.” The distribution of lexical stresses is predictable to a great extent from the text of an utterance and stands in arbitrary relation to meaning<sup>7</sup>. In contrast, there are a number of possible accentual patterns for any given utterance, and the accentual choice made by a speaker provides information regarding the pragmatic interpretation of the utterance (Ladd 1996). Whether the unified approach to prominence taken in metrical phonology has utility in other areas is beyond the scope of this discussion; as the accounts of the durational consequences of lexical and phrasal stress in Chapter 2 indicate, for timing purposes the two-tier approach advocated by Bolinger appears more appropriate.

A further, terminological distinction should be made at this point: the term “rhythm” is used widely in discussion of speech timing, sometimes apparently to encompass all systematic durational variation, sometimes more narrowly to refer to the durational effects of all levels of prominence, and sometimes to refer only to the distribution of lexical stresses. The latter approach is taken here: rhythm means the pattern of alternation of strong syllables (carrying primary or secondary lexical stress) and weak (unstressed) syllables. The term “pitch accent” as used here relates solely to the occurrence of phrasal stress, which is intonationally marked. To avoid any confusion, the term “phrasal stress” is generally avoided henceforth.

## 1.4 Research questions

This dissertation examines a number of durational effects at the suprasyllabic level and attempts to characterise the domain and the locus in each case. Constituents of syntactic and prosodic hierarchies and prominence-delimited units are considered as potential domains of edge and span processes in the review of research presented in Chapter 2. The new experimental work described in Chapters 3 and 4 examines domain-edge and domain-span processes at two structural levels: the word and the utterance. Specific questions raised in the research review and in the experimental work include:

- Do all constituents serve as domains of edge and span processes or are such processes specific to particular levels of some putative hierarchy?

---

<sup>7</sup>There are, of course, a small number of minimal pairs such as *'accent* and *ac'cent* in which the noun carries primary stress on the first syllable and the verb carries primary stress on the second syllable. In the great majority of cases, however, the placement of stress within a word is of no syntactic, semantic or pragmatic significance.

- If domain-edge or domain-span processes are associated with more than one constituent, is the locus the same in each case?

The durational effects of pitch accent are also considered: firstly, the locus of accentual lengthening is examined; secondly, given the likely presence of pitch accent in many previous studies of durational effects, the interaction between accentual lengthening and domain-edge and domain-span processes is examined.

### 1.4.1 The structure of the dissertation

These research questions are applied in Chapter 2 to previous experimental studies. The work discussed—like the experimental work presented in later chapters—is that which arises from what may be characterised as the “laboratory phonology” approach to the study of speech timing: particular hypotheses are tested about specific durational effects by examining recordings of subjects reading prepared materials, wherein certain factors are controlled whilst others are experimentally varied. Evidence for domain-edge and domain-span processes from such work is discussed with regard to syntactic constituents and prosodic constituents; evidence for the existence of such processes in prominence-delimited units is also examined. Previous studies of the durational effects of pitch accent are reviewed.

Two particular aspects of this previous experimental work are examined further in Experiment 1, presented in Chapter 3. This experiment is designed to determine: firstly, whether the word is a domain of durational processes or whether previous experimental findings regarding the word are a result of domain-edge or domain-span processes at other levels of linguistic structure; secondly, whether the presence or absence of pitch accent has an influence on observed effects at the word level. The findings of Experiment 1 indicate that the word is a domain of durational variation and that some of this variation may be affected by the presence of pitch accent.

Experiment 2, presented in Chapter 4, is designed to determine the type of processes operating at the word level, whether domain-edge or domain-span processes, or some combination; the existence of domain-edge and domain-span processes at the utterance level is also tested. Examination of the loci of variation indicates that there are domain-edge processes, but not domain-span processes, at the word level and the utterance level; there is some evidence of a domain-span effect within a sub-word constituent, but this may also be interpreted as a word-final effect. The distribution of accentual lengthening within monosyllables, disyllables and trisyllables is found to be responsible for apparent word-span compression. Domain-edge processes do not appear to interact with the durational effects of pitch accent.

The conclusions drawn from Experiments 1 and 2 are summarised in Chapter 5, and compared with other experimental findings regarding domain-edge and domain-

span processes. A model of English suprasyllabic speech timing is outlined in Chapter 6, based upon localised lengthening effects at domain edges and within pitch-accented words.

### **Conventions used in this dissertation**

The terms “polysyllable” and “polysyllabic” are used here to refer to a word of two or more syllables, typically in the contrast “monosyllables vs polysyllables”, rather than in the standard sense of a word of many, or more than three, syllables.

Where results are given as percentage of lengthening or shortening, this is calculated by a standard formula. For lengthening, the difference between the longer and shorter duration is expressed as a proportion of the duration in the shorter context. For shortening, the difference between the longer and shorter duration is expressed as a proportion of the duration in the longer context.

Where comparisons of segmental duration between two contexts are made in the text, underlining is used to indicate the measured segments where these are not otherwise identified, thus: “a stressed syllable is longer in an accented monosyllable than in an accented disyllable: for example, thank fulfil vs thankful Phil.”