1033

Phonology 13 (1996) 139–169

*A phase window framework for articulatory timing**

Dani Byrd
University of California, Los Angeles/Haskins Laboratories

1 Introduction

One of the most significant challenges in the study of speech production is to acquire a theoretical understanding of how speakers coordinate articulatory movements. A variety of work has demonstrated that articulatory, prosodic and extralinguistic factors all influence speech timing in a complex and interactive way. Models such as Articulatory Phonology that stipulate the relative timing of articulatory units must be revised to allow for this variability. Such a revision is outlined below.

The following work should be viewed as a presentation of a new framework for conceptualising articulatory timing. This approach, meant to be programmatic rather than conclusive, is productive if it motivates research that might not otherwise have been undertaken. §1 overviews Articulatory Phonology. The implementation of articulatory timing in terms of phasing relations is discussed. Speech production data bearing on timing variability are discussed in §2. §3 argues for an alternative to Articulatory Phonology's current rule-based approach to intergestural timing that can allow for linguistic and extralinguistic variables to systematically influence phasing relations. §3.2 introduces the PHASE WINDOW framework, which allows the degree of articulatory overlap between linguistic gestures to vary within a constrained range. Finally, §4 concerns the relation of intergestural timing to the postulation of the segment as a primitive unit in phonology. It is hypothesised that certain intergestural timing relations are stable and lexically specified. Gestures whose coordination is constrained by lexical PHASE WINDOWS seem to bear a close relation to those conglomerates of gestures that constitute what is traditionally considered to be a segment.

1.1 A theoretical framework

In Articulatory Phonology, developed by Browman & Goldstein (1986, 1988, 1989, 1990a, b, 1991, 1992a, b, 1995a, b), articulatory gestures are the units of phonological representation. Gestures are coordinated, goal-

They liken the choice of a few invariant phase relationships to the choice of values for the dynamic parameters of the gestures (Browman & Goldstein 1995b). In more complex formulations, Articulatory Phonology allows groups of gestures, syllable-sized and smaller, to be marshalled into an organisation that may in turn be coordinated to another gesture (Browman & Goldstein 1988). Phasing rules may synchronise a point in one gesture to an arithmetically defined abstract index calculated from specific points in another set of contiguous gestures, e.g. the C-centre (Browman & Goldstein 1988).

There is additional information to which phasing rules have access. The rules proposed in Browman & Goldstein (1988, 1990b) refer to whether a gesture is a consonant or a vowel – that is, whether it is on the consonant or vowel tier – and to gestural contiguity on a particular tier. They also refer to ASSOCIATION. In Articulatory Phonology association encodes precedence relations (Browman & Goldstein 1990b) and determines which gestures are phased with respect to one another.[1] Furthermore, Browman & Goldstein (1995a, citing Krakow 1989 and Sproat & Fujimura 1993) characterise the phasing relationships for lips and velum in nasals, and for tongue body and tip in laterals, by making crucial reference to the LINKING of two gestures. They say that 'in the language of Articulatory Phonology... [l] *consists* of two gestures' (1995a: 21; emphasis added) and that 'syllable-final *linked* gestures are phased so that the gesture with the wider constriction degree comes earlier' (1995a: 25; emphasis added). The relation between linking and association is unclear. In our discussion of the nature of segmenthood in §4, the question of whether some element exists that 'consists' of gestures, and the integral role of phasing in expressing this idea, will be important.

The representation of an utterance, or gestural score, explicitly specifies which gestures are phased with respect to each other and what that phase relationship is (Browman & Goldstein 1990b). Currently, the organisation of gestures comprising a gestural score is not derived dynamically:

> Once a gestural score is specified, it remains fixed throughout a given simulation, defining a unidirectional, rigidly feedforward flow of control from the intergestural to interarticulator levels of the model. The gestural score acts, in essence, like the punched paper roll that drives the keys of a player piano... [Various] data imply that activation patterns are not rigidly specified over a given sequence. Rather, such results suggest that activation trajectories evolve fluidly and flexibly over the course of an ongoing sequence governed by an intrinsic intergestural dynamics. (Saltzman 1995b: 167)

In Articulatory Phonology, lexical distinctions are made in a limited number of ways (Browman & Goldstein 1995b). One, a gesture may be present or absent. Two, the temporal coordination between gestures may differ. This information is found in the gestural score. Lexical gestural scores may be altered in two ways after their creation. One, a gesture's dynamic characteristics may change (e.g. resulting in a smaller or shorter gesture) and two, changes in phasing, or intergestural timing, may occur. Phasing may occur in the lexicon and external to the lexicon; the latter, yielding for example a variety of casual speech processes (Browman & Goldstein 1990b). Lastly, utterances must be coordinated above the word level such that there is some mechanism for phasing one word relative to another word. This, by definition, must also take place outside the lexicon. Currently in Articulatory Phonology, some provision will need to be made for rules changing phasing relationships, for example due to rate, stress and style. In the discussion to follow, the status of invariant (i.e. a limited and invariant set of) pointwise values that define stable phasing relationships, implemented by rule, is an important topic.

1.2 Keating's window model of coarticulation

Certain concepts developed in a very different approach to phonetic implementation, namely Keating's targets-and-interpolation approach, are also relevant to the discussion below and, for this reason, are introduced now. In Keating's window model of coarticulation (Keating 1990a, b), articulatory and/or acoustic targets are projected temporally and spatially by the feature specification of the relevant segments. Crucially for our interest, a feature projects a target window that allows the realisation of a featural specification to vary within a specified range (Keating 1990a). She explains:

> this window is not a mean value with a range around that mean, or any other representation of a basic value and variation around that value. It is an undifferentiated range representing the contextual variability of a feature value. For some segments this window is very narrow, reflecting little contextual variation; for others it is very wide, reflecting extreme contextual variation. Window width thus gives a metric [of] variability. There is no other 'target' associated with a segment; the target is no more than this entire contextual range...
>
> Windows are determined empirically on the basis of context, but once determined are not themselves contextually varied. That is, a feature value...does not have different windows for different contexts. Information about the possibilities for contextual variation is already built into that one window. (Keating 1990b: 455, 456)

This concept will be fruitful in §3, where a new framework for considering variability in interarticulator timing is presented.

2 Speech timing and phonology

Before leaving the topic of Articulatory Phonology, a few words should be said about the interest of articulatory timing to the field of phonology. If

Articulatory Phonology is taken as a theory of phonology, as its conceivers intend, then intergestural timing is of fundamental importance because it is one of a very limited number of ways that (i) lexical representations (gestural scores) can differ from one another and (ii) gestural scores can be altered postlexically. If, however, Articulatory Phonology is taken to be of interest as a framework for phonetic implementation (of some other sort of phonological representation), then interarticulator timing, while perhaps not of immediate importance for representational issues, is certainly relevant to the phonological theory for several reasons: (i) it serves as one component of the interface between the phonological representation and its overt realisation as speech; (ii) it is the main means by which certain phonological contrasts (/p/ *vs.* /p^h/, /m͡b/ *vs.* /m/) are realised; and (iii) it provides empirical evidence for prosodic structure such as phrase boundaries. The discussion presented here actually makes the first assumption – that is, Articulatory Phonology is discussed as a theory of phonology – and one particular component of the theory, the assignment of temporal relations between gestures, is critically evaluated. However, the remarks made below are, for the most part, unaffected by assuming Browman & Goldstein's model to be a phonological model, as opposed to a phonetic one. The important point is that a research programme in phonology can be directed from the top down or from the bottom up. That is to say, one way of gaining insight into representational issues and phonological processes is to understand the general articulatory patterns that are their output. The discovery of a small number of parameters governing the speech system becomes most interesting when the relation of those parameters to information required for phonological specification becomes close. When this happens, it is suggestive that the two research programmes (top-down and bottom-up) are converging on what is significant in the phonology.

Considerable progress has been made in understanding certain phonological processes, for example assimilation (e.g. Barry 1985; Browman & Goldstein 1990b; Byrd 1992; Nolan 1992), by examining low-level articulatory patterning. Furthermore, just as articulatory detail informs phonology, substantive consideration of phonological structure is necessary to understand articulation. As research in speech production becomes more and more integrated with linguistic theory, it has become apparent that articulatory detail cannot be understood except in the context of linguistic structure; such structure includes, but is surely not limited to, the domains of gestures (or segments), syllables and phrases. Effects of such phonological structure pervade low-level articulatory behaviour. Despite the pervasiveness of these effects, only a very few phonological 'signatures' have been elucidated at the level of articulatory patterns. §2.1 is concerned with the relation between linguistic structure and one particular aspect of articulation – the temporal coordination of articulatory gestures. I review evidence that speech timing is affected by syllabic and phrasal structure. §2.2 discusses the experimental literature that speaks to differences in degree of timing stability.

2.1 Articulatory timing and suprasegmental structure

Recent studies have shown that intergestural timing differences exist as a function of syllable position. For example, Krakow (1989), examining nasal consonant production, finds that in syllable-initial position the oral and velic gestures are reliably coordinated in a roughly synchronous relationship, while in syllable-final position the velum-lowering gesture consistently preceded the oral closure. Similarly, Sproat & Fujimura (1993) find that in word-initial [l]'s the peak tongue tip movement precedes the tongue body movement valley, while in word-final position the body lowering leads the peak. Hardcastle (1985) reports less coarticulation in /kl/ clusters in syllable onset position than across a word boundary. Byrd (1996) finds the word-initial consonant cluster [sk] to have less temporal overlap than word-final clusters and sequences spanning a word boundary.

Just as position in the syllable affects intergestural timing, so does position in the utterance. For example, McLean (1973) investigates the effects of a variety of prosodic boundaries occurring between the vowels of /CV#Vn/ sequences. He finds that the anticipatory coarticulation of the nasal consonant (i.e. onset of velum lowering) was consistently delayed relative to the preceding vowel when particular phrasal boundaries were present. Hardcastle (1985) reports on the coordination of /kl/ sequences in which a variety of juncture types occur between the consonants. In general, he finds that 'the condition least favourable to coarticulation [between [k] and [l]] is the prosodically marked clause or sentence boundary at the [normal slow utterance rate]'. Holst & Nolan (1995) studied assimilation in [sʃ] sequences. They present acoustic data that they divide into four categories ranging from most like a [sʃ] sequence, indicating an absence of assimilation, to a sequence with spectrally stable [ʃ] characteristics indicative of assimilation. In her commentary on the Holst & Nolan data, Browman (1995) presents evidence that this continuum is indicative of a progressively increasing degree of overlap between [s] and [ʃ], and comments that the presence of a clause boundary intervening between the consonants is negatively correlated with the degree of gestural overlap between them. Similarly, Byrd *et al.* (1996) report less overlap in Tamil [C#C] sequences when a large phrase boundary intervenes between the consonants.

2.2 Articulatory timing variability

The search for invariant timing properties in speech has had mixed success. In examining movements composing a *single gesture*, many stable aspects of temporal coordination have been found. Within a single articulatory movement, some studies of articulatory kinematics have suggested that the relation of peak velocity to displacement (or in some studies the relationship of this ratio to duration) is the dynamic *intra-*

gestural property that remains stable across variation in linguistic and extralinguistic contexts (Kozhevnikov & Chistovich 1965; Ohala *et al.* 1968; Mermelstein 1973; Sussman *et al.* 1973; Kuehn & Moll 1976; Ostry & Munhall 1985; Gracco 1988; Gracco & Abbs 1989; Vatikiotis-Bateson & Kelso 1993). However, work by Ostry *et al.* (1983), Munhall *et al.* (1985) and Kelso *et al.* (1985) has documented systematic rate and stress effects on the peak velocity to displacement relationship.

Between gestures that are traditionally considered to constitute a segment, researchers have found instances of invariant and bimodal timing relations (Löfqvist & Yoshioka 1981; Munhall *et al.* 1986; Krakow 1989; Löfqvist 1991, citing Löfqvist & Yoshioka 1984; Sproat & Fujimura 1993). Saltzman & Munhall (1989) note that this argues for the existence of a higher multigesture unit in speech production. They conceive of implementing such a unit in terms of dynamical coupling.[2] Note that it is not just *any* two gestures that display this tight timing relationship but rather gestures that have long been considered to belong to the *same* segment. The implications of this are discussed further in §4.

However, the existence of invariant phasing relationships between gestures composing what would traditionally be considered *different* segments is not evidenced. The few studies of such timing relationships have had methodological flaws or have not found evidence of stable, i.e. invariant, timing. See Löfqvist (1991) and Keller (1990) for overviews. Studies by Tuller *et al.* (1982) and Tuller & Kelso (1984) examined the articulatory coordination of intervocalic consonants with their adjacent vowels (see Keller 1990 for a concise review of this research programme). This work reported stable consonant latencies with respect to the vowel cycle. However, further work (Barry 1983; Munhall 1985; Benoit 1986; Sock & Jah 1986) indicated that some of this correlation is due to a statistical artifact. Remaining effects reported in Munhall (1985) are suggested by Keller to be due to large speech rate variation affecting all coupled articulatory measures similarly, and are not replicated within cells (see Keller 1987, 1990). Keller (1990) also notes difficulties in replicating the results of the original study, citing Lubker (1986) and Nittrouer *et al.* (1988).

Nittrouer *et al.* (1988) conclude that while interarticulator timing may well be controlled in terms of phase relations, the specific amount of overlap may vary continuously as a function of linguistic and non-linguistic factors. This is a concept that we will develop here. They state:

> In contrast to the findings of Kelso *et al.* (1986a), we found no support for the notion that the relative phasing of jaw vowel gestures and upper lip consonant gestures are stable across manipulations in linguistic and non-linguistic factors. In fact, the evidence from the present experiment suggests that the intersegmental organisation of gestures is a function of the utterance being produced. In other words, the *phase relations between articulatory gestures used in the production of adjacent segments varies* [*sic*] *systematically based on linguistic and nonlinguistic structure,*

> which includes speaking rate, stress pattern, syllable structure, and consonant identity. (Nittrouer *et al.* 1988: 1659; emphasis added)

This finding of timing variability receives further support for jaw and tongue tip gesture phase relations in Nittrouer (1991), in which vowel quality, consonant voicing, stress and rate were manipulated. Lubker (1986), also in an effort to examine the results of Kelso *et al.* (1986a), finds that:

> observed across all subjects and utterances, the upper lip movement toward /b/ began anywhere from 115° to 230° on the jaw-phase portraits with a good deal of clustering around 180°...There was greater consistency within subjects, but still not enough to fit any very tight model or warrant any strong theory...Within each 'stressed-unstressed' word pair, little variation of lip onset relative to the jaw-phase portrait can be seen...However, a great deal of variation can be seen across the four words...If the upper lip...is somehow triggering its movement onset on jaw phase, then it must be doing so based upon a different 'trigger' for each word. (Lubker 1986: 133–134)

He goes on to note the possibility of specifying a timing such as '$180° \pm 25°$' but rejects this option as lacking precision and elegance.

Löfqvist (1991, examining data from Löfqvist 1986) reports results parallel to those of Nittrouer *et al.*, i.e. no evidence of invariant intersegmental timing. He argues that this may indicate a difference in gestural cohesion within and across segments. (This is discussed further in §4.) Shaiman & Porter (1991) examine effects of stress and vowel/diphthong identity and report that vowel–consonant phase relations varied systematically. They note that their findings on the effect of stress concur with those of Nittrouer *et al.* (1988). Shaiman *et al.* (1995) report on intergestural coordination in a VCV sequence as a function of rate and find that the coordination varies with speaking rate systematically and speaker-specifically. Research presented in Byrd (1996) and Byrd & Tan (1996) describes significant effects of a number of factors on the timing of English consonant sequences, including consonant identity, syllable structure and speaking rate. De Jong (1991) finds that VC phasing is affected by consonant voicing. Finally, Hardcastle (1985) reports rate and phrasing effects (and an interaction between these effects) on the coproduction of /kl/ sequences.

In sum, we can see that the coordination of articulatory gestures varies in systematic ways. Phonological structure such as syllable and phrasal position plays an active role, along with extralinguistic information, in determining the temporal organisation of speech. An adequate theory of speech timing is impossible without a substantive means of incorporating the systematic effects of such linguistic structure.

3 The phase window framework

to what extent [*can*] *the notion of an 'ill-formed' word be reduced to that of a 'statistically improbable' word* (Pierrehumbert 1994: 184)

Next, we turn to a discussion of the theoretical formulation of timing implementation. This section will outline a new framework for describing speech timing that allows timing variability to be better conceptualised. The mechanism for timing proposed here, called the PHASE WINDOW framework, adopts the notion of relative phasing, as does Articulatory Phonology. However, I argue for crucial differences from Articulatory Phonology's approach.

3.1 Phasing rules

Articulatory Phonology currently implements each particular phasing relationship with a rule or rules specifying an invariant coordination. 'Readjustment' rules, changing phasing relations from those specified in the gestural score, would have to be admitted to allow for timing variation due, for example, to rate, stress or register. This offers little explanatory insight into the linguistic regularities of articulatory timing, not facilitating generalisability across similar rules. In discussing their proposed phasing rule for consonant clusters, Browman & Goldstein (1988) suggest that it may need to be refined to include syllabification effects and articulator-specific effects. It is not clear what they foresee as the nature of this rule refinement, but this seems to suggest concerns similar to those voiced here about the status of invariant phasing rules. In fact, their analysis of assimilation (see for example their response (1992b) to Kohler 1992) relies crucially on variation in gestural overlap within words. Keating (1995: 31) comments that if within-word variation in phasing as a function of postlexical structure is pervasive, then 'we cannot say that lexical specification tells us how to pronounce a word, only how to pronounce it in some particular context'. Either the number of phasing rules needs to be proliferated, e.g. specific to particular sequences and prosodic conditions, or phasing relations must be implemented in such a way that variability is possible. I pursue the latter course here and reject a rule-based approach to intergestural coordination.

Browman & Goldstein (1991: 319) state that 'there is a potential continuum [of overlap] ranging from complete synchrony...through partial overlap...to minimal overlap' and that 'there are no *a priori* constraints on intergestural organization within the gestural framework. The relative "tightness" of cohesion among particular constellations of gestures is a matter for continuing research.' In principle, any point in a gesture could be phased to any point in another gesture, thereby yielding an infinite number of possible phasing relationships. The lack of principled constraints on possible phasings makes this approach overpowerful. However, the postulation of phasing rules that have access only to three points in a gesture – the onset, target and perhaps release (Browman & Goldstein 1990a, 1995b) – is empirically overly constraining and lacks theoretical motivation. While no-one can doubt that certain timing relationships give rise to qualitative differences (see Goldstein 1989, 1990; Ohala 1990), why would exactly these three phase angles and no others, e.g. why 0° but not 1°, exist for timing rules? These facts lead us to believe that the instantiation of linguistic timing in terms of a set of phasing rules is of limited predictive value and is mostly useful as restatements of empirical observations. I formulate an alternative below that allows, but constrains, variability in phasing relationships. The various factors that influence coordination can be seen as competing simultaneously, each contributing to the final intergestural phasing relation.

3.2 The phase window

The phase window framework allows variability in a single assignment of phasing relationship rather than using a set of timing rules that operate sequentially to coordinate the articulatory units. As in Articulatory Phonology, precedence relations are encoded (transitively) in the gestural score in terms of *association*, which specifies which gestures are to be phased to which other gestures. That is, the lexical representation specifies the requisite gestures and specifies which gestures are to be coordinated relatively. In contrast with Articulatory Phonology, this approach to timing assumes that the intergestural phase relations are not all specified lexically, e.g. in the gestural score. (Lexically specified timing relations are discussed in §4.) Coordination between associated gestures is assumed to be variable but constrained to particular ranges specific to the types of gestures involved, for example: V-to-C (reflexively), C-to-C and V-to-V. I refer to these ranges as PHASE WINDOWS. The ultimate timing relations are actualised concurrently with, or within (see §3.4), a dynamic model that converts the output of the linguistic gestural model (the phonology) into articulator movements.

Temporal organisation can create meaningful contrasts in the lexicon, e.g. voice onset time. In light of this, it seems reasonable to assume that temporal relations that are specified lexically are discrete and/or stable. This will be discussed further in §4. However, *outside* the lexicon, I assume that interarticulator phasing relationships depend on a wide range of linguistic and extralinguistic factors. A probabilistic approach to intergestural phasing (leaving aside the question of which points stand in the phasing relationship) is proposed. A general tenet of this research programme is that a given phasing relationship is constrained both physically (by biology) and language-specifically (by learning) to occur within a certain permissible window – the phase window. I propose that there are upper and lower limits placed on a particular phase window that are determined by both system constraints (motor, auditory and cognitive) and language constraints (language-specific, learned permissible phasing relationships). Clearly the window defined by the latter constraints will be properly contained in that defined by the former. Utterance-specific (task-specific) INFLUENCERS then act to weight the window but do not constrain it further. The weighting of this phase window can be considered to take place in a probabilistic manner. These linguistic and non-linguistic influencers determine where in the range of permissible overlap relationships a specific phasing is likely to be implemented. This is in the spirit of Keller's (1990) idea of mutual competition between codeterminers of speech timing, including perceptual and prosodic factors. Like Keating's spatial windows approach, the temporal phase window framework assumes that some kind of optimisation occurs on line. Of course, this optimisation is complex because of the large number of simultaneously extant influencers that weight the window. The proposed phase window framework is envisioned as obviating the need for phasing rules within a model using relative timing such as Articulatory Phonology, and as a means of conceptualising variability in intergestural timing.

Docherty (1992) presents an analysis of voice onset time in the same vein as our proposal, in which he also extends Keating's concept of spatial windows to the temporal domain. That is, his temporal windows 'define sets of acceptable inter-articulator temporal relationships' (1992: 217). He hopes that window placement in the temporal space will help delineate categorical differences within a language as well as between languages. Window width or variability of a particular temporal interval would be language-specific and context-specific. Distribution of timing patterns within his temporal window is hypothesised to be context-specific and speaker-specific. Finally, Docherty suggests that a related approach might be possible in Articulatory Phonology: 'the relatively abstract phase angles which specify inter-gestural sequencing...could be translated (by language specific rules) into equivalence classes of phase angles, or "windows" within the syllable cycle' (1992: 223). While our proposal differs from Docherty's in significant ways, it is clear that both address similar concerns regarding handling variability in speech timing through the use of constraints.

With respect to the phase window framework, let's consider what the phase window for the oral constriction gestures in a sequence of two consonants might look like. First, given the assumption that the two consonants are timed with respect to one another, we need to define what points stand in a phasing relationship. This question is not addressed here. Let's assume that the onset of C2 is phased to some point in C1. Thus the relevant phasing relationship for this phase window is $C1(x°) = C2(0°)$. Further, as outlined above, some cross-linguistic, e.g. universal motor, constraints exist that limit the value of x. Let's postulate for our example that these limits are minimal and are something like a lower bound of 0° and an upper bound of 360°. That is, C2 may not start before C1 is activated or after its activation ceases entirely. The phase window is, however, also additionally constrained in a way specific to English. Although very little cross-linguistic work has been done on this timing relationship, it does not seem unreasonable to assume that English

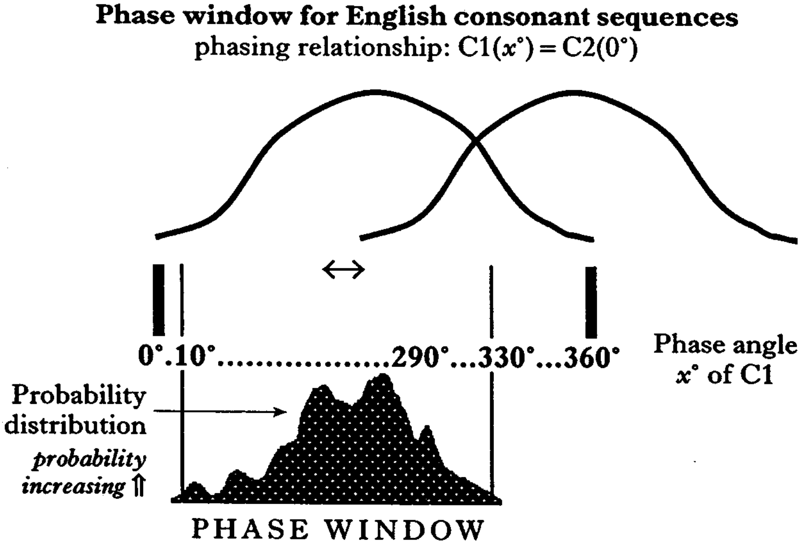

Figure 1
The phase window framework: a C-C phase window showing the combined effects of influencers within the phase window.

consonant sequences are typically quite overlapped in comparison to other languages, as we find systematic perceptual assimilations (Byrd 1992) and generally no acoustic releases of consonants in sequence (Jones 1956; Catford 1977; and others). A language with systematically released consonants, like Tsou (Tung 1964) or Salish (Flemming *et al.* 1993), would have a higher lower bound, and perhaps upper bound, on the window, thereby yielding less overlapped sequences than in English. Furthermore, let's suppose that English allows a wide range of possible timing relationships, as in fact appears to be the case. So in the case of our supposition, English has a large amount of variation in CC timing as compared to many other languages. These language-specific constraints are learned by the child acquiring English.[3] She learns that no more than a certain degree of overlap is allowable in consonant sequences for words to be intelligible to other speakers and that overlap is required to be at least a certain amount such that acoustic releases aren't present between the consonants. These constraints yield the consonant cluster phase window for English. The hypothetical universal and language-specific boundaries are marked in the diagram in Fig. 1 by thick and thin lines respectively.

We see that there is some probability, however small, of any value of *x* occurring in the phase window (otherwise it wouldn't be in the window).[4] The combined influence of the linguistic and extralinguistic conditions existing for the particular consonant cluster token of a single utterance determines the final probability density for the window. The more alike the contextual effects are from token to token, the more alike the combined influencer distributions will be. This will yield a high probability of similar organisations being realised in similar contexts – i.e. low token-to-token variability. Of course, an interesting empirical and theoretical question is how this determination of the combined weighting of the phase window is arrived at.

The system-specific and language-specific constraints embodied in the phase window reflect issues previously considered in the literature. While not precisely analogous to our proposal, Turvey's discussion of ecological psychology is relevant in considering a theory of timing that responds to task-specific requirements within a framework delimited by physical, systemic, and learned constraints. He says:

> Laws identify real possibilities. When circumstances – boundary conditions, constraints – are appended, actual events result... Nature, however, is not very economical with respect to patterns of coordination. There is a great diversity, with each pattern giving expression to the general laws and principles in very specific ways... Furthermore, in the province of coordinated movements, the circumstances appended to laws include intentions, plans, goals, and so on. Intentions function as exceptional boundary conditions on natural law. (Turvey 1990: 941)

Gracco has recognised similar goals in the study of speech production, noting that 'it is becoming increasingly clear that any behaviour is a reflection of multiple overlapping and interacting influences, each of which needs to be identified. The purpose of identifying the subcomponents is not strictly to assign function to structure but to evaluate their potential contribution to the overall process, and hence allow development of realistic and biologically plausible working models of the system' (1991: 53–54). MacNeilage (1970) describes Hebb's (1949) contention that motor equivalence requires the use of learned perceptual information in addition to moment-to-moment information about ongoing motor activity. The conception of the phase window is motivated by such concerns.

3.3 Effects of influencers on the phase window

Let's consider how variables may influence the probability density of a particular phase window. There are really three 'dimensions' to consider here. First, a variable may cause a preference for a particular region of the phase window. This is related to how much overlap a contextual variable is associated with. For example, a fast speech rate will favour the 'more overlapped' end of the window, and a slower speech rate the 'less overlapped' end. Secondly, variables may differ in the extent of the window over which they have an influence. This corresponds to how much variability an influencer will allow. Lastly, the level of weighting or activation contributed by particular variables may differ. As an example, the speech register could have a greater influence on the final probability density than the speaking rate.

Many variables affect intergestural timing, including intrinsic influences such as constriction location and degree, influences of adjacent contexts, structural influences such as syllable constituency and boundary location, and extralinguistic influences such as speaking rate. Such factors can, tentatively (because of the unknown complexity of interactions between variables), be hypothesised to be active in weighting the phase window. For example, in Byrd (1996) it was found that front consonants followed by back consonants tend to be more overlapped than back consonants preceding front ones. A front–back order in a consonant sequence might therefore weight a region in the 'more overlapped' end of the phase window. Similarly, stops were found to be more overlapped by a following stop than was an [s]. The manner of C1 would then influence the final probability density of the phase window accordingly. Byrd (1996) also presents some evidence that an onset cluster is less overlapped and less variable than a like coda cluster and heterosyllabic sequence. This would reflect a more narrow region weighted in the onset cluster context and a region more in the 'less overlapped' end of the window than for the other sequence types. In addition, other work (Hardcastle & Roach 1979; Browman & Goldstein 1988) has suggested that timing of the same gestures between words is more variable than within words. The presence or absence of a word boundary, or other prosodic boundaries, may influence the window over more or less narrow regions.

Let's consider, for example, the potential influence of speech rate on the phase window for consonant sequences. We observe in Byrd & Tan (1996) that rate has a roughly linear relationship to overlap in heterosyllabic consonant clusters. The schema in Fig. 2 indicates the strength and region of influence in the phase window for particular speaking rates. The *x*-axis is the phase window, the *y*-axis shows sample planes representing probability distributions in a continuum of speech rate and the *z*-axis indicates probability. (Only a single plane in the diagram is relevant for any particular utterance. This schema is intended to represent a three-dimensional space, although only slices through the *z*-axis (speaking rate axis) of that space are shown to simplify presentation.) The increase in overlap at faster rates can be seen by the movement of the peak across the window. The linear nature of this change is emphasised by the line overlaid on the peaks. Suppose also, for the sake of illustration, that variability in timing was found to decrease (i.e. coordination was more stable) at faster rates. Such a difference in timing variability would correspond to differences in the broadness of the distributions as rate changes. Such a difference is also illustrated in Fig. 2.

Let's consider how the timing of a consonant cluster in a particular utterance would be described. Suppose the following: (i) the consonant cluster is [sk]; (ii) it is in syllable onset position; and (iii) it is being spoken somewhat faster than 'normal'. Thus we might find the following influences, each of which is represented by a plane schematically in the top

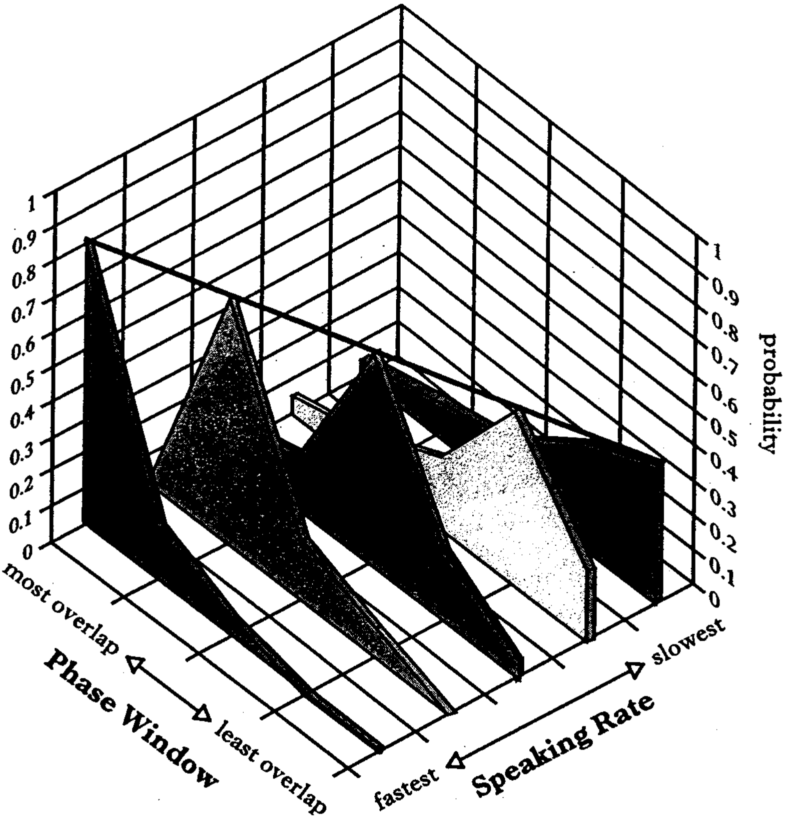

Figure 2
Schematic figure showing the effect of speech rate on the C-C phase window.

panel of Fig. 3: (i) consonant sequences of fricative–stop prefer less overlap; (ii) onset clusters prefer less overlap; and (iii) medium fast clusters prefer somewhat more overlap. (Alternatively, the second influence could be seen as a mapping from a continuous variable of prosodic cohesiveness to the probability distribution in the phase window, in much the same way as rate. See §3.3.1 for more on continuous *vs.* categorical variables.) Note that, unlike Fig. 2, the planes in Fig. 3 represent the influencers affecting a *single* utterance.

The most difficult question is how the individual probability functions for each variable combine to determine the final probability density for the phase window for the utterance. It is likely that these interactions are quite complex. Answering this complicated question is beyond the scope of this work and will require detailed computational modelling. The effects of each influencer could be added, convolved, overlaid[5] and subjected to a

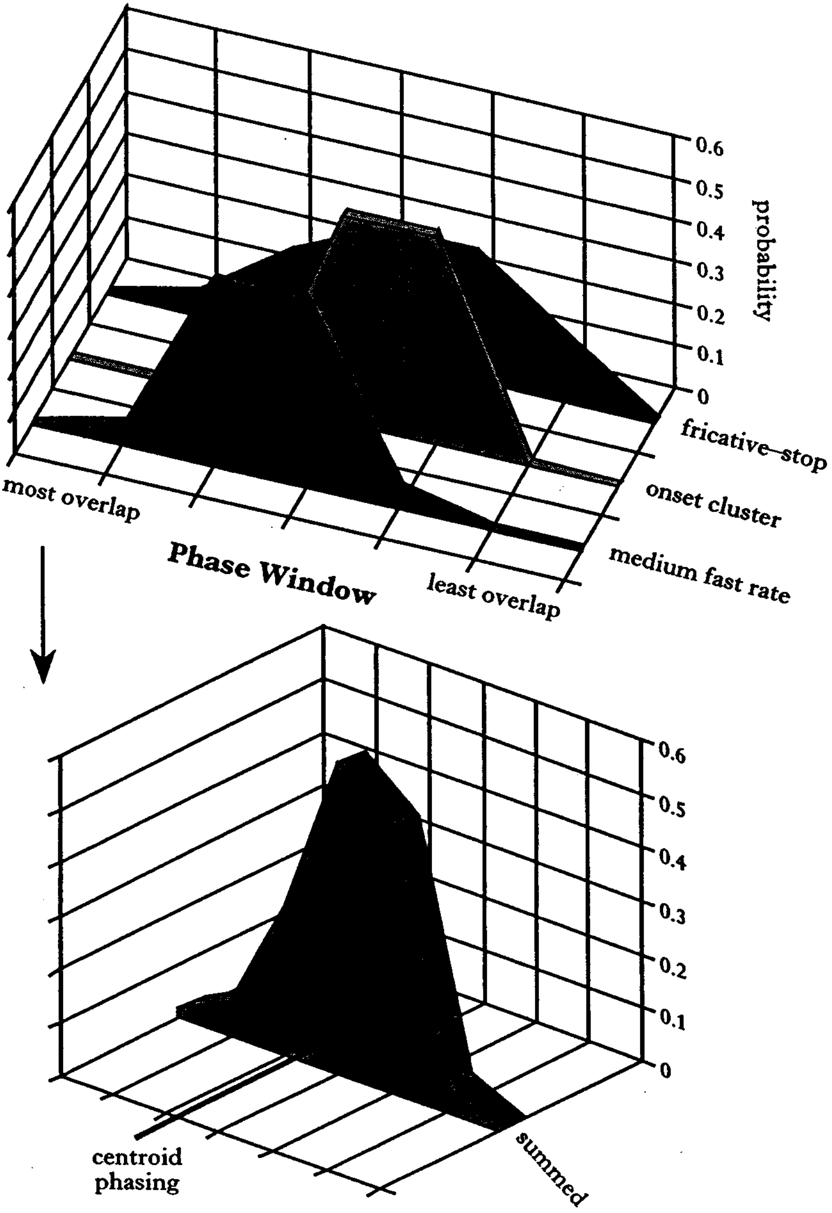

Figure 3
Schematic figure showing multiple influences on the C-C phase window for an [#sk] token at a medium fast rate.

peak-picking algorithm, or combined by any number of other possibilities. Two reasonable combinatorial possibilities are outlined here. First, the probability distributions of each influencer may simply be additive. The phase angle with the highest weighting or activation after the effects of all variables are added up specifies the phasing relationship with the highest probability of being realised. A second approach to combining effects of different influencers is to add the individual distributions, as before, and then take a weighted average in the form of the centre of mass or centroid value. This value would be the output phasing. (This is the approach taken by Kosko 1993 for combining fuzzy sets.) These two approaches yield different empirical predictions. The first admits unpredictability and the possibility of multimodal resultant distributions. The latter does not. Unfortunately, testing for outliers and biomodality under identical influencers is, in practice, difficult. It is impossible for an experimenter to ensure identical conditions on the part of the speaker for many tokens of an utterance.

The final phase relation for the sequence shown in Fig. 3 depends on the procedure used to combine the influencers – the additive or the additive-centroid. (Note that because most experimental studies on intergestural timing examine the *main* effects of certain variables on timing, I make no claim about the nature of their interactions and can infer nothing about the relative weighting of the influences that might exist. However, it would not be at all surprising to find that certain influencers are stronger than others.) In Fig. 3 the additive process yields a timing with a high probability of being realised in the middle of the phase window but some chance as well of being realised in the more overlapped portion of the window. The additive-centroid combination does not admit any uncertainty in the outcome; it yields a single phase angle that specifies the CC timing.

Cases in which substantially different behaviour is predicted by the two combinatorial processes include tokens in which two disparate influencers are competing at opposite ends of the phase window. The additive process predicts a bimodal distribution of the output timing across a large number of repeated tokens. The additive-centroid process predicts an output phasing consistently between the areas of the phase window preferred by each of the influencers. However, it's likely that there will nearly always be many synchronously operating influencers that prefer the middle range of the phase window, thus making instances of bimodality highly unlikely if an additive procedure is used to combine the effects of the influencers.

3.3.1 *Different kinds of influencers.* Gracco (1991) suggests two rationales for research focusing on developing models of motor control: 'first, there is an inherent richness and intricacity [*sic*] to even the simplest problem of sensorimotor control, and second...an implicit assumption that higher functions such as cognition are not discontinuous with the lower level sensorimotor functions that implement them' (Gracco 1991: 54). These are both very important points.

In the exploration of motor control, much attention has been given to physiologic factors determining timing. The efficiency of timing patterns has been determined in large part by physical factors such as energy and work requirements. There is an important way, however, in which speech movement differs from other types of (non-communicative) body movement. For speech, the determination of efficient movement patterns must take into account a perceiver.[6] That is, unlike control of the limbs in the study of gait, theories of articulatory coordination in speech timing must be able to account for the *communicative* goals, and hence communicative efficiency, of the movements. Is the listener in the room, in another room, hard of hearing, a non-native speaker, an infant? All these factors could conceivably influence how the articulators are coordinated. The approach outlined above incorporates *extra*linguistic influences in the same way as linguistic variables, even though the quality of their effects on timing may differ.

The proposal that auditory goals may affect intergestural coordination is in the spirit of Ladefoged *et al.* (1972) and Johnson *et al.* (1993), who have outlined an auditory theory of speech production in which speech movements are directed by auditory goals. Although acoustic influences on *timing* are not their focus, Johnson *et al.* do suggest that 'the acoustic product of speaking is the crucial determinant of the *organisation* of speech articulation' (1993: 712; emphasis added). Ohala (1990) also states that the ultimate goals in speech are acoustic-auditory events and that acoustic goals can influence timing. While I don't assume that acoustic output is the only goal appropriate to a speech production model, the admission of listener-oriented influences on speech motor control seems plausible (see for example Lindblom 1990). Kohler's (1992) comments regarding reduction are also applicable to our discussion of timing:

> Speakers have to tune their performance to... [different communicative] conditions to guarantee a successful language interaction... In order to increase the explanatory power of 'articulatory phonology' it has to be supplied with an auditory control unit, because speakers not only control gestures with regard to the physiological and articulatory potentials contained in the dynamics of sound production but also take listeners into account and adapt to their needs (Lindblom 1990), i.e. 'articulatory phonology' cannot be blind to acoustic consequences, it must be input- *and* output-oriented. (Kohler 1992: 209–210)

Of course there are options other than an 'auditory control unit' to address this concern. Theories incorporating a mechanism for speech planning that has access to a speaker's knowledge about the crucial perceptual cues of his language offer greater insight into speech timing than those that rely solely on the physical properties of the movements themselves. Löfqvist remarks that 'gestures are intentional, are made for a purpose, and have to be adapted to the environment... One important aspect of the environment during speech is the listener' (1990: 316). In the phase window framework, timing is still implemented intrinsically, i.e. in terms of the dynamic characteristics of the linguistic unit, but can be influenced by other factors as well.

We know that both linguistic and extralinguistic variables may affect speech timing. Here the word 'linguistic' is used in a very narrow sense to mean categorical or phonological. 'Extralinguistic' is used to refer to those variables that change continuously. In the proposal outlined, both types of influencers have commensurate means of affecting the phase window, although the shape of their probability distributions may differ. It is also not always clear whether a contextual influence is categorical or continuous. For example, resyllabification of the second consonant in a coda cluster to a following syllable might be a matter of *degree* or might be *all or nothing*. Conversely, changes in speech style and rate might behave in categorical fashions or vary gradiently. Allowing linguistic and extralinguistic factors to influence speech timing through the functioning of a single mechanism is preferred both by Occam's Razor and because it seems in agreement with the difficulties phoneticians have had in delimiting separate, non-interacting influences of such factors.

Finally, recall that Browman & Goldstein (1990b) infer a phasing relationship for consonant sequences based on their observation that in X-ray microbeam data for three speakers, C2 seemed to start at 290° (i.e. the release) of C1. Among the clusters examined are four onset clusters and four heterosyllabic clusters, most of which start with [s], in an [i__a] environment (the number of tokens considered isn't specified). If this preference does in fact extend to other clusters, syllabic positions, speakers and environments, their observation might be captured here in one of two possible ways. The first possibility is that a motor or perceptual advantage proffered by this phasing relationship creates a strong, narrow weighting in this part of the phase window such that the combined preferences of other variables are generally much less strong. This would result in a high probability of the consonant sequence being realised with this phasing relationship. The second, and to my mind, more appealing possibility is that the *typical* probability distributions for the contextual variables in their data set combine to yield a high probability peak around 290° in the phase window. This would result in an empirical bias for this phasing. However we should remember that such a bias may or may not be found for other consonants, syllabic positions, vowel contexts or speakers. Indeed, data in Byrd (1996), Byrd & Tan (1996) and Hardcastle (1985) suggest that it is not consistently found.

In summary, this framework for speech timing has two elements – phase windows and influencers. It is hypothesised that only a small number of phase windows exist, for example of the type consonant-to-vowel (reflexively), consonant-to-consonant and vowel-to-vowel. This assumption is desirable in a research programme designed to determine a small number of controllable parameters in speech production and their relation to phonological structure. The myriad factors that have been shown to influence articulatory timing have no place in phonological representation; in fact many of them are commonly considered to be extralinguistic. The postulation of a small number of very general phase windows is made preliminarily as the most conservative assumption with which to guide the research programme.

However, the question arises of whether the influencers acting alone, that is, unconstrained by any limits on variability, would not be sufficient to capture the articulatory patterns. Ideally, to demonstrate the usefulness of phase windows one would want to show that the variability in phasing is less for some types of gestures (VC, for example) than for some other type (CC, for example) in the same context (same rate, prosodic position, etc.). It is hoped that such experimental data can be evaluated in the future. The phase window acts to limit the temporal compressibility or disassociation of gestures. It prevents the joint influence of contextual variables from going beyond a certain range. The window acts as the absolute (language-specific) limit beyond which no amount of influence can have an effect. By distinguishing windows, i.e. ranges, from influencers, a given influencer can have a consistent effect on all phase windows. The shape of the influencer may be constant, but different phasings will result depending on the particular window to which it's applied (as well as depending on other influencers of course). Despite the general nature of an influencer's effect, the relative coordination of the gestures cannot be forced beyond the constraints imposed by the particular phase window required in the case of those gestures. For example, when we talk our fastest, adjacent consonantal gestures do not increase so much in overlap that they continue to 'slide' right by one another; nor, when we talk our slowest, do epenthetic vowels appear between adjacent consonants. There are limits on phasing variability.

3.4 Programs for modelling speech timing

There are of course many possible routes that can be taken in modelling speech timing. The probabilistic framework described above is one avenue for exploration. I am sympathetic with Diehl's impression (1991) that phonological and phonetic knowledge might best be understood by research strategies employing a probabilistic approach. He comments that 'in most cases scientists must at least provisionally settle for probabilistic forms [of explanation], because the full intricate skein of laws and relevant conditions is not completely known' (1991: 129). The proposal outlined above responds to Gracco's view that 'perhaps speech perception and production should be appropriately represented as stochastic processes based on probability statements implemented through an adequate but imprecise control system. Strict determinism, invariance, and precision are most likely relegated to man-made machines working under rigid tolerance limits or simplified specifications, not to complex biological systems' (Gracco 1992: 20). However, a probabilistic approach to speech timing is not the only way to allow contextual variables to affect intergestural coordination. One promising alternative model of intergestural timing is being pursued by Saltzman and his colleagues Mitra, Hogden, Levy and Rubin (Saltzman 1995a, b; Saltzman & Munhall 1989). Their work is attempting to integrate task dynamics with intergestural timing, thereby extending the dynamics now intrinsic to the interarticulator level of Articulatory Phonology to the intergestural level. Saltzman (1995a: 85) describes the use of the recurrent, sequential network architecture of Jordan (1986) as a means of 'patterning gestural activation trajectories for the task-dynamic model', thus implementing intergestural timing dynamically. The hope is that this approach will allow rate and boundary effects to be accurately modelled. If accomplished, this would represent an appealing advance.

4 Lexically specified timing

To this point, I have discussed only intergestural relations of the V-to-C, V-to-V or C-to-C type – that is, those relations typically considered to obtain between the articulations of different segments. I have argued that the phase window framework is useful in capturing the timing variability observed in the coordination of these gestures. Under this approach the precedence relations between the gestures is considered to be specified lexically, but the small number of phase windows is argued to be implemented postlexically, operating, for example, on the output of the linguistic gestural model of Articulatory Phonology.

However, there are many temporal relations that are crucial in making a phonological contrast. For example, the differences between /p/ and /p^h/ or /m͡b/ and /m/ lie significantly in the timing relationship between an oral and a non-oral gesture – glottal opening in the first case and velum lowering in the second. Articulatory Phonology makes no distinction in the mechanism for coordinating two such gestures as compared to coordinating gestures, for example, between words. No differences in variability are predicted, and there is no theory-internal reason to expect the degree of overlap to be different.

In §2.2 it was claimed that some intergestural coordinations have been found to be relatively stable; these are the coordinations between gestures that are traditionally considered to constitute a segment. These particular coordinations were excluded from the general discussion of the phase window framework above. It's unlikely to be coincidental that stable timing relations have been found for those pairs of gestures that are traditionally considered to belong to a single segment, but not between those of different segments. It has been suggested within Articulatory Phonology that syllable structure is a characteristic pattern of gestural coordination (Browman & Goldstein 1995a; see also Kelso *et al.* 1986a). I extend this and consider that the percept and functionality of the segmental unit, to whatever extent it exists, results from *its* characteristic pattern of coordination. I propose that this characteristic pattern of coordination is stability, i.e. a narrow phase window that is lexically specified. An independent speculation by Nittrouer *et al.* (1988) hypothesises that it may be the case that intergestural overlap is more stable within segments than between segments; I infer that they mean that the presence of a segment will cause timing of its gestures to be stable. Similarly, Löfqvist (1991: 346) offers the possibility that 'gestures forming a segment may show a greater degree of internal stability in the form of coherence of patterns of muscular activity and/or movement than those associated with different segments'. My proposal is crucially different – it is not the case that the quality of being a segment causes stable timing, but rather that stable timing causes the quality of being a segment. Thus, segmenthood might be epiphenomenal.[7]

This approach has more in common with Saltzman & Munhall (1989: 365), who 'assume that gestures cohere in bundles corresponding, roughly, to traditional segmental descriptions, and that these segmental units maintain their integrity in fluent speech'. They hypothesise that this cohesion is due to dynamical coupling of the gestures. It is reasonable to assume that an appropriate type of coupling could yield a limited window of relative phase relations between the coupled gestures.

Like Articulatory Phonology, this discussion sees the gesture as a basic phonological primitive. In contrast to Articulatory Phonology, this approach to intergestural coordination proposes that only certain timing relationships are lexically specified. These lexically specified phase windows are narrow, that is, very tightly constrained in the variability they allow. Furthermore, it is claimed that the gestures whose coordination, or phase window, is part of a word's lexical representation bear a close relation to those conglomerates of gestures that constitute what is traditionally considered to be a 'segment'. Crucially, the narrowness of the phase window, or constraint on permissible coordinations, ensures that the recoverability of the gestures is preserved for the listener by this tight cohesiveness and not jeopardised by an inappropriate degree of overlap. Certain aspects of gestural structure might not be recoverable unless coordinated in a specific way. For example, in certain cases a precise temporal coordination may be necessary to yield aerodynamic properties that typify a sound, such as ingressive airflow in a click (see Mattingly 1981 for an insightful discussion of the importance of intergestural coordination in accounting for restrictions on gestural overlap and perceived sequential order of phonetic elements).

The postulation of certain narrow lexical phase windows makes several predictions. This approach to the nature of 'segmenthood' predicts that a doubly articulated stop like [k͡p] should have a very stable timing, while a [kp] sequence should be more variable in timing.[8] Although few articulatory movement data have been gathered on doubly articulated stops (but see Maddieson 1993), the descriptions of these stops in the phonetic literature do seem to support the prediction of stable timing. Westermann & Ward (1933: 58) state in discussing their production that 'the two articulations *must* be simultaneous' (cited in Maddieson 1993; emphasis added). While the simultaneity (or lack thereof, cf. Maddieson & Ladefoged 1989) isn't relevant here, the use of 'must' suggests a stable timing pattern. Other similar cases of relatively stable coordination within 'segments' are also predicted. (i) The coordination of labial and velar gesture in English [w] should be less variable than the English sequence of [kp] (although the latter differs in constriction degree from the glide). (ii) The timing of the closure and constriction gestures constituting an affricate should be less variable than a comparable stop–fricative sequence. (iii) The coordination of larynx raising/lowering gestures with oral gestures in ejectives/implosives should be more stable than the co-ordination of those gestures with adjacent vowels. (iv) The same would be expected of the timing of tongue backing and oral closure gestures in clicks. Other multigesture 'segments' to consider are the labial, pharyngeal and tongue tip gestures used in certain productions of American [ɹ] (Uldall 1958; Delattre & Freeman 1968; Lindau 1985), velum and oral gestures in nasals, and tongue backing and tip raising in [l].

This approach also predicts that interword timing (by definition postlexical) will not exhibit the stability characteristic of lexically contrastive timing relationships. Additionally, it suggests that if a timing relationship were to become diachronically more and more stable, it would be likely to be lexicalised by the language learner. (We should also note that in at least one case, a regular bimodal pattern of coordination has been interpreted by linguists as resulting in two allophones of a single 'segment', the case in point being [l] *vs.* [ɫ]: Delattre 1971; Sproat & Fujimura 1993.[9]) In terms of language acquisition, it would not be surprising if language learners lexicalise the stable timing relationships to which they are exposed; but learn as general principles of speech coordination the relationships that are systematically influenced by a variety of factors.

Lastly, the proposal outlined here predicts that while certain *postlexical* timing relationships may appear more stable than others (recall the results summarised above for the onset cluster), their timing will still be affected by other influencers, like rate, in a way that a *lexical* timing relationship will not be, due to the difference in the width of the phase window that constrains the range over which influencers may have an effect. That is, there is a difference between a narrow phase window and a narrow influencer. Put more succinctly, this approach predicts that the relative timing of gestures constituting a 'segment' will be less affected by contextual variables than that of other gestures not constituting a 'segment'. That is, a smaller effect of rate or other variables is expected on the phase relations of gestures having a lexical phase window as compared to a greater influence on gestures that are only *associated* (i.e. have their precedence relation specified) lexically.

Some findings that would argue *against* the proposal outlined above include: (i) comparable variability in the relative timing of all associated gestures – for example, comparable variability between oral and glottal gestures for stops or tip and body gestures for clicks as compared to V-C, V-V or C-C timing; and (ii) greater or comparable effects of influencers such as speech rate or phrasal accent on the relative timing of the postulated lexically specified relationships as compared to those postulated to be specified by postlexical phase windows. Regarding (i), the data presently available (see §2.2) suggest that there are in fact systematic differences in variability. Additional data of interest are provided by Saltzman *et al.* (1995), whose perturbation experiments examining lip and larynx gestures find greater intergestural temporal stability within segments than between syllables (see also Saltzman *et al.* in preparation). Further data are needed to evaluate (ii).

Browman & Goldstein (1991) express the hope that lexically contrastive patterns of gestural overlap can be understood by extending recent research on bimanual rhythmic movements that has demonstrated stable coordinative modes (citing Kay *et al.* 1987; Turvey *et al.* 1986). Additionally, they suggest that critical differences in amount of overlap may yield qualitatively different acoustic, and hence perceptual, consequences. These are among the mechanisms by which they plan to partition the 'potential' continuum of overlap. In the case of contrastive phasing, I agree with Browman & Goldstein (1991) and Goldstein (1989) that quantal perceptual effects and natural oscillatory modes will determine the types of contrastive phase windows that may exist. (This is a separate issue from the supposition that phase windows exist and that they are narrow.) Clements (1992) remarks that phasing relations as currently formulated in Articulatory Phonology predict more 'types' of lexical contrasts than are actually attested. Neither does the proposal outlined here make any predictions regarding the number or 'type' of lexically contrastive timing relationships that might be possible. It only predicts that they share a common quality – stability. In fact, these frameworks *shouldn't* make such a prediction. As properly put, this question is not a structural (representational) question but a functional question. That is, these types of gestural organisations are limited by the dynamic and perceptual systems that produce and perceive them. Lindblom (1990) emphasises that speech motor control is prospectively organised to allow discriminability and recoverability by the listener, as well as efficiency on the part of the speaker. Browman & Goldstein (1991) note that lexically contrastive patterns of gestural overlap are 'designed' to allow gestural recovery by a listener. Specifically, I agree with Goldstein (1989, see also 1990) that quantal (Stevens 1989) perceptual effects are important in determining contrastive timing relationships but not in actively constraining other types of intergestural timing.

5 Conclusion

The implementation of articulatory timing within the Articulatory Phonology framework was discussed. It was argued that systematic variability in intergestural timing necessitates a means of allowing linguistic and extralinguistic variables to influence phasing relations. I propose the phase window framework, in which variability is allowed in the assignment of phasing relations. Competing influencers that differ from utterance to utterance weight a phase window, determining where in the range of permissible overlap relationships a token is likely to be realised. Additionally, it is hypothesised that certain phasing relations are lexically specified and stable. It is argued that gestures having such a lexical phasing relation bear a close correspondence to what has traditionally been considered a segment. It is hypothesised that the percept and functionality of what has traditionally been called a segment result from this characteristic stable timing, i.e. a narrow phase window that is lexically specified.

This approach to intergestural timing argues for the pursuit of a research programme designed to identify the factors influencing phasing relations and the nature of these effects. The main goals of this presentation were to offer a framework in which variability is allowed but constrained and, to some degree, predictable, and to motivate discussion as to the nature of segmenthood and its relation to intergestural timing.

NOTES

* The author is very grateful to Catherine Best, Edward Flemming, Oliver K. Foellmer, Louis Goldstein, Abigail Kaun, Patricia Keating, Peter Ladefoged, Ian Maddieson, Elliot Saltzman, Donca Steriade and several anonymous reviewers for their insightful comments, advice and discussion. This work is largely based on the author's PhD dissertation (Byrd 1994). This research and writing was supported by an NSF Graduate Fellowship, by an NSF Dissertation Improvement Grant #DBS 9213604 to the author and Patricia Keating, by the Department of Linguistics at UCLA and by NIH grant DC-00016.

[1] The concept of association remains somewhat vague, as demonstrated by suggestions such as that a 'statement' of ambisyllabicity 'applies' that associates a coda consonant to the vowel of a following word, thereby forcing the 'reapplication' of the phasing rule that phases a vowel to the leftmost preceding associated consonant (Browman & Goldstein 1990b).

[2] Structure which is associated with multiple gestures in Articulatory Phonology includes nodes on a rhythmic tier (Browman & Goldstein 1990b) and gestural constellations. Browman & Goldstein (1986) describe 'the phonological structure of a lexical item as a "constellation" of gestures, that is, a stable organisation among gestures' (see also Browman & Goldstein 1990b). In other references a constellation is 'syllable-sized' and associated with a stress node on the rhythmic tier, e.g. Browman & Goldstein (1990b: 351).

[3] Regarding the acquisition of speech timing by children, two points bear mention. First, Nittrouer (1993) notes that young speakers may exhibit greater overlap among articulatory gestures compared to more experienced speakers (Goodell & Studdert-Kennedy 1993; Nittrouer *et al.* 1989). Nittrouer (1993), citing Kent (1983), suggests that the initiation of gestures composing a syllable might take place synchronously for a child who has not yet established phasing relationships between the gestures. Kent comments that 'synchronous patterning may be a default principle that is overridden by phonetic and motor learning' (1983: 71). Second, research suggests that a child's 'loosely coordinated gestures [evolve into] the tightly coordinated patterns of articulatory movement characteristic of adult speech' and that 'young speakers may exhibit great variation in (perceived) phonetic structure across attempts at the same utterance' (Nittrouer 1993: 960). On greater timing variability in children's speech, see also Kewley-Port & Preston (1974); Tingley & Allen (1975); Kent & Forner (1980); Sharkey & Folkins (1985); Chermak & Schneiderman (1985); Kent (1986); Katz *et al.* (1991). This gives some support to the idea that a speaker in learning to coordinate his speech movements is narrowing in on acceptable phase windows

for particular relationships. We might speculate on how the data on child speech timing acquisition could be understood in the phase window framework. First, before phase windows have been created, i.e. before phase relations are implemented, children may be likely to initiate gestures (composing some unit like a syllable/word) synchronously. That is, they may know what movements are involved but not have established any organisational relationship among them. Next, after phasing relationships have been established but before they are subject to all the influences and limitations exhibited in adult timing, the coordination between gestures may be highly variable. Later, as coordination becomes more mature, systematic influences on phase relations are expected to yield more consistency in the timing of the phased gestures, although no single phasing relationship is required.

[4] This framework implies that an intergestural timing relationship realised with a coordination outside the constraints of the appropriate phase window would be perceived as 'abnormal' by a listener (e.g. possible examples of this might include foreign accented speech or apraxic speech).

[5] Simply adding the distributions is in fact unrealistic, as it would ultimately yield a flat line output.

[6] We would not wish to suggest that work such as that of Kelso *et al.* (1986a) does not admit non-motoric constraints on speech timing (see Kelso *et al.* 1986b), but rather that their research has focused on obtaining a thorough understanding of the motoric constraints.

[7] While this discussion concerns speech production, it remains an open question whether the segment acts as a unit in speech perception.

[8] Note that Maddieson (1993) finds that the underlying [k] and [p] gestures are probably the same in both [k͡p] and [kp], with any movement trajectory differences being due mainly to aerodynamic conditions.

[9] Let's digress for a moment to consider two cases of reported 'bistability', i.e. two stable patterns of coordination. Krakow (1989), examining nasal consonant production for two speakers, found that in syllable-initial position the oral and nasal gestures were reliably coordinated in a roughly synchronous relationship while in syllable-final position the velum lowering gesture consistently preceded the oral closure, again in a consistent fashion. Sproat & Fujimura (1993) considered movement data from five speakers' [l] productions. They found that in word-initial [l]'s the tongue tip peak led the body lowering while in word-final position the body lowering led the peak. (Note that Goldstein (1994) describes these data as paralleling Krakow's (1989) – that is, that the tip and body are roughly synchronous in word-initial [l]. The 'tip delay' measure reported in Sproat & Fujimura (1993: Table III) suggests at least for peak movements that neither word-initial nor word-final [l]'s were consistently synchronous, only that there is a consistent difference in which articulator led.) These examples of 'bistability' suggest that the intergestural coordination of such 'intrasegment' gestures should be characterised differently from the 'intersegment' phase relations described in §2.2, where a range of variability in the timing between vowel and consonant, or consonant and consonant oral gestures, is observed. The marked difference in these behaviours – stability (even if specific to syllable position) for [n] and [l] *vs.* a range of variation for VCV or CC coordination) – lends support to the hypothesis that there may be two types of timing subject to different constraints, one stable type specified lexically and one variable type implemented postlexically.

REFERENCES

Barry, M. (1985). A palatographic study of connected speech process. *Cambridge Papers in Phonetics and Experimental Linguistics* 4. 1–16.

Barry, W. J. (1983). Some problems of interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance* 9. 826–828.

Bell-Berti, F. & L. J. Raphael (eds.) (1995). *Producing speech: contemporary issues for Katherine Safford Harris*. New York: AIP Press.

Benoit, C. (1986). Note on the use of correlations in speech timing. *JASA* **80**. 1846–1849.

Browman, C. (1995). Assimilation as gestural overlap: comments on Holst and Nolan. In Connell & Arvaniti (1995). 334–342.

Browman, C. & L. Goldstein (1986). Toward an articulatory phonology. *Phonology Yearbook* **3**. 219–252.

Browman, C. & L. Goldstein (1988). Some notes on syllable structure in articulatory phonology. *Phonetica* **45**. 140–155.

Browman, C. & L. Goldstein (1989). Articulatory gestures as phonological units. *Phonology* **6**. 201–251.

Browman, C. & L. Goldstein (1990a). Gestural specification using dynamically-defined articulatory structures. *JPh* **18**. 299–320.

Browman, C. & L. Goldstein (1990b). Tiers in articulatory phonology, with some implications for casual speech. In Kingston & Beckman (1990). 341–376.

Browman, C. & L. Goldstein (1991). Gestural structures: distinctiveness, phonological processes, and historical change. In I. Mattingly & M. Studdert-Kennedy (eds.) *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Erlbaum. 313–338.

Browman, C. & L. Goldstein (1992a). Articulatory phonology: an overview. *Phonetica* **49**. 155–180.

Browman, C. & L. Goldstein (1992b). Response to commentaries. *Phonetica* **49**. 222–234.

Browman, C. & L. Goldstein (1995a). Gestural syllable position effects in American English. In Bell-Berti & Raphael (1995). 19–33.

Browman, C. & L. Goldstein (1995b). Dynamics and articulatory phonology. In Port & van Gelder (1995). 175–193.

Byrd, D. (1992). Perception of assimilation in consonant clusters: a gestural model. *Phonetica* **49**. 1–24.

Byrd, D. (1994). *Articulatory timing in English consonant sequences*. PhD dissertation, UCLA.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *JPh* **24**. 209–224.

Byrd, D., A. R. Kaun & S. Narayanan (1996). Prosodic boundary effects in Tamil: an articulatory study. Paper presented at the 70th Annual Meeting of the Linguistic Society of America, San Diego.

Byrd, D. & C. C. Tan (1996). Saying consonant clusters quickly. *JPh* **24**. 263–282.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Bloomington: Indiana University Press.

Chermak, G. D. & C. R. Schneiderman (1985). Speech timing variability of children and adults. *JPh* **13**. 477–480.

Clements, G. N. (1992). Phonological primes: features or gestures. *Phonetica* **49**. 181–193.

Connell, B. & A. Arvaniti (eds.) (1995). *Phonology and phonetic evidence: papers in laboratory phonology IV*. Cambridge: Cambridge University Press.

Delattre, P. (1971). Consonant gemination in four languages: an acoustic, perceptual, and radiographic study. Part I. *IRAL* **9**. 31–52.

Delattre, P. & D. Freeman (1968). A dialect study of American R's by x-ray motion picture. *Linguistics* **44**. 29–68.

Diehl, R. L. (1991). The role of phonetics within the study of language. *Phonetica* **48**. 120–134.

Docherty, G. J. (1992). *The timing of British English obstruents*. Berlin: Foris.

Docherty, G. J. & D. R. Ladd (eds.) (1992). *Papers in laboratory phonology II: gesture, segment, prosody*. Cambridge: Cambridge University Press.

Elenius, K. & P. Branderud (eds.) (1995). *Proceedings of the 13th International Congress of Phonetic Sciences*. Vol. 3. Stockholm: Congress Organizers at KTH and Stockholm University.

Flemming, E., P. Ladefoged & S. Thomason (1993). The phonetic structures of Montana Salish. *UCLA Working Papers in Phonetics* **87**. 1–34.

Goldstein, L. (1989). On the domain of the quantal theory. *JPh* **17**. 91–97.

Goldstein, L. (1990). On articulatory binding: comments on Kingston's paper. In Kingston & Beckman (1990). 445–450.

Goodell, E. W. & M. Studdert-Kennedy (1993). Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: a longitudinal study. *Journal of Speech and Hearing Research* **36**. 707–727.

Gracco, V. L. (1988). Timing factors in the coordination of speech movements. *Journal of Neuroscience* **8**. 4628–4634.

Gracco, V. L. (1991). Sensorimotor mechanisms in speech motor control. In H. F. M. Peters, W. Hulstijn & C. W. Starkweather (eds.) *Speech motor control and stuttering*. Amsterdam: Elsevier. 53–76.

Gracco, V. L. (1992). Characteristics of speech as a motor control system. *Haskins Laboratories Status Report on Speech Research* **SR-109/110**. 13–26. To appear in G. Hammond (ed.) *Cerebral control of speech and limb movements*. Amsterdam: Elsevier.

Gracco, V. L. & J. H. Abbs (1989). Sensorimotor characteristics of speech motor sequences. *Experimental Brain Research* **75**. 586–598.

Hardcastle, W. J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication* **4**. 247–263.

Hardcastle, W. J. & A. Marchal (eds.) (1995). *Speech production and speech modelling*. Kluwer: Dordrecht.

Hardcastle, W. J. & P. Roach (1979). An instrumental investigation of coarticulation in stop consonant sequences. In H. Hollien & P. Hollien (eds.) *Current issues in the phonetic sciences*. Amsterdam: John Benjamins. 531–540.

Hawkins, S. (1992). An introduction to task dynamics. In Docherty & Ladd (1992). 9–25.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Holst, T. & F. Nolan (1995). The influence of syntactic structure on [s] and [ʃ] assimilation. In Connell & Arvaniti (1995). 315–333.

Johnson, K., P. Ladefoged & M. Lindau (1993). Individual differences in vowel production. *JASA* **94**. 701–714.

Jones, D. (1956). *An outline of English phonetics*. 8th edn. Cambridge: Heffer.

Jong, K. de (1991). An articulatory study of consonant-induced vowel duration changes in English. *Phonetica* **48**. 1–17.

Jordan, M. (1986). *Serial order in behaviour : a parallel distributed processing approach*. San Diego: Institute for Cognitive Science, University of California.

Katz, W., C. Kripke & P. Tallal (1991). Anticipatory coarticulation in the speech of adults and young children: acoustic, perceptual, and video data. *Journal of Speech and Hearing Research* **34**. 1222–1232.

Kay, B. A., J. A. S. Kelso, E. Saltzman & G. Schöner (1987). Space-time behavior of single and bimanual rhythmical movements: data and limit cycle model. *Journal of Experimental Psychology : Human Perception and Performance* **13**. 178–192.

Keating, P. A. (1990a). Phonetic representations in a generative grammar. *JPh* **18**. 321–334.

Keating, P. A. (1990b). The window model of coarticulation: articulatory evidence. In Kingston & Beckman (1990). 451–470.

Keating, P. A. (1995). Segmental phonology and non-segmental phonetics. In Elenius & Branderud (1995). 26–32.

Keller, E. (1987). The variation of absolute and relative measures of speech activity. *Journal of Speech and Hearing Research* **30**. 223–229.

Keller, E. (1990). Speech motor timing. In Hardcastle & Marchal (1990). 342–364.

Kelso, J. A. S., E. Saltzman & B. Tuller (1986a). The dynamical perspective on speech production: data and theory. *JPh* **14**. 29–59.

Kelso, J. A. S., E. Saltzman & B. Tuller (1986b). Intentional contents, communicative context, and task dynamics: a reply to commentators. *JPh* **14**. 171–196.

Kelso, J. A. S. & B. Tuller (1987). Intrinsic time in speech production: theory, methodology, and preliminary observations. In E. Keller & M. Gopnik (eds.) *Sensory and motor processes in language*. Hillsdale, NJ: Erlbaum. 203–222.

Kelso, J. A. S., E. Vatikiotis-Bateson, E. Saltzman & B. Kay (1985). A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling. *JASA* **77**. 266–280.

Kent, R. D. (1983). The segmental organization of speech. In P. F. MacNeilage (ed.) *The production of speech*. New York: Springer. 57–89.

Kent, R. D. (1986). Is a paradigm change needed? *JPh* **14**. 111–115.

Kent, R. D. & L. L. Forner (1980). Speech segment duration in sentence recitations by children and adults. *JPh* **8**. 157–168.

Kewley-Port, D. & M. S. Preston (1974). Early apical stop production: a voice onset time analysis. *JPh* **2**. 195–210.

Kingston, J. & M. E. Beckman (eds.) (1990). *Papers in laboratory phonology I: between the grammar and physics of speech*. Cambridge: Cambridge University Press.

Kohler, K. J. (1992). Gestural reorganization in connected speech: a functional viewpoint on 'articulatory phonology'. *Phonetica* **49**. 205–211.

Kosko, B. (1993). *Fuzzy thinking*. New York: Hyperion.

Kozhevnikov, V. & L. Chistovich (1965). *Speech: articulation and perception*. Translated and distributed by Joint Publications Research Service, Washington D.C. Originally published in Russian, 1965.

Krakow, R. A. (1989). *The articulatory organization of syllables: a kinematic analysis of labial and velar gestures*. PhD dissertation, Yale University.

Kuehn, D. P. & K. L. Moll (1976). A cineradiographic study of VC and CV articulatory velocities. *JPh* **4**. 303–320.

Ladefoged, P., J. DeClerk, M. Lindau & G. Papcun (1972). An auditory-motor theory of speech production. *UCLA Working Papers in Phonetics* **22**. 48–75.

Lindau, M. (1985). The story of /r/. In V. A. Fromkin (ed.) *Phonetic linguistics: essays in honor of Peter Ladefoged*. Orlando, Fl.: Academic Press. 157–168.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H & H theory. In Hardcastle & Marchal (1990). 403–439.

Löfqvist, A. (1986). Stability and change. *JPh* **14**. 139–144.

Löfqvist, A. (1990). Speech as audible gesture. In Hardcastle & Marchal (1990). 289–322.

Löfqvist, A. (1991). Proportional timing in speech motor control. *JPh* **19**. 343–350.

Löfqvist, A. & H. Yoshioka (1981). Interarticulator programming in obstruent production. *Phonetica* **38**. 21–34.

Löfqvist, A. & H. Yoshioka (1984). Intrasegmental timing: laryngeal–oral coordination in voiceless consonant production. *Speech Communication* **3**. 279–289.

Lubker, J. (1986). Articulatory timing and the concept of phase. *JPh* **14**. 133–137.

McLean, M. (1973). Forward coarticulation of velar movement at marked junctural boundaries. *Journal of Speech and Hearing Research* **16**. 286–296.

MacNeilage, P. (1970). Motor control of serial ordering of speech. *Psychological Review* **77**. 182–196.

Maddieson, I. (1993). Investigating Ewe articulations with electromagnetic articulography. *Forschungsberichte, Institut für Phonetik und Sprachliche Kommunikation* **31**. 184–214.

Maddieson, I. & P. Ladefoged (1989). Multiply-articulated segments and the feature hierarchy. *UCLA Working Papers in Phonetics* 72. 116–138.

Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver & J. Anderson (eds.) *The cognitive representation of speech.* Amsterdam: North-Holland. 415–420.

Mermelstein, P. (1973). Articulatory model for the study of speech production. *JASA* 53. 1070–1082.

Munhall, K. G. (1985). An examination of intra-articulatory relative timing. *JASA* 78. 1548–1553.

Munhall, K. G., A. Löfqvist & J. A. S. Kelso (1986). Laryngeal compensation following sudden oral perturbation. *JASA* 80. S109.

Munhall, K. G., D. J. Ostry & A. Parush (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology* 4. 457–474.

Nittrouer, S. (1991). Phase relations of jaw and tongue tip movements in the production of VCV utterances. *JASA* 90. 1806–1815.

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: evidence from an acoustic study. *Journal of Speech and Hearing Research* 36. 959–972.

Nittrouer, S., K. G. Munhall, J. A. S. Kelso, B. Tuller & K. S. Harris (1988). Patterns of interarticulator phasing and their relation to linguistic structure. *JASA* 84. 1653–1661.

Nittrouer, S., M. Studdert-Kennedy & R. S. McGowan (1989). The emergence of phonetic segments: evidence from the spectral structure of fricative–vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research* 32. 120–132.

Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In Docherty & Ladd (1992). 261–289.

Ohala, J. (1990). The generality of articulatory binding. In Kingston & Beckman (1990). 435–444.

Ohala, J., S. Hiki, S. Hubler & R. Harshman (1968). Photoelectric methods of transducing lip and jaw movements in speech. *UCLA Working Papers in Phonetics* 10. 135–144.

Ostry, D. J., E. Keller & A. Parush (1983). Similarities in the control of the speech articulators and the limbs: kinematics of tongue dorsum movements in speech. *Journal of Experimental Psychology: Human Perception and Performance* 9. 622–636.

Ostry, D. J. & K. G. Munhall (1985). Control of rate and duration of speech movements. *JASA* 77. 640–648.

Pierrehumbert, J. (1994). Syllable structure and word structure: a study of triconsonantal clusters in English. In P. Keating (ed.) *Phonological structure and phonetic form: papers in laboratory phonology III.* Cambridge: Cambridge University Press. 168–187.

Port, R. F. & T. van Gelder (eds.) (1995). *Mind as motion: explorations in the dynamics of cognition.* Cambridge, Mass.: MIT Press.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: a preliminary model. In H. Heuer & C. Fromm (eds.) *Generation and modulation of action patterns.* New York: Springer. 129–144.

Saltzman, E. (1995a). Intergestural timing in speech production: data and modeling. In Elenius & Branderud (1995). 84–91.

Saltzman, E. (1995b). Dynamics and coordinate systems in skilled sensorimotor activity. In Port & van Gelder (1995). 149–173.

Saltzman, E. & J. A. S. Kelso (1987). Skilled actions: a task dynamic approach. *Psychological Review* 94. 84–106.

Saltzman, E., A. Löfqvist, B. Kay, J. Kinsella-Shaw & P. Rubin (in preparation). Dynamics of intergestural timing: a perturbation study of lip–larynx coordination.

Saltzman, E., A. Löfqvist, J. Kinsella-Shaw, B. Kay & P. Rubin (1995). On the dynamics of temporal patterning in speech. In Bell-Berti & Raphael (1995). 469–487.

Saltzman, E. & K. G. Munhall (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology* **1**. 333–382.

Shaiman, S., S. G. Adams & M. D. Z. Kimelman (1995). Timing relationships of the upper lip and jaw across changes in speaking rate. *JPh* **23**. 119–128.

Shaiman, S. & R. J. Porter Jr. (1991). Different phase-stable relationships of the upper lip and jaw for production of vowels and diphthongs. *JASA* **90**. 3000–3007.

Sharkey, S. G. & J. W. Folkins (1985) Variability of lip and jaw movements in children and adults: implications for the development of speech motor control. *Journal of Speech and Hearing Research* **28**. 8–15.

Sock, R. & O. Jah (1986). Old wine in new bottles, new wine in old bottles: action theory and (para)metric phonology views in the domain of Wolof. *Bulletin de l'Institut de Phonétique de Grenoble* **15**. 117–133.

Sproat, R. & O. Fujimura (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *JPh* **21**. 291–312.

Stevens, K. N. (1989). On the quantal nature of speech. *JPh* **17**. 3–45.

Sussman, H. M., P. F. MacNeilage & R. J. Hanson (1973). Labial and mandibular dynamics during the production of bilabial consonants: preliminary observations. *Journal of Speech and Hearing Research* **16**. 397–420.

Tingley, B. M. & G. D. Allen (1975). Development of speech timing control in children. *Child Development* **46**. 186–194.

Tuller, B. & J. A. S. Kelso (1984). The timing of articulatory gestures: evidence for relational invariants. *JASA* **76**. 1030–1036.

Tuller, B., J. A. S. Kelso & K. S. Harris (1982). Interarticulator phasing as an index of temporal regulatory in speech. *Journal of Experimental Psychology: Human Perception and Performance* **8**. 460–472.

Tung, T. (1964). *A descriptive study of the Tsou language, Formosa*. Taipei: Institute of History and Philology.

Turvey, M. T. (1990). Coordination. *American Psychologist*. August 1990. 938–953.

Turvey, M. T., L. D. Rosenblum, P. N. Kugler & R. C. Schmidt (1986). Fluctuations and phase symmetry in coordinated rhythmic movements. *Journal of Experimental Psychology: Human Perception and Performance* **12**. 564–583.

Uldall, E. T. (1958). American 'molar' *r* and 'flapped' *r*. *Revista do Laboratorio de Fonética Experimental (Coimbra)* **4**. 103–106.

Vatikiotis-Bateson, E. & J. A. S. Kelso (1993). Rhythm type and articulatory dynamics in English, French and Japanese. *JPh* **21**. 231–265.

Westermann, D. & I. C. Ward (1933). *Practical phonetics for students of African languages*. Oxford: Oxford University Press, for the International African Institute.