563

Phonology Yearbook 3 (1986) 219–252
Printed in Great Britain

*Towards an articulatory phonology**

Catherine P. Browman
Louis M. Goldstein
Haskins Laboratories and Yale University

ABSTRACT
We propose an approach to phonological representation based on describing an utterance as an organised pattern of overlapping articulatory gestures. Because movement is inherent in our definition of gestures, these gestural 'constellations' can account for both spatial and temporal properties of speech in a relatively simple way. At the same time, taken as phonological representations, such gestural analyses offer many of the same advantages provided by recent nonlinear phonological theories, and we give examples of how gestural analyses simplify the description of such 'complex segments' as /s/–stop clusters and prenasalised stops. Thus, gestural structures can be seen as providing a principled link between phonological and physical description.

1 Introduction

The gap between the linguistic and physical structure of speech has always been difficult for phonological theory to bridge. Until recently, theories have encapsulated the linguistically relevant structure of speech in a sequence of segmental units, each of which corresponds to a feature bundle. Under this strict segmental hypothesis (formulated in terms of features), the sequence of feature bundles that constitute segments forms a feature matrix, whose cells are organised into non-overlapping columns. Linguistically relevant contrast between utterances, in this approach, requires that at least one feature value differ between contrasting strings. The bridge to the continuous nature of speech is made by assuming that 'each segment is characterized in terms of a state of the vocal organs, and the transitions between these states are ... predictable in terms of very general linguistic and physiological laws' (Anderson 1974:5).

This strictly linear view of the relation between linguistic units and speech has come under attack in recent years from two different directions. Phonologists have found the constraint imposed by linear sequences of non-overlapping segments to be too extreme to capture a variety of phonological facts. Recognition of the importance of allowing feature specifications to overlap was made, e.g., by Anderson (1974). He presented an alternative approach that decomposed articulation into four subsystems (an energy source, a laryngeal system, an oral system, and a nasal system), and noted that 'it is possible ... [that] the boundaries of a specification in one system will not coincide with the boundaries of a specification in another' (1974:274). The other direction of attack has come from phoneticians (e.g. Lisker 1974), who have shown the linguistic relevance of the detailed temporal structure of speech. For example, as discussed below, interarticulator temporal organisation may vary from language to language in a way that cannot be predicted (by any universal principles) from existing phonetic feature characterisations, and must thus be specified somehow in language descriptions. These developments suggest a need for a revised conception of phonological/phonetic structure, one that incorporates overlapping phonological units and one which allows temporal relations among articulatory structures to emerge from the description. We consider the two lines of attack in greater detail.

The linearity assumption has been challenged (if not completely discarded) by attempts over the last ten years to formalise more enriched conceptions of phonological structure. Like Anderson's (1974) proposal, these efforts were undertaken in response to the failure of the segmental model to account adequately for certain facts. These conceptions include explicit incorporation of syllable structure (Kahn 1976; Hooper 1972, 1976), hierarchical metrical structures (Liberman & Prince 1977; Selkirk 1980; Hayes 1981), dependency structures (Anderson & Jones 1974; Ewen 1982), independent structural or autosegmental tiers (Goldsmith 1976; Clements 1980), and explicit incorporation of a consonant–vowel skeleton (Halle & Vergnaud 1980; McCarthy 1981, 1984; Clements & Keyser 1983; Prince 1984). While these approaches have increased the range of facts that can be adequately formalised in phonological theory, they are inexplicit with respect to the relation between the revised conception of phonological structure and the physical structure of speech. The traditional link between phonological and physical structure has vanished along with strictly linear segmental analyses, and a new link has yet to be forged. It is this task we attempt in this paper, by accounting as simply as possible for the organisation of speech in both space and time. We will show that the structures that emerge from such an account can also be used as a basis for phonological description – indeed, a kind of phonological description that is much in the spirit of the above-mentioned theories.

From the phonetic side, there has been growing evidence that systematic phonetic feature representations cannot adequately describe phonetic differences among languages. Ladefoged (1980), for example, argues that the specification of features at the systematic phonetic level is neither 'necessary nor sufficient to specify what it is that makes English sound like English rather than German' (1980:495). As both Ladefoged and Anderson (1974) point out, phonetic differences between languages may involve aspects of speech that do not serve as the basis for phonological contrast within any one language. For example, Anderson shows that languages differ with respect to whether stops are released in clusters and in word final position, even though no single language contrasts released *vs.* unreleased stops. Thus, he suggests a feature [±release] to differentiate the phonetic representations in these languages.

The difference between released and unreleased stops can be seen as part of a more general problem: differences among languages in the relative timing of articulatory gestures. The 'unreleased' initial stops in clusters are, presumably, released, but only after the occlusion for the second stop has formed. There would be little acoustic evidence, therefore, of their release (cf. Catford 1977). Thus language differences in stop release may be analysed as differences in the temporal overlap of adjacent closure gestures. It is possible, in general, to describe such cross-language differences in gestural timing within the *SPE* framework (Chomsky & Halle 1968) by means of features such as [±release]. However, the potential number and variety of such differences would lead to the proliferation of features that have no contrastive function within languages. (A similar point about proliferation of phonetic features, but not specifically about timing relations, is made by Keating 1984.)

It is not difficult to find documented examples of cross-language differences in gestural timing. A number of writers (e.g. Flege & Port 1981; Port 1981; Mitleb 1984; Keating 1985) have demonstrated such differences in voicing contrasts, specifically in the duration of vowels preceding voiced and voiceless stops. While the acoustic duration of a pre-consonantal vowel is generally longer before a voiced than before a voiceless stop, the effect is larger in some languages than in others (as was earlier noted by Lehiste 1970), and can be virtually absent (e.g. in Polish, Czech and Arabic). These differences in vowel duration presumably reflect differences in the relative timing of vowel and consonant gestures. In this case, a different feature, probably [±long], would be used in an *SPE* treatment to describe cross-language differences in gestural timing.

Such phenomena are not restricted to voicing. For example, languages may also show differences in ejective consonants in the time between release of glottal closure and release of oral closure (Catford 1977:69; Lindau 1984). There is no evidence that such differences in ejectives contrast in any language, and therefore some ad hoc feature, similar to [±release], would have to be proposed to account for this difference. Finally, Fourakis (1980) has shown that the occurrence of so-called epenthetic stops in English words like *tense* is dialect-dependent. If such 'stops' are to be analysed in terms of variation in the relative timing of oral and velic gestures (rather than actual segment insertion; cf. Ohala 1974; Anderson 1976), then such timing relations are also not universal, but are a property of a particular language or dialect.

In general, then, languages can differ from one another in the timing of (roughly the same) articulatory gestures. The above examples are meant simply to illustrate the variety of phenomena that can be analysed in this way. An *SPE* characterisation of such examples not only proliferates features in the grammar; more seriously, it misses a generalisation: timing of articulatory gestures is linguistically relevant, at least in terms of how languages are distinguished from one another.

One approach to the description of language-particular patterns of articulatory timing has involved positing rules that specifically convert segmental feature matrices to temporally continuous physical parameters (e.g. Port 1981; Keating 1985). As described by Port & O'Dell (1984), such 'implementation' rules are like phonological rules in that they must be assumed to differ from language to language, but unlike such rules, they do not map feature values onto feature values. Rather they take features (or matrices thereof) as input, and output 'a complex pattern of graded commands distributed over time to the various articulators' (Port & O'Dell 1984:122). However, such rules for implementing patterns of articulatory timing have not been made explicit. This is not surprising, perhaps, since the supposed output of the rules – control parameters for articulators – requires reference to the organisation of articulatory movements. Yet no linguistic approach has provided a vocabulary for describing such organisation. In the implementation rule view, the organisation remains outside speech itself, in the segmental (and metrical, etc.) structure. Speech itself has no organisation, but is rather seen as a plastic medium that somehow serves to code the information present externally in the linguistic structure. It is worth noting that the implementation rules that have been successfully made explicit by Liberman & Pierrehumbert (1984) characterise intonation. Here, the relevant physical parameter is univariate in the acoustic domain (Fo), and no assumptions about articulatory organisation are made.

Rather than positing inexplicit implementation rules or proliferating ad hoc features, we propose to base phonological representation on an explicit and direct description of articulatory movement in space and over time. As argued by Fowler (1980), incorporation of time (in particular) into the basic definition of phonetic units can simplify much of the complex translation that is required in an approach like that of implementation rules. Moreover, describing speech in terms of overlapping (relatively invariant) articulatory units with inherent time courses can account in a simple way for observed patterns of acoustic variation (Liberman *et al.* 1967; Fowler 1980; Bell-Berti & Harris 1981; Fujimura 1981; Mattingly 1981; Liberman & Mattingly 1985). Such articulatory descriptions are particularly promising for nonlinear phonological analyses that require overlapping features, because there is a clear physical reality underlying the decomposition of articulation into quasi-independent systems whose movements are not always synchronised.

While there is, of course, a long history of referring to aspects of articulation in phonological and phonetic representations (e.g. Jespersen 1914; Pike 1943; Abercrombie 1967; Chomsky & Halle 1968; Ladefoged 1971), such representations have often been forced to rely on impressionistic descriptions, and have emphasised the static aspects of articulation.

Two recent developments in speech research make it feasible, we believe, to incorporate explicit characterisations of articulatory movement into phonological representation. The first of these is the development of improved technologies for tracking continuous articulatory movement (e.g. Fujimura *et al.* 1973), which reduces the need to rely on impressionistic observations by providing more explicit physical measurements. The second, the development of a theoretical framework for the analytical and mathematical description of coordinated movements (e.g. Bernstein 1967; Turvey 1977; Fowler *et al.* 1980; Kugler *et al.* 1980; Saltzman & Kelso 1983; Kelso & Tuller 1984a, b), simplifies the description of movement sufficiently to make it tractable for phonological purposes, and also provides a framework in which to explore generalisations regarding coordinated articulatory movements.

As we will attempt to show in this paper, an explicit description of articulatory movement can serve as the basis for phonological representation. The basic units in this framework are articulatory gestures (§ 1.1). We will first define gestures simply as characteristic patterns of movement of vocal-tract articulators, or articulatory systems, and will suggest phonological analyses of two linguistic problems in terms of these gestures and the relations among them (§ 2). Such analyses have many of the advantages found in recent nonlinear phonological theories, while at the same time providing a possible solution to the problem of the missing link between phonological and physical structures. We will then show how these characteristic patterns of movement can emerge from an abstract mathematical formalisation of gestures and the lexical structures composed of such gestures (§ 3). This mathematical formalisation of a gesture, being developed in cooperation with our colleagues at Haskins Laboratories, uses a dynamical model to explicitly characterise the coordinated patterns of articulatory movement. In this approach, based on the concept of coordinative structures (e.g. Turvey 1977) as instantiated in the task dynamic model of Saltzman & Kelso (1983), gestures are autonomous structures that can generate articulatory trajectories in space and time without any additional interpretation or implementation rules.

1.1 Gestural structure of speech

We begin our discussion of gestures with a simple example, the utterance [abə]. During this utterance, the lower lip moves up gradually toward the upper lip, reaches some peak upward displacement, and then moves downward again, as can be seen in Fig. 1. The lip is constantly in motion, except instantaneously at its maximum displacement – it does not necessarily achieve any steady-state configuration that could be associated unambiguously with the /b/. The absence of steady states characterises all observations of speech, whether articulatory or acoustic, and this may be one of the reasons why phonologists have posited a complex relation between phonology and speech. We argue, however, that it is the assump-

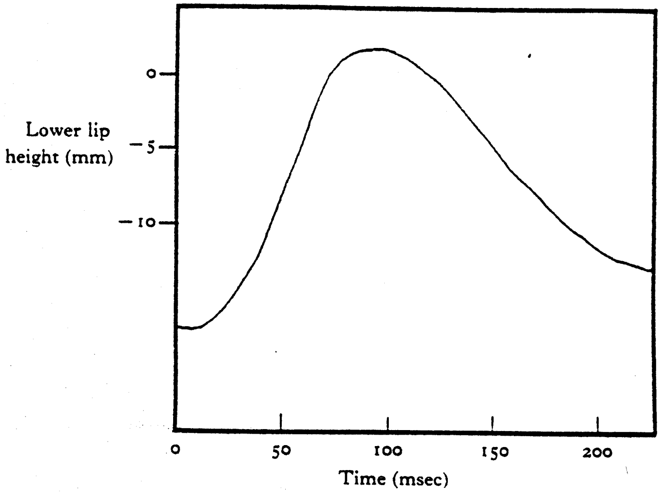

Figure 1

Trajectory of the lower lip in [abə], as measured by tracking infra-red LED placed on subject's lower lip.

tion of a steady-state specification that leads to the apparent complexity, and that a phonology which inherently incorporates movement in its descriptions will simplify much of this apparent complexity.

Instead of looking for a steady-state correlate of the /b/ segment, then, we take the trajectory of the lower lip in [abə] as a pattern in space and time that characterises utterances transcribed with a /b/ in them. That is, there is no particular spatial coordinate value of the lower lip (or any other articulator) that is held for some time and that is characteristic of /b/. Rather, it is the movement of the articulators through space over time that constitutes an organised, repeatable, linguistically relevant pattern. We can refer to this pattern as a bilabial closure gesture. (Not every utterance of a word transcribed with a /b/ will display exactly the trajectory of Fig. 1: the trajectory will vary with vowel context, syllable position, stress, speaking rate and speaker. We must, therefore, ultimately characterise a /b/ as a family of patterns of lip movement. In §3, we will suggest how this family can be formally defined using an abstract set of equations that can generate the variant trajectories. For the present, we can think of a gesture as an instance of a family of related trajectories.)

If every segment in a traditional phonological (or phonetic) representation could be described as one gesture, much as /b/ can be described as a bilabial closure gesture, then the implications of the gestural approach for phonology would be limited. However, the relation between segments and gestures is not always one-to-one. English voiced stops can, to a first approximation, be characterised as single gestures of bilabial, alveolar or velar closure; but other segments, such as the voiceless stops, require more than one gesture. In /p/, for example, we have a bilabial closure gesture much like that for /b/. In addition, however, the glottis must be opened for /p/, and then narrowed again. That is, from the point of view of spatio-temporal speech structure, /p/ is an organisation of two gestures – a bilabial closure gesture plus a glottal opening (and closing) gesture. Thus, there is no one-to-one relation between gestures and segments.

Nor do gestures bear a one-to-one relation to traditional phonological features. A single bilabial closure gesture would correspond to a number of features, such as [−continuant], [+anterior], [−coronal], [+consonantal], [−vocalic], etc. In general, differences in the presence or absence of glottal or velic (opening and closing) gestures correspond to single feature differences, while supraglottal constriction gestures correspond to multiple feature differences. Thus, gestures do not bear a one-to-one relation to either phonological segments or phonological features. Rather, they represent organised patterns of movement within oral, laryngeal and nasal articulatory systems.[1]

In addition to the gestures themselves, the relations among gestures also play an important role in the articulatory description, similar to the role of the associations among autosegments in autosegmental phonology. As an example, consider the bilabial closure and glottal opening gestures in words transcribed as beginning with /p/. These gestures are not temporally simultaneous, but repeated observations of words beginning with /p/ reveal tight spatial and temporal relations between the two gestures (Löfqvist 1980; Löfqvist & Yoshioka 1985; cf. Lisker & Abramson 1964). The incorporation of such spatio-temporal coordination within our description can be seen as having two different functions. On the one hand, the representation of a tight relation between the two gestures defines a phonological class – the class traditionally described as words beginning with /p/. On the other hand, the relation is specified in explicit enough fashion to capture the systematic, language-particular aspects of the timing between the gestures.

In the particular example of /p/, the gestural structure specified corresponds to a segment. In general, however, the interdependencies among gestures are not restricted to those that constitute single segments in traditional approaches. Rather, the pattern of relations among a set of gestures, the gestural CONSTELLATION, can serve the functions typically filled by other phonological structures, ranging from complex segments to syllables and their constituents. In §2, we show how such constellations can provide the basis for phonological analyses in cases where featural overlap has been invoked, of the type explored by Anderson (1974) and developed further within autosegmental, including CV and X-tier, phonologies.

2 Gestural analysis of two linguistic problems

As noted in the Introduction, segmental and gestural analyses differ minimally for gesturally simple segments like /b d g/. Even in such cases, however, a gestural analysis has the additional advantage that it accounts for the physical movements of the articulators as well as for the phonological structure. For gesturally more complex structures, the gestural analysis differs from a segmental analysis. In this section, we present two instances in which the analyses diverge: English /s/–stop clusters (§2.1) and prenasalised stops (§2.2). Both are examples of 'complex segments' (e.g. Ewen 1982); that is, they behave in some ways like single segments and in some ways like clusters. Using the gestural approach, we attempt to answer the question of 'one' *vs.* 'two' units by analysing the observed articulatory movements themselves. As we shall see, this gestural analysis can provide structures that allow linguistic facts to be stated more generally and simply than in a segmental analysis. We will also see how these same gestural structures can account for some of the observed patterns of timing.

2.1 Glottal gestures and /s/–stop clusters

In a segmental phonology, the description of initial /s/–stop clusters in English is problematic. There are at least two facts about these clusters that require specific statements within the phonology – statements that apply only to these clusters. First, the phonotactics must state that there is no contrast between voiced and voiceless stops following initial /s/; that is, there is a 'defective' distribution. Second, the realisation of this stop as voiceless unaspirated must be specified by a separate phonetic (or phonological) rule. In current approaches, these facts do not follow from any more general characterisations of English phonology or phonetics.[2] In addition, other problematic aspects of such clusters have led phonologists to argue that they are more 'unitary', i.e. more nearly describable as single segments, than are other clusters. (Ewen 1982 summarises the evidence for a monosegmental analysis, which includes the /s/–stop clusters' violation of the sonority hierarchy and the failure of these clusters to alliterate with /s/ in Germanic verse.) In this section, we will show not only how, in an articulatory phonology, the phonetic and distributional facts about the clusters follow from a more general constraint on the articulatory structure of English words, but also how the gestural structure might account for the clusters' ambiguous status as one or two units.

Crucial to this account is an understanding of the behaviour of the glottis in voiceless stops and clusters. This has recently been investigated in English (Yoshioka *et al.* 1981), Swedish (Löfqvist & Yoshioka 1980a), Icelandic (Löfqvist & Yoshioka 1980b; Pétursson 1977) and Danish (Fukui & Hirose 1983). Like English, these other Germanic languages contrast an initial voiceless aspirated stop with either an unaspirated or voiced initial stop, but neutralise the contrast after /s/. In all cases, a single glottal opening/closing gesture is found for words beginning with /sC/ clusters (where C is a stop). This single gesture is similar to the one that occurs either with /s/ alone initially, or with one of the initial voiceless aspirated stops, although the magnitude of the gesture tends to be a bit smaller when accompanying the stops. As Pétursson (1977) argues, the failure to find two glottal gestures in the initial /sC/ clusters cannot be due to a principle of economy of movement under which the glottis remains open for as many voiceless segments as are required. That is, the glottis is not, apparently, held in an open position during long periods of voicelessness (Yoshioka *et al.* 1981), but rather exhibits a sequence of opening and closing movements. This can be found, for example, in /s#C/ sequences (i.e. sequences containing a word boundary), which can show two glottal opening and closing gestures. Thus, the failure to find two glottal gestures for initial /sC/ clusters points to a generalisation about the linguistic organisation in these languages – words begin with, at most, a single glottal gesture.

These observations form the basis for the gestural analysis of /sC/ clusters. Here, contrasts are described in terms of different characteristic gesture constellations – that is, one or more gestures in a specific spatio-temporal relation. For example, /pa/ and /ba/ differ in the presence *vs.* absence of the glottal opening/closing gesture. /pa/ and /sa/ differ in that one has a bilabial closure gesture, the other an alveolar fricative gesture, both in constellations with a glottal gesture (with characteristic spatio-temporal relations). The initial /sp/ cluster is a constellation of three gestures – an alveolar fricative gesture, a bilabial closure gesture and a single glottal opening/closing gesture. Thus, the glottal gesture can occur in a constellation with a single oral constriction gesture (as in /pa/ or /sa/), with two (as in /spa/), or alone (as in /ha/). (Cf. Hockett's 1955 proposal of an immediate constituent analysis for /sC/ clusters along similar lines.) The phonotactics of English in this approach is a statement of the possible constellations of gestures. The generalisation of interest here is that, in word-initial position, English has at most one glottal gesture, of roughly constant magnitude, regardless of the other gestures with which it co-occurs.

This generalisation accounts for the lack of word-initial contrast in English between /sp/ and /sb/. That is, a word-initial contrast would require either two glottal gestures in the constellation for /sp/, or a different, much smaller, glottal gesture for /sb/. Both of these possibilities are ruled out by the single-glottal-gesture generalisation. Moreover, in conjunction with a fact about intergestural coordination in English, it also accounts for the realisation of the /p/ in initial /sC/ clusters as voiceless unaspirated. This additional fact is that peak glottal opening typically occurs at the midpoint of a fricative gesture, if there is one present in the constellation, or, if not, at the release of a stop closure gesture (Yoshioka *et al.* 1981). The generalisation is presented in its current form in (1); it might ultimately be simplified by referring to the coordination of the glottal gesture with the vowel:

(1) *Glottal gesture coordination in English*
  a. If a fricative gesture is present, coordinate the peak glottal opening with the midpoint of the fricative.
  b. Otherwise, coordinate peak glottal opening with the release of the stop gesture.

Statement (1) holds for both single consonants and consonant clusters. For single consonants, it accounts for the fact that initial voiceless stops are aspirated (since, by (1b), the peak glottal opening occurs at the release of closure), while initial fricatives are not (1a). For clusters like /sp/, it accounts for the lack of aspiration. That is, since the peak glottal opening occurs during the fricative (1a), by the time the following stop closure is released the glottis is already narrowed, producing a voiceless unaspirated stop. (The above analysis of aspiration is similar to a proposal by Catford 1977, who did not, however, explicitly discuss gestural organisation.)

Statement (1) can also account for another aspect of the phonetic structure of English – the devoicing of sonorants following initial voiceless stops but not following initial voiceless fricatives. For initial /pl/, for example, (1b) predicts that the peak of the glottal gesture is timed to occur at the release of the stop gesture, regardless of the presence or absence of other gestures. As Catford (1977) notes, the alveolar lateral (impressionistically, at least) has already been achieved by the time of release of the stop closure, so that the wide-open glottis co-occurs with the lateral, producing a voiceless lateral. Thus, the voicelessness of the lateral follows directly from the nature of the gestures and the independently required generalisation about gestural coordination, and does not have to be stated as a separate allophonic rule. In contrast, (1a) predicts that sonorants will be only slightly devoiced following initial voiceless fricatives, and not devoiced at all in clusters such as /spl/, since the peak glottal opening occurs at the midpoint of a fricative gesture regardless of the number of following consonantal gestures. Thus, the intergestural coordination generalisation captures a number of facts about English phonetics that would otherwise require separate statements.

Returning to the single-glottal-gesture generalisation, we note that it also has implications beyond its explanation of the defective distribution of /sC/ clusters. The ambiguous nature of such clusters is inherent in their proposed gestural constellations, consisting of a single glottal gesture with two overlapping oral gestures. These clusters might act as single units under the influence of the single glottal gesture, or as sequences of two units under the influence of the two oral gestures. A similar analysis has been proposed for ejective clusters such as [t'p'] in Kabardian (Anderson 1978). He argues that Kuipers' (1976) analysis of Kabardian as phonologically vowelless can be maintained if the ejective clusters are treated in a unitary way as complex segments. Their unitary phonological behaviour is related by Anderson to their articulatory nature, consisting of a sequence of two oral articulations associated with a single laryngeal gesture. Thus, he proposes an autosegmental-type analysis, in which a single laryngeal specification is associated with a sequence of specifications for oral articulators. The parallel with the gestural constellations proposed here for /sC/ clusters is obvious. Both cases independently support our contention that gestural structures, derived from observing articulatory movements, provide an appropriate basis for stating phonological generalisations. Moreover, taken together, they suggest a general principle that a particular type of gestural structure (one laryngeal gesture organised with two oral ones) may be associated with ambiguous phonological behaviour.

2.2 Prenasalised stops and nasal–stop clusters

Prenasalised stops constitute another class of complex segments whose analysis has been used to enrich the strictly segmental phonological model (e.g. Anderson 1976; Feinstein 1979; Ewen 1982). In this section, we will show that the difference between prenasalised stops and nasal–stop sequences posited in such analyses cannot predict the kinds of temporal regularities shown by nasal–stop sequences in English, unless certain ad hoc rules are added. These temporal regularities lead us to hypothesise a similar gestural analysis for prenasalised stops and English clusters, and we will present articulatory data to support this analysis. This gestural analysis, we will argue, captures the relevant phonological generalisations while allowing the temporal regularities to be predicted directly from the gestural organisation.

Anderson (1976) has presented arguments for analysing prenasalised stops as single segments, but with a sequence of values for the feature [nasal] (this is consistent with Herbert's 1975 acoustic data showing that prenasalised stops have roughly the same duration as simple stops). Thus, the domain of value-assignment for the nasal feature is not coterminous with the boundaries between segments. In this way, the ambiguous nature (unitary *vs.* sequential) of such stops can be directly captured. His representation for a prenasalised stop is shown in (2a) and for a sequence of homorganic nasal + stop in (2b):

(2)	a. m b	b. m b
cons	+	+ +
nasal	+ −	+ −
ant	+	+ +
cor	−	− −
⋮		

The structures represented in (2a) and (2b) might be expected to lead to different phonetic entities. From the gestural point of view, we would expect to find a difference between the bilabial closure gestures in (2a) and (2b), with the prenasalised stop (2a) having a single bilabial closure gesture, and the nasal–stop cluster (2b) having either two bilabial closure gestures, or, possibly, a single longer bilabial closure gesture. Since in

English, words like *camper* and *canker* are analysed as having nasal–stop sequences, we would expect them to have structures like that in (2b). (The analysis is supported by distributional considerations: /mp/, /mb/, etc., cannot occur in syllable-initial position where no sonorant–stop sequences are allowed, but they can occur post-vocalically, along with other sonorant–stop sequences.) However, certain durational properties of the words containing these clusters, specifically the duration of the preceding vowels, are not correctly predicted by representations such as (2b).

The durational characteristics of the English words *camper* and *camber* were analysed acoustically by Vatikiotis-Bateson (1984) and compared to words containing single segments, *capper*, *cabber* and *cammer*. His results showed, as expected from other studies (e.g. Haggard 1973; Lindblom & Rapp 1973; Walsh & Parker 1982), shortening of the nasal and stop segments when they occurred in a cluster, compared to their durations as single consonants. However, contrary to expectations based on other studies (e.g. Lindblom & Rapp 1973; Fowler 1983), the stressed vowels preceding the nasal–stop clusters did not shorten when compared to the vowels before single consonants (i.e. /mp/ *vs.* /p/ and /mb/ *vs.* /b/). That is, the labial nasal–stop sequences behaved like single consonants in terms of their effects on preceding vowel duration. Moreover, this similarity of behaviour between the clusters and the single consonants extended to the effect on the preceding vowel duration of the consonantal voicing. In the Vatikiotis-Bateson data (as has also been shown by Raphael *et al.* 1975 and Lovins 1978), vowels were shorter before the clusters containing a voiceless stop as well as before the single voiceless consonants, in spite of the fact that the nasal immediately following the vowel was voiced. That is, it was the voicing of the oral stop that determined the vowel length differences.

In the above analyses of acoustic duration, the movement of the velum appeared to be irrelevant, since nasal–stop clusters behaved like single consonants. The similarity between the effects of clusters and singletons could be accounted for if there were a single bilabial gesture in English bilabials and nasal–stop sequences, regardless of the movement of the velum. Such a specification is best captured by representation (2a). This implies in turn, however, that nasal–stop sequences in English should have the same gestural structure as prenasalised stops.

To test this hypothesis about the gestural structure of English nasal–stop sequences, we collected articulatory data for the same set of words (containing bilabial stops and nasal–stop sequences) that Vatikiotis-Bateson analysed, and for similar sequences in a language with prenasalised stops, kiChaka (Chaga), a Bantu language spoken in Tanzania (Nurse 1979). Chaga /mb/ is analysed as a prenasalised stop and can occur word-initially in contrast to /m/ and /p/. In addition, word-initial /mp/ occurs in Chaga, but here the /m/ is analysed as constituting a separate syllable. To investigate the labial sequences, we recorded a female speaker of American English and a male speaker of Chaga. The same overall format was used for both speakers. Each spoke the selected words containing bilabial sequences in a carrier phrase, repeating the words in the phrases five times. The selected words and carrier phrases are listed in (3). Notice that there is no /baka/, since all voiced obstruents in Chaga are prenasalised:

(3)		*English*	*Chaga*
	phrase:	it's a __ tomorrow.	/wia mboka __ kimbuho/ 'say to the starter __ slowly'
	words:	capper	/paka/ 'cat'
		cammer	/maka/ 'year'
		cabber	
		camper	/mpaka/ 'boundary'
		camber	/mbaka/ 'curse'

The articulatory information for both speakers was gathered with a Selspot system at Haskins Laboratories. To track the movement of the lips and jaw, miniature infra-red light-emitting diodes (LEDs) were attached to the midpoint of the upper lip, the lower lip, and just under the chin (or, in the case of the bearded Chaga speaker, slightly above the chin). A modified video camera positioned to capture the profile of the speaker then tracked the movement of the diodes. In addition to the lip and jaw movement data gathered by the Selspot equipment, a gross measure of nasal airflow (during voiced speech) was obtained using an accelerometer attached to the bridge of the nose.

Fig. 2 shows the data for the English word *camper* in a carrier phrase. The acoustic signal is displayed in the bottom panel. The markings in each panel are derived from the acoustic signal – for example, NAS to indicate the onset of nasal murmur associated with the /m/, CLO for the onset of silence during the stop gap of the /p/, RL for the acoustic release of closure, AE and ER for voiced vocalic onsets. The top panel shows the information about nasality gathered by the accelerometer. Notice that, in addition to the nasal murmur (from 390 ms to 440 ms), the entire vowel /æ/ (from 290 to 390 ms) is nasalised. The middle panels display information about the vertical position of the upper and lower lips over time. Here we are most concerned with the labial closure gesture, extending approximately from 350 to 500 ms. The upper lip has very little vertical movement, while the lower lip smoothly raises and then lowers for closure and release. The actual acoustic closure encompasses most of the peak of the labial gesture, from 390 ms (beginning of the nasal murmur) to 470 ms (release of the stop). (Note that closure is achieved before the highest point in the articulatory movement is achieved, and is not released (point labelled RL) until after the lip begins to move down again. The lips are, therefore, continuing to move even during acoustic closure, presumably compressing the tissues involved.) We will be concentrating on this lower lip movement as we investigate the articulation of the various labial sequences.

Our primary interest in the labial gestures is in the similarities and

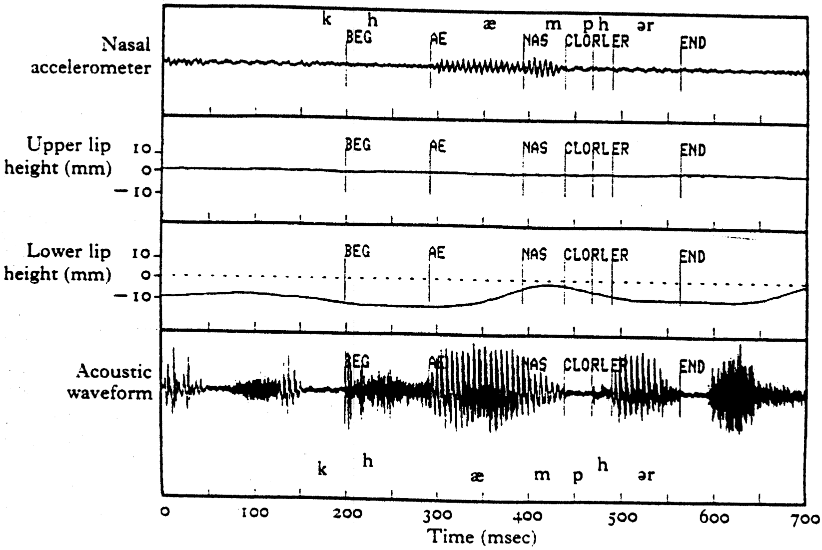

Figure 2

Acoustic waveform and articulatory measurements for single token of *camper*.

differences among gestures in different phonological categories, that is, single consonants, prenasalised consonants, nasal–stop clusters, and syllabic nasal plus stop. Both speakers proved to be quite regular across tokens within the same phonological category. Typical examples are shown in Fig. 3a for the English speaker (the medial /p/ of *capper*), and Fig. 4a for the Chaga speaker (the initial /m/ of /maka/). In both figures, lower lip gestures for two repetitions of the word are superimposed and displayed above the acoustic signal from one of the tokens. The vertical lines indicate the onset and release of closure, as measured in the acoustic waveform. To compare tokens between phonological categories, a single repetition was chosen from each set of five for each of the words to be compared. In each case, this representative item (selected from the second, third or fourth repetition) was identical, as determined by visual inspection, to at least one other repetition, both in terms of the pattern over time of the lower lip gesture, and in terms of the timing of the gesture relative to the surrounding vowels. These representative items are used in the rest of the figures; the conclusions based on these items have been confirmed by comparisons among all the repetitions.

The between-category comparisons indicate that, contrary to expectations based on segmental descriptions such as (2b), all of the phonological categories except for the syllabic nasal + stop are represented by a single labial gesture. That is, there is no systematic difference among the labial gestures associated with a single consonant, a prenasalised consonant,

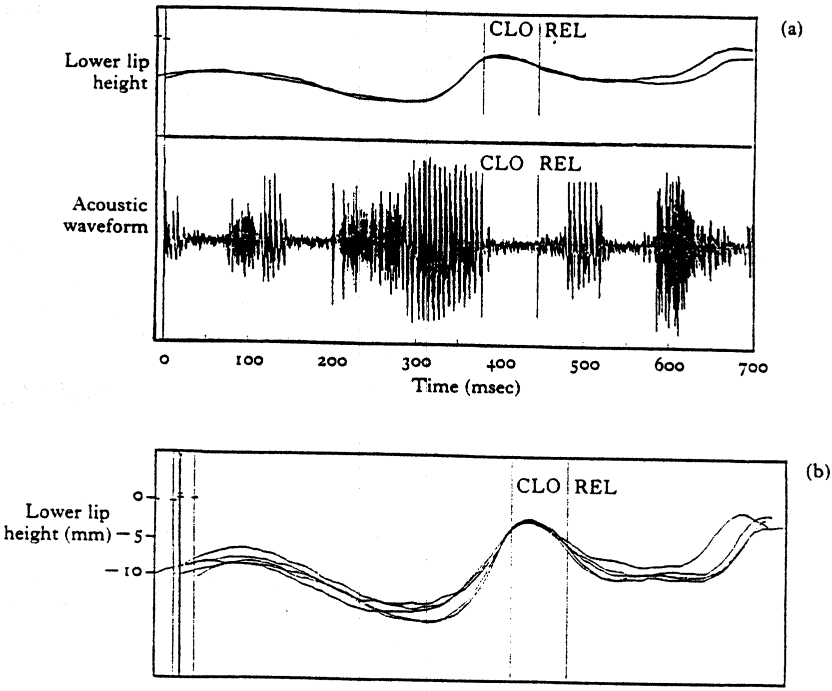

Figure 3

Comparison of lower lip trajectories for English words. (a) Two superimposed tokens of English *capper*; (b) superimposed English *capper*, *cabber*, *cammer*, *camper* and *camber*.

and a consonant cluster. This can be seen in Fig. 3b for English, and Fig. 4b for Chaga.

Fig. 3b shows the lower lip traces for English *cabber*, *cammer*, *camper* and *camber* superimposed on the trace for *capper*. (The gestures have been slightly offset, both horizontally and vertically, to facilitate comparison of their overall forms. The extent of the horizontal offsets can be determined from the lines on the left; the vertical offsets are represented by the tick marks on these lines.) While there are small differences in the amplitude and in the slope of the onset and offset of the gesture, which correspond to similar differences among /p/, /b/ and /m/ reported in the literature (e.g. Kent & Moll 1969; Sussman *et al.* 1973), the overall envelope of the gestures is similar, particularly in the central portion demarcated by the lines on the *capper* trace. That is, regardless of whether the consonantal portion is described as a single consonant (/b/, /p/ or /m/) or as a consonant cluster (/mp/ or /mb/), in English there appears to be a single labial gesture.

Fig. 4b shows the superimposed lower lip gestures for the Chaga words

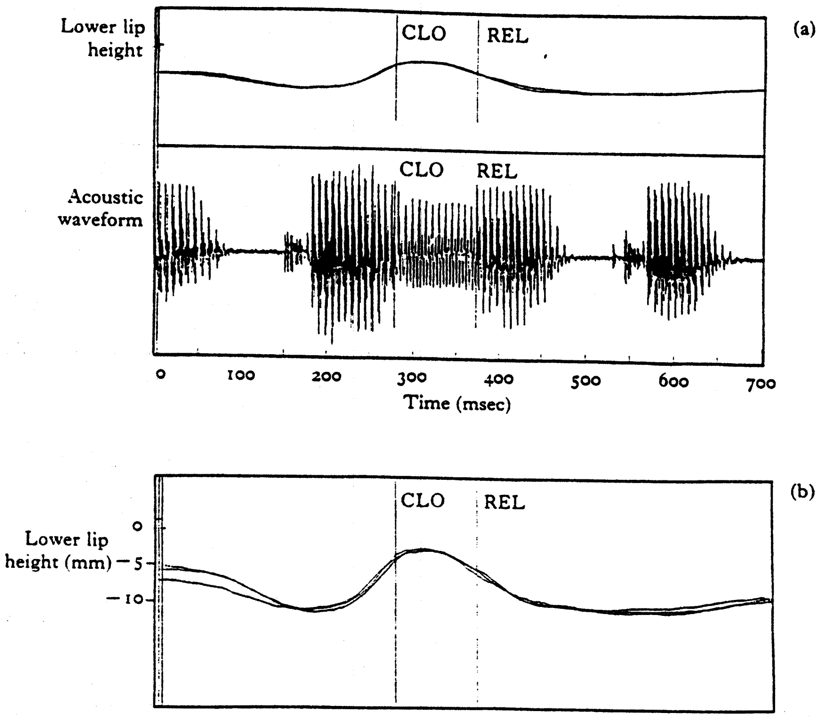

Figure 4

Comparison of lower lip trajectories for Chaga words (except /mpaka/). (a) Two superimposed tokens of Chaga /maka/; (b) superimposed Chaga /maka/, /paka/ and /mbaka/.

/paka/, /maka/ and /mbaka/. Again, as in English, there is a single gesture, quite similar in overall envelope, and particularly in the central portion. That is, in Chaga there is a single labial gesture associated with single and prenasalised consonants.

The syllabic nasal + stop /mp/ in Chaga, however, presents a different picture. Comparing /maka/ and /mpaka/ in Fig. 5a, we see for the first time a clear difference in the overall duration of the envelope of the lower lip gesture. The gesture for /mpaka/ is clearly longer, as can be confirmed by checking the /maka/–/paka/ comparison in Fig. 5b. This difference in duration, we argue, is the result of two overlapping labial gestures. To see that this might be the case, consider Fig. 5c, in which the gestures for /maka/ and /paka/ are superimposed in an overlapping fashion, and Fig. 5d, in which the gesture for /mpaka/ is superimposed on the overlapping gestures for /maka/ and /paka/. The close correspondence observable in the figure holds across all the repetitions for these utterances. That is, the gesture for syllabic /m/ + /p/ corresponds closely to the individual gestures for /m/ and /p/ arranged sequentially with partial overlap. An alternative description could be suggested, namely that the

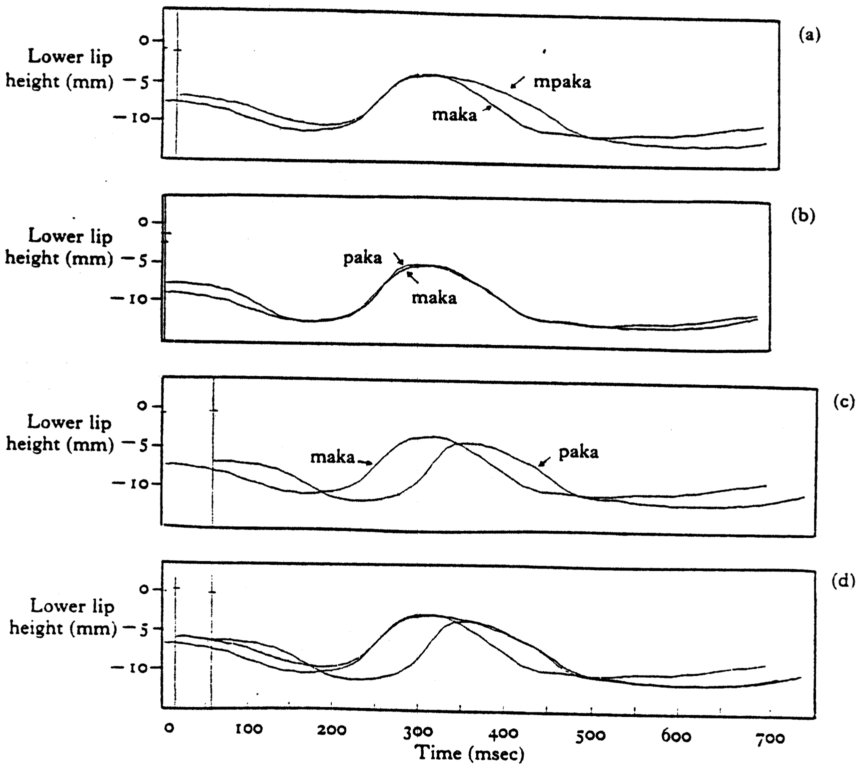

Figure 5

Comparison of lower lip trajectories for Chaga /mp/, /m/ and /p/. (a) /maka/, /mpaka/; (b) /maka/, /paka/; (c) /maka/, /paka/ (shifted); (d) /maka/, /paka/ (shifted), /mpaka/.

bilabial closure gesture in /mp/ was simply 'larger'. Note, however, that both the amplitude and the slopes of the onset and offset are unchanged from the single consonant case. This argues for overlap, or else another mechanism which simply holds for the peak of a gesture, rather than for a larger gesture, since, as reported by Kelso *et al.* (1985), larger gestures (due to changes in stress and rate, at least) typically have both increased amplitude and steeper slopes. We provisionally prefer the analysis of overlap to that of a held peak, since it requires mechanisms that must in any case occur in the phonology, namely overlap among gestures involving different articulators.

Thus the articulatory evidence suggests that syllabic /mp/ is a gestural constellation including two partially overlapping labial gestures. This distinguishes it from Chaga prenasalised stops and English nasal–stop clusters, both of which are constellations involving a single bilabial closure gesture. That is, as we hypothesised, the English nasal–stop clusters and Chaga prenasalised stops would both, using Anderson's (1976) framework, be represented as (2a). Representation (2b), however, would be more appropriate for the Chaga syllabic /mp/.

How, then, given the similarity between their gestural structures, do we capture the distinction between prenasalised stops in Chaga and nasal–stop sequences in English? The simplest statement is as a distributional, or phonotactic, difference. That is, in Chaga such gestural structures can occur in word (and/or syllable) initial position, whereas in English the same gestural structures cannot occur in initial position. Thus, we can account for the difference between prenasalised stops and nasal–stop clusters as different distributional characteristics of the same physical structure. However, such a distributional difference can only serve to distinguish prenasalised stops and nasal–stop clusters in two different languages. We still need to address the issue of how prenasalised stops differ from nasal–stop clusters in a language where they contrast. In such a case, we expect an articulatory difference between the two.

Feinstein (1979) describes one such language, Sinhalese. We know that there is in fact a physical difference here between the nasal–stop clusters and prenasalised stops: Feinstein reports that the nasal is longer in /nd/ clusters than in prenasalised /^{n}d/. This difference might reside either in the relative timing of the oral and velic gestures or in the oral closure gesture itself, which could be longer, or doubled, for the clusters. We would in fact expect the latter to be true, because the nasal–stop clusters are part of a morphological class of inanimate nouns in which gemination is an active process. That is, members of this class containing oral stops alternate between single and geminate stops in the plural and definite singular (e.g. [potu] and [pottə] 'core', Feinstein 1979), while members containing nasal–stop sequences alternate between prenasalised and nasal–stop clusters (e.g. [ka^{n}du] and [kandə] 'hill', Feinstein 1979). Such a classification would be simply explained if the identifying characteristic of the class were the lengthening, or doubling, of the oral gesture. For the oral stops, this would result in a geminate, while for the prenasalised stops, it would result in a lengthened nasal, i.e. a nasal–stop cluster.

Such a characterisation, combined with a gestural reformulation of their syllable template, also directly captures Feinstein's (1979) and Cairns & Feinstein's (1982) analysis of the difference between the prenasalised and nasal–stop sequences in Sinhalese as a difference in syllable structure, with the prenasalised stops being tautosyllabic (syllable-initial) and the nasal–stop sequences heterosyllabic. In the gestural reformulation, the terminal nodes of the syllable template would be oral gestures, rather than segments, and syllable onsets would be restricted to single oral gestures. Since the velic gesture under this formulation is not relevant to the syllable structure, either prenasalised or single oral stops (assuming both have a single oral gesture) could occur in the syllable onset. That is, lengthened/doubled oral gestures would be heterosyllabic, while single oral gestures would be syllable-initial, regardless of their co-occurrence with velic gestures. Such a reformulation correctly captures the syllabification difference in the nasal–stop sequences, and eliminates the need for a separate language-specific statement about the priority of prenasalisation.

As we have shown, then, the gestural structures for prenasalised stops and nasal stop clusters lead to both a simple statement of their physical properties, and a satisfactory description of their phonological properties. Anderson's (1976) analysis of prenasalised stops predicts their temporal properties fairly well, but for nasal–stop sequences in English, some rule would be required to collapse a structure like (2b) into a structure like (2a). Moreover, the structure in (2a) then must be mapped onto an articulatory representation, much like that which we take as our basic phonological representation. In contrast, in an articulatory phonology, the representation in terms of a spatio-temporal organisation of gestures directly captures the articulatory structure as well as providing simple statements of phonological generalisations.

3 Preliminary formalisms

The analyses of §2 suggest that spatio-temporal descriptions of articulatory movements can, in fact, provide the basis for stating phonological regularities. In order to state such generalisations explicitly, the gestural structures must be formalised in some way, and it is to such formalisations that we turn in this section. We should note that these suggestions for formalisation are preliminary and incomplete, and their implications for the description of a wide range of phonological data have not yet been investigated. Our preliminary attempts here are intended to show simply that the kind of structures that we have been arguing for can be rigorously formalised and that their adequacy in accounting for complex phonological data can, therefore, be tested.

We begin by describing one promising approach to a dynamical specification of gestures and their coordination in §3.1. This dynamical specification is meant to be detailed enough to account for phenomena such as coarticulation and language-particular timing patterns. In §3.2, we show that a simplified, more qualitative notation can be used to index these dynamically defined structures. These indices are a simplified way of representing gestural structures, appropriate to such linguistic functions as lexical contrast and description of phonological generalisations.

3.1 Specification of gestures and inter-gestural relations

We have been using the notion of an articulatory gesture as a characteristic pattern of movement of an articulator (or of an articulatory subsystem) through space, over time. How can we precisely define such spatio-temporal patterns? We might attempt to specify the values of articulator position at successive points in time. Such an approach, however, in which time is explicitly one dimension of the description and spatial position another, has trouble with the complex variations in articulatory trajectories introduced by changes in speaking rate and prosodic factors. It would be preferable to view the change in position over time as the output of a more abstract system, such as a dynamical system, capable of generating a variety of related trajectories.

Dynamical systems (e.g. Rosen 1970; Abraham & Shaw 1982) have been applied to problems of motor coordination in biological systems in general (e.g. Fel'dman 1966; Polit & Bizzi 1978; Cooke 1980; Kelso & Holt 1980; Kugler *et al.* 1980; Kelso *et al.* 1981; Kelso & Tuller 1984a), and to the organisation of speech articulators in particular (Lindblom 1967; Fowler 1977; Fowler *et al.* 1980; Kelso *et al.* 1983; Kelso & Tuller 1984b; Ostry & Munhall 1985). Space and time are not specified in a point-by-point fashion in a dynamical system, but the system is capable of specifying characteristic patterns of movement that are organised in space and time. There are two properties of such systems that make them useful to the description of linguistic gestures. First, for a given fixed specification of system parameters, the system can output an infinite number of different (but related) trajectories, as a function of the initial conditions of the articulators, and as a function of other dynamical systems (for other gestures) that might be simultaneously active. At least some trajectory context-dependence (i.e. coarticulation) can be handled in this way, by characterising a gesture in terms of the invariant *input* parameters to such a system. Second, although the articulators are moving throughout such a gesture, the equation itself does not vary over time, but rather characterises the whole pattern of movement. Thus, the traditional notion of a discrete element, imposed on speech from the outside, is replaced by the notion of coherent gestural movements that can be described by a single system of equations (cf. Fowler *et al.* 1980).

To exemplify such equations, consider a physical example of a dynamical system, a mass attached to a spring. If the mass is pulled, stretching the spring beyond its rest length (equilibrium position), and then released, the system will begin to oscillate. Assuming that the system is without friction, the resulting movement trajectory of the mass can be described by the solution to equation (4):

(4) $m\ddot{x} + k(x - x_0) = 0$

where m = mass of the object

k = stiffness of the spring

x_0 = rest length of the spring (equilibrium position)

x = instantaneous position of the object

$\ddot{x}$ = instantaneous acceleration of the object

Thus, a time-varying trajectory is generated by an equation that does not itself change over time. Different trajectories can be obtained from this same system by different choices of values for the dynamical parameters m, k and x_0, and by different initial conditions for x and $\ddot{x}$. Changing the stiffness k of the spring will affect the frequency of oscillation of the mass, while changing the rest length (equilibrium position) of the spring x_0 and the initial position x will affect the amplitude of oscillation.

Recently, it has been shown that such dynamical systems can account for systematic trajectory differences associated with linguistic variations in stress (Browman & Goldstein 1985; Kelso *et al.* 1985; Ostry & Munhall 1985). In these papers, mass–spring models such as (4) provided an abstract description of the articulatory movements associated with lip closure. Thus, the x in equation (4) was taken to represent the vertical position of the lower lip, instead of the length of a spring. The lower lip in the stressed syllables took more time to move between the displacement peaks and valleys, which was modelled by decreasing the stiffness (k) in equation (4). The lower lip in stressed syllables also moved a greater distance, which can be modelled by increasing the difference between the rest length of the lower lip (x_0) and the initial position (although this aspect of the modelling was less thoroughly explored in the above papers).

A gesture such as that for bilabial closure cannot be fully described by the movement of a single articulator; the coordination of a number of articulators is required, i.e. the jaw, the lower lip and the upper lip. The TASK DYNAMICS of Saltzman & Kelso (1983) offers a promising approach to modelling this coordination. Task dynamics provides an organisation of articulatory movement that is defined in terms of a particular task to be performed – in our example, lip closure. It relates this task closure to the movement of the various articulators involved in its performance, in particular the jaw, the lower lip and the upper lip. The positions of these articulators are anatomically linked – as the jaw moves, for example, the lower lip can move along with it. Because of this fact, lip closure can be achieved in a number of different ways, from moving only the jaw, to moving just the lower lip with relatively little jaw movement. It is this flexibility that allows a phonetic task such as bilabial closure to be achieved even when the movement of one of the component articulators is mechanically restrained during speech (Kelso *et al.* 1984; Abbs *et al.* 1984). Such compensatory behaviour has been successfully modelled by the task dynamics approach (Saltzman forthcoming). In addition, this flexibility can account for aspects of coarticulation such as the fact that a bilabial closure gesture is produced with a higher jaw position in [bi] than in [bæ] (Sussman *et al.* 1973). The default contributions of the component articulators to a given task, in the absence of perturbation or coarticulation, can be specified in terms of characteristic weightings for these articulators. These weightings may vary for different gestures: for example, the upper lip is weighted quite differently in bilabial closure and labiodental fricative gestures.

A gesture, then, is defined by specifying (i) a dynamic equation (or a set of them), (ii) a motion variable or variables, i.e. the variable(s) to substitute for x in equation (4) or other dynamic equation, (iii) values for the coefficients of the equation (the dynamic parameters), and (iv) weightings for individual articulators. An initial application of this definition of gestures was presented by Browman *et al.* (1984). They employed a single undamped second-order system (such as (4)) defined for two motion variables, lip aperture (vertical distance between the upper and lower lips) and lip protrusion. Different values for the dynamic parameters (stiffness and equilibrium position) were employed on alternate motion cycles, so as to generate the articulatory trajectories appropriate for an alternating stress sequence /'mama'mama.../. The computed output trajectories of the upper lip, lower lip and jaw were then used to control a vocal-tract simulation (Rubin *et al.* 1981) that synthesised speech. Thus a very simple specification of a bilabial closure gesture in terms of dynamically defined variables for lip aperture and protrusion successfully captured the information necessary to produce convincing speech. At the same time, as we have seen, such gestural descriptions are useful as a basis for phonological description.

In order to generate the complete inventory of speech sounds, gestures must be combined into constellations. Again, as with the gestures themselves, the relations among gestures can be specified abstractly using spatio-temporal phase relations (Kelso & Tuller 1985). A specification in terms of phase is neither solely spatial nor solely temporal, because the exact point in time associated with a particular phase angle will change as the frequency (stiffness) of the gesture changes, and the exact point in space will change as the amplitude (rest length) of the gesture changes. Rather, phasing specifies relations among characteristic spatio-temporal patterns. Empirical evidence in favour of a spatio-temporal approach has been presented by Tuller *et al.* (1982) and Tuller & Kelso (1984). For example, Tuller *et al.* (1982) have shown that in sequences like ['papip] and [pa'pip], the time of onset of lip activity for the intervocalic consonant relative to jaw or tongue activity for the initial vowel is quite variable across the two different stress patterns and across changes in speaking rate. This indicates that the purely temporal approach cannot specify gestural relations in a sufficiently general way. However, Tuller *et al.* go on to show that the onset of lip activity for the intervocalic consonant can be quite precisely (and linearly) related to the period of the vocalic cycle, defined as the time between the onset of activity for the first vowel and the onset of activity for the second. This linear relationship remains invariant across changes in speaking rate and stress. (Similar constancies in relative timing of acoustic events, across changes in rate, have been reported by Weismer & Fennell 1985.) Kelso & Tuller (1985) have further analysed their movement data in terms of phase and have shown that the consonant gesture begins at a fixed phase angle in the vowel cycle. As the vowel period changes with stress and rate, the absolute time corresponding to that phase angle will also change, in a systematic way.

We take, then, as a first hypothesis that gestures can be characterised in terms of a dynamical system and its associated motion variables and parameter values, and that intergestural relations can be specified in terms of their phasing. This framework can accommodate the cross-linguistic timing differences discussed in §1 quite naturally, although the analysis in any particular case (e.g. phase differences *vs.* stiffness differences) remains to be determined by the relevant spatial and temporal data.

3.2 Contrastive articulatory structures

So far, gestural organisations have been described in terms of attributes of the motions of physical articulators, including the more abstract and general physical descriptions provided by dynamics, specifically task dynamics. That is, we have shown in the last section how it is possible to capture spatio-temporal articulatory structure, using a dynamical framework. In this section, we continue to develop a formalism for articulatory phonological representation by laying out some sample lexical descriptions.

One function of a lexical description is to provide information about the physical structure of an item so that linguistically significant similarities and differences among lexical entries can be observed in as simple and direct a way as possible. Since we are dealing solely with articulatory structure, the domain in which linguistic facts such as distinctiveness can be stated consists of the set of articulatory gestures and their relations. The descriptions in this section differ from those in the previous section only in terms of the degree of detail in the description. It is as if, in the present section, we have decreased the resolution on our microscope so that the descriptions are coarser-grained and more qualitative. Thus we are referring to the same set of dynamically specified gestures, but this time using symbols which serve as indices to entire dynamical systems. These symbolic descriptions highlight those aspects of the gestural structures that are relevant for contrast among lexical items. In our discussion of lexical representations, then, there are three important considerations to keep in mind. First, the minimal units in the lexical representation are dynamically defined articulatory gestures. Second, these gestures are spatio-temporal in nature. Third, the gestures are organised asynchronously, with varying degrees of overlap among the gestures.

Symbol	Gesture
β	bilabial closing and opening
γ	glottal opening and closing (returns to voicing position)
$+\mu$	velic opening
$-\mu$	velic closing
V	vowel

[Table I. *Gestural symbols*]

Table I shows the symbolic notation we are suggesting to index the gestures relevant to the English and Chaga words discussed in §2.2. The symbols are shorthand for specific sets of dynamical equations and their associated motion variables and parameter values that can generate the kinds of articulatory trajectories seen in Fig. 6. These gestural trajectories are based on the articulatory data for *camper*, an example of which was shown in Fig. 2. The bottom panel is the acoustic signal. The middle articulatory panel is the actual recorded trace of the vertical movement of the lower lip. This closing and opening gesture of the lips is indexed by

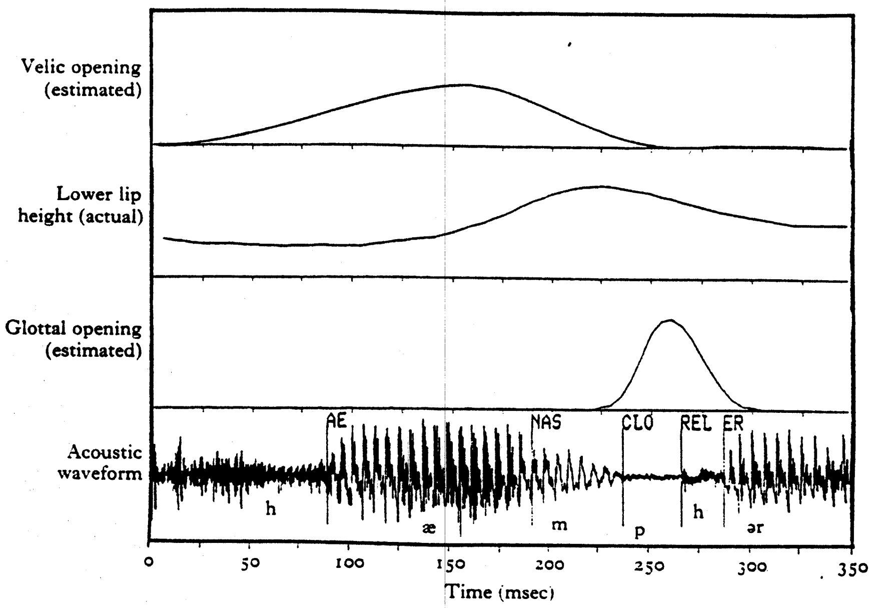

Figure 6
Acoustic waveform and bilabial, velic and glottal gestures for *camper*.

the β in Table I. The other two panels of articulators are estimates only, and show the amount of opening associated with the gesture, rather than actual vertical height. The bottom articulatory panel displays a representative glottal gesture associated with voicelessness: the peak indicates the maximum opening of the vocal folds. (The shape of the glottal gesture was estimated from Sawashima & Hirose 1983, while the timing was based on the acoustic signal, combined with information from various studies on glottal timing (Löfqvist 1980; Löfqvist & Yoshioka 1985).) Here, a γ in Table I represents the glottal opening and closing gesture. Note that the presence of the glottal gesture means an open glottis, i.e. voicelessness.

The top panel estimates the opening of the velo-pharyngeal port, based on the accelerometer record and published data on velum movement (e.g. Kent *et al.* 1974; Vaissiere 1981). These data indicate that in an utterance like *camper*, velum lowering (i.e. velic opening) begins at the release of the initial consonant, and velum raising (i.e. velic closing) begins at some time before the achievement of articulatory closure for the /mp/. The velum movement is represented by two separate gestures in Table I, an opening gesture indicated by a $+\mu$ (nasal), and a closing gesture indicated by a $-\mu$ (non-nasal, i.e. oral). The decision to treat velic opening and closing as two separate gestures, as compared with the glottal and oral gestures that incorporate both opening and closing, is based on the fact that each velic gesture may act as a word-level phenomenon, so that the velum can possibly be held in either a closed or an open position indefinitely. Kent *et al.* (1974) provide an example of the latter phenomenon: a long sentence with many nasal consonants in which the velum remains lowered throughout. This can also be seen in nasal harmony as well as in non-distinctive nasalisation. In addition, the velum opening and closing gestures may require different amounts of time; for example, Benguerel *et al.* (1977) show that for a French talker, opening gestures are slower than closing gestures. It is possible that a comparable internal gestural structure will also be needed for oral and glottal gestures.

Fig. 6 also illustrates the timing relations among gestures. Note that the overlap among the gestural trajectories reflects the inherent spatio-temporal properties of the gestures as well as their asynchronous organisation. That is, the gestures that form the /mp/ constellation all require a certain amount of time to unfold, but they are not necessarily synchronised in the sense of their onsets (or peaks, or offsets) being aligned. The velic opening, for example, begins at least 100 ms before the labial gesture begins, and velic closure begins sometime in the middle of the labial gesture. This asynchrony results in the vowel being nasalised, and the labial gesture being partly nasal and partly oral. The glottal gesture is delayed with respect to the labial gesture, so that the glottis most likely reaches its peak opening after the peak closure of the labial gesture, but before its offset. Thus, a single glottal gesture, requiring a certain amount of time to unfold and asynchronously aligned with the labial closure gesture, generates both voicelessness and aspiration (as originally proposed by Lisker & Abramson 1964). In addition to the particular gestures, then, our symbolic representation must capture aspects of the phase relations among the gestures, because as we shall see, contrasting items may include the same set of gestures, but in different relations.

For the sake of our symbolic representations, we project gestural constellations into a two-dimensional representation which captures qualitative aspects of these relations that are important for contrast (or for certain kinds of phonological generalisations). Examples of such constellation projections for the English words investigated in §2.2 are shown in Fig. 7, with gestures for the initial consonant omitted. Chaga words beginning with /m/, /p/ and /mb/ have representations like those of the comparable English words, except that the initial V is not present for these words. The representation of Chaga /mpaka/ (with syllabic /m/) is also shown in the figure. Note that vocalic gestures are indexed simply as V, because their gestural aspects have not been investigated here.

The vertical dimension of these representations is organised by articulatory subsystem. Gestures of the oral subsystem are found on the top two lines, gestures of the laryngeal subsystem are found on the third line, and velic gestures are at the bottom. The particular ordering (from top to bottom) is meant to relate the gestural structures to the more global (syllable and foot) rhythmic organisation of speech. (This rhythmic organisation, corresponding to, e.g. metrical trees or grids, or CV skeleta, is itself not yet incorporated in these structures.) The closer to the top a gesture is, the more relevant it is presumed to be in carrying the overall

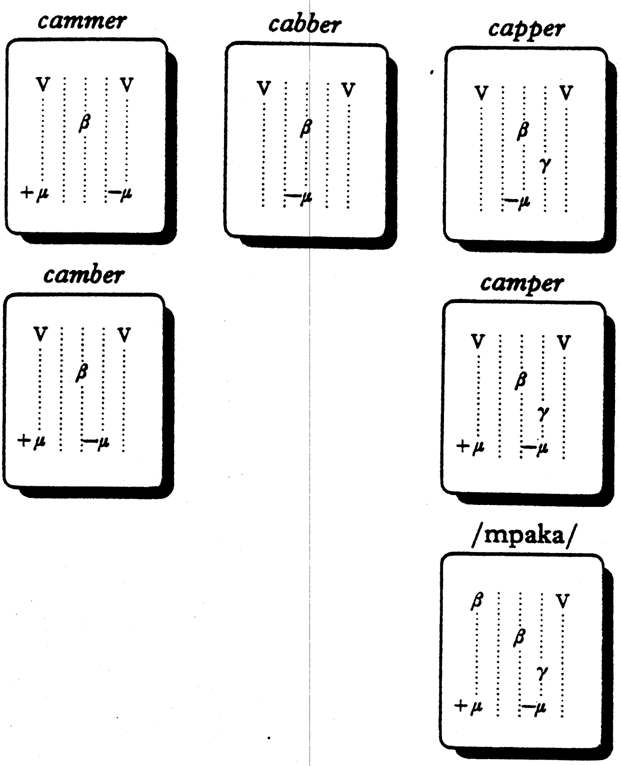

Figure 7
Gestural constellations for English words, and for Chaga /mpaka/ (see text for interpretation).

rhythm. Thus, vowel gestures are found on the top line, as they seem to be most important in carrying the speech rhythm, with other gestures being coproduced with them (cf. Fowler 1983). Velic gestures, by contrast, are placed at the very bottom, because they contribute very little, by themselves, to the rhythmic structure.

The horizontal dimension of the representations in Fig. 7 consists of a grid that can be used to give a qualitative indication of the relative phase relations of the gestures. The lines of the grid represent roughly 90 degree (quarter cycle) phase intervals. Two gestures that are lined up on the same grid line are assumed to be relatively synchronous. For example, their onsets, or their maximum displacements, might coincide in time. The grid lines are not, however, meant to indicate any special structural relation between lined-up gestures. For example, it is not necessarily the case that one of the gestures governs the other, or that there is a particularly cohesive (or tightly invariant) relationship between the gestures. Gestures on successive grid lines are assumed to have approximately a 90 degree phase relation (e.g. the displacement maximum of one gesture synchronised with the velocity maximum of another); those two lines apart are assumed to have a 180 degree relation; etc.

As an example, consider the constellation for English *camper*. The placement of the two V symbols indicates that they are phased 360 degrees apart (they are separated by four grid intervals), and they thus represent one complete vowel cycle. For expository purposes, we can think of this vowel cycle in terms of the action of the jaw – high for the consonant, low for the vowel. The first grid line can be thought of as corresponding to the onset of the first vowel gesture – the beginning of the descent of the jaw from its maximum height towards its low position for the vowel. The third grid line can then be used for gestures that are 180 degrees out of phase with respect to the vowel gesture, i.e. a gesture whose onset occurs 180 degrees into the vowel cycle, when the minimum jaw height is reached for the vowel. This is (roughly) the phase relation between the bilabial closure gesture (on the third grid line) and the vowel gestures reported by Kelso & Tuller (1985). Note that this representation also directly captures the notion (Öhman 1966; Fowler 1983) that consonant gestures overlap a continuous vowel production cycle.

The glottal gesture in *camper* is positioned 90 degrees out of phase with the bilabial gesture. This would be the case, for example, if the peak glottal opening were synchronised with the point of peak velocity during the opening portion of the bilabial gesture, as is consistent with data on the timing of glottal opening for aspirated stops (e.g. Löfqvist 1980), and as can be seen in Fig. 6. An unaspirated stop, as contrasted with an aspirated one, would be represented by synchronising the bilabial and glottal gestures. Turning to the velic gestures, we note that the velic closing gesture ($-\mu$) is lined up with the glottal gesture. This could correspond to the maximum displacement of the two gestures being synchronised – peak glottal opening with the maximum velic closure. Note also that the velic opening gesture ($+\mu$) is positioned on the first grid line, directly capturing the fact that the velum begins to open sometime close to the onset of the vowel (as indicated by the nasal accelerometer, and also as seen in Kent *et al.* 1974).

Such representations not only provide qualitative information about the articulatory structure for individual items, but also serve to differentiate items from each other. Compare, for example, the representations for English *camper* and Chaga /mpaka/. Here, based on our articulatory measurements, the distinction between the English /mp/ and the Chaga syllabic nasal + stop is represented by a second labial gesture for the Chaga. This gesture is positioned on the top line (where normally only vowel gestures occur), in order to capture the fact that this bilabial closure gesture assumes the syllabic function within rhythmic structure that is more usually filled by vowel gestures.

Another pair of lexical items, *cammer* and *camber*, demonstrates the distinctive use of phase structure in the representations. The only difference between the representations for *cammer* and *camber* lies in the phasing of the velic closing gesture. In *camber*, the velum closes during the labial gesture, so that there is a brief period of non-nasalised closure. In *cammer*, on the other hand, the velum closes sometime after the labial gesture is released. This is captured by different grid positions of the velic closure gesture. (The phasing of the gestures is inferred from evidence provided by the nasal accelerometer combined with the acoustic signal.)

It is important to note that the physical descriptions provided by these lexical entries are not complete descriptions of the articulatory actions. Other physical or physiological events regularly occur in these utterances but are not part of these descriptions. For example, larynx lowering and overall expansion of the oral cavity (Westbury 1983) typically accompany the bilabial closure for voiced stops. However, at this early stage of our investigation, we wish to focus on characterising those aspects of the physical structure that are most relevant to capturing linguistic generalisations and specifying contrast among lexical items. In addition, these qualitative representations may omit certain detailed differences that are present in the quantitative specification of the gestures' dynamic parameter values and phasing. For example, the phase angle for /b/ relative to the vowel is somewhat greater than it is for /p/ (approximately 205 degrees *vs.* 180 degrees in Kelso & Tuller 1985). This difference corresponds to the durational differences (discussed in § 1) between vowels before voiced and voiceless stops. Such differences in gestural parameters and phasing are directly represented in the more quantitative description.

The gestural constellations in Fig. 7 represent contrast in a physically realistic way; however, the representations are clearly understructured when compared to recent forms of phonological representation. We expect, however, that additional structuring will emerge as we learn more about patterns of intergestural phasing, e.g. as we discover whether relations are best captured by phasing individual gestures to one another, or whether there are coherent sub-constellations of gestures that should be phased with respect to one another. As such knowledge becomes available, it will be possible to look for convergences between such gestural structures and structures hypothesised on the basis of strictly phonological evidence. For example, different syllable structures may correspond to different characteristic patterns of phasing. Again, we want to account for as much of phonological structure as possible in terms of the organisations required to describe articulatory structure.

The phase relations among gestures are reminiscent of association lines among autosegments on different tiers in autosegmental and CV phonology. Gestural relations and autosegmental associations share the same advantage of allowing gestural overlap (in gestural terms) or multiple associations among autosegments (in autosegmental terms). From this perspective, an articulatory phonology and autosegmental phonology can be seen as converging on the same type of lexical representation. There is nothing in a gestural framework that contradicts autosegmental representations.

Rather, autosegmental phonology and the present framework differ first in their starting points (phonological patterns *vs.* articulatory measurements), and second in the aspects of the representation that are more highly structured. In particular, the gestures have an explicit internal structure: they are dynamical systems that serve to structure the movements of articulatory subsystems.

Thus, the gestural framework can provide a basis for making some principled predictions about the likelihood of a particular phonological feature (or set of features) being split off as a separate tier in autosegmental phonology. We would expect, in general, that the more articulatorily independent of other features a particular feature is, the more likely it is to become a separate autosegmental tier. In particular, we would predict that phonological features associated with gestures of the three articulatory subsystems would be the most likely to be segregated onto separate tiers, a view that is compatible with the prevalence of autosegmental analyses for nasalisation and tone. Within the oral subsystem, the gestures involving the lips are relatively independent of the tongue gestures (although both sets share the jaw as an articulator). Thus features involving lip gestures would be the next most likely to be segregated onto a tier. More generally, we expect that successive gestures specified with common motion variables but with different dynamic parameter values would be more likely to be grouped together on a separate tier than gestures with similar parameter values but specified for different motion variables. Thus we expect that place of articulation features (which correspond to motion variables) should readily split off onto separate tiers but that manner or constriction degree features like [continuant] or [high] (which correspond to particular values of the dynamic parameters) should do so more rarely. Formally speaking, there is no distinction in traditional feature systems between the features for place and manner, whereas in the gestural analysis, they correspond to distinct aspects of the representation. In general then, the explicit model of articulatory organisation provided by the gestural model can lead to specific hypotheses about the hierarchy of expected tier independence.

4 Concluding remarks

We have outlined an approach to phonological representation based on constellations of articulatory gestures, and explored some consequences of this approach for lexical organisation and statements of phonological generalisations. In particular, we showed how working within a gestural framework led to simple generalisations about initial /s/–stop clusters as well as nasal–stop sequences in English and Chaga. We additionally discussed the benefits of formalising these gestures and their relationships in terms of dynamical systems. In general, we suggested that constellations of dynamically defined articulatory gestures can capture articulatory facts in a simple and elegant fashion and show promise of providing more highly constrained and explanatory phonological descriptions. We intend to pursue this line of investigation further – to develop phonological rules that refer to gestural structures, and to discover the range of phonological phenomena that can be accounted for using gestural structures and rules.

Even within other phonological approaches, a gestural description of speech could be used as the basic data for which the phonology attempts to account. That is, gestural structures could replace phonetic transcriptions as the 'output' of the phonology. There are several reasons for doing so. First, it is easier to verify empirically the gestural structure of an utterance: the relation between gestural structures and physical observables is simple and constrained, compared to the relation between a segmental transcription and speech. Second, the gestural structure incorporates temporal information that is usually absent from segmental transcriptions. This is not only important for its own sake (in accounting for cross-language differences, for example), but the increased resolution of the representation may sharpen the process of comparing competing phonological analyses. Finally, as we have argued above, certain aspects of phonological representations, such as tier decomposition, can be rationalised or explained with respect to such gestural structures.

NOTES

* This paper has benefited greatly from the comments and criticisms of several dozen colleagues, primarily from Haskins, Yale and UCLA. We only wish we could thank each of them individually here. Any weaknesses remaining in the paper are due solely to our own intransigence in the face of their patient and generous critiques. (This work was supported in part by NIH grants HD-01994, NS-13870 and NS-13617 to Haskins Laboratories.)

[1] The notion of a gesture has been used in a somewhat similar way as a basic phonological unit in recent versions of dependency phonology (Ewen 1982; Lass 1984). However, our use of the term gesture is restricted to patterns of articulatory movement, while in dependency phonology it can refer to units that cannot, in any obvious sense, be defined in that way, such as the 'categorial' gestures for major classes.

[2] It might be possible to link the two statements by means of markedness conventions. Keating (1984), reanalysing Trubetzkoy's markedness theory, has argued that voiceless unaspirated stops tend to appear in various languages in positions of neutralisation.

REFERENCES

Abbs, J. H., V. L. Gracco & K. J. Cole (1984). Control of multimovement coordination: sensorimotor mechanisms in speech motor programming. *Journal of Motor Behavior* **16**. 195–231.

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Abraham, R. H. & C. D. Shaw (1982). *Dynamics – the geometry of behavior*. Santa Cruz: Aerial Press.

Anderson, J. M. & C. Jones (1974). Three theses concerning phonological representations. *JL* **10**. 1–26.

Anderson, S. R. (1974). *The organization of phonology*. New York: Academic Press.

Anderson, S. R. (1976). Nasal consonants and the internal structure of segments. *Lg* **52**. 326–344.

Anderson, S. R. (1978). Syllables, segments and the Northwest Caucasian languages. In A. Bell & J. B. Hooper (eds.) *Syllables and segments*. Amsterdam: North-Holland. 47–58.

Aronoff, M. & R. T. Oehrle (eds.) (1984). *Language sound structure*. Cambridge, Mass.: MIT Press.

Bell-Berti, F. & K. S. Harris (1981). A temporal model of speech production. *Phonetica* **38**. 9–20.

Benguerel, A.-P., H. Hirose, M. Sawashima & T. Ushijima (1977). Velar coarticulation in French: a fiberscopic study. *JPh* **5**. 149–158.

Bernstein, N. (1967). *The coordination and regulation of movements*. London: Pergamon Press.

Browman, C. P. & L. M. Goldstein (1985). Dynamic modeling of phonetic structure. In Fromkin (1985). 35–53.

Browman, C. P., L. M. Goldstein, J. A. S. Kelso, P. E. Rubin & E. L. Saltzman (1984). Articulatory synthesis from underlying dynamics. *JASA* **75**. S22–S23.

Cairns, C. E. & M. H. Feinstein (1982). Markedness and the theory of syllable structure. *LI* **13**. 193–225.

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

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

Clements, G. N. (1980). Vowel harmony in nonlinear generative phonology: an autosegmental model (1976 version). Indiana University Linguistics Club.

Clements, G. N. & S. J. Keyser (1983). *CV phonology: a generative theory of the syllable*. Cambridge, Mass.: MIT Press.

Cooke, J. D. (1980). The organization of simple, skilled movements. In Stelmach & Requin (1980). 199–212.

Ewen, C. J. (1982). The internal structure of complex segments. In H. van der Hulst & N. Smith (eds.) *The structure of phonological representations*. Vol. 2. Dordrecht: Foris. 27–67.

Feinstein, M. H. (1979). Prenasalization and syllable structure. *LI* **10**. 245–278.

Fel'dman, A. G. (1966). Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III: Mechanographic analysis of execution by man of the simplest motor tasks. *Biophysics* **11**. 766–775.

Flege, J. E. & R. Port (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech* **24**. 125–146.

Fourakis, M. S. (1980). A phonetic study of sonorant–fricative clusters in two dialects of English. *Research in Phonetics, Department of Linguistics, Indiana University* **1**. 167–200.

Fowler, C. A. (1977). *Timing control in speech production*. Indiana University Linguistics Club.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *JPh* **8**. 113–133.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General* **112**. 386–412.

Fowler, C. A., P. E. Rubin, R. E. Remez & M. T. Turvey (1980). Implications for speech production of a general theory of action. In B. Butterworth (ed.) *Language production*. New York: Academic Press. 373–420.

Fromkin, V. A. (ed.) (1985). *Phonetic linguistics*. New York: Academic Press.

Fujimura, O. (1981). Temporal organization of articulatory movements as a multidimensional phrasal structure. *Phonetica* **38**. 66–83.

Fujimura, O., S. Kiritani & H. Ishida (1973). Computer controlled radiography for observation of movements of articulatory and other human organs. *Computers in Biology and Medicine* **3**. 371–384.

Fukui, N. & H. Hirose (1983). Laryngeal adjustments in Danish voiceless obstruent production. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo* **17**. 61–71.

Goldsmith, J. A. (1976). *Autosegmental phonology*. Indiana University Linguistics Club.

Haggard, M. (1973). Abbreviation of consonants in English pre- and post-vocalic clusters. *JPh* **1**. 9–24.

Halle, M. & J.-R. Vergnaud (1980). Three dimensional phonology. *Journal of Linguistic Research* **1**. 83–105.

Hayes, B. (1981). *A metrical theory of stress rules*. Indiana University Linguistics Club.

Herbert, R. K. (1975). Reanalyzing prenasalized consonants. *Studies in African Linguistics* **6**. 105–123.

Hockett, C. F. (1955). *A manual of phonology*. Baltimore: Waverly Press.

Hooper, J. B. (1972). The syllable in phonological theory. *Lg* **48**. 525–540.

Hooper, J. B. (1976). *An introduction to natural generative phonology*. New York: Academic Press.

Jespersen, O. (1914). *Lehrbuch der Phonetik*. Leipzig: Teubner.

Kahn, D. (1976). *Syllable-based generalizations in English phonology*. Indiana University Linguistics Club.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Lg* **60**. 286–319.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In Fromkin (1985). 115–132.

Kelso, J. A. S. & K. G. Holt (1980). Exploring a vibratory systems analysis of human movement production. *Journal of Neurophysiology* **43**. 1183–1196.

Kelso, J. A. S., K. G. Holt, P. E. Rubin & P. N. Kugler (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: theory and data. *Journal of Motor Behavior* **13**. 226–261.

Kelso, J. A. S. & B. Tuller (1984a). A dynamical basis for action systems. In M. Gazzaniga (ed.) *Handbook of cognitive neuroscience*. New York: Plenum Press. 321–356.

Kelso, J. A. S. & B. Tuller (1984b). Converging evidence in support of common dynamical principles for speech and movement coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology* **246**. R928–R935.

Kelso, J. A. S. & B. Tuller (1985). Intrinsic time in speech production: theory, methodology, and preliminary observations. *Haskins Laboratories Status Report on Speech Research* **SR-81**. 23–39.

Kelso, J. A. S., B. Tuller & K. S. Harris (1983). A 'dynamic pattern' perspective on the control and coordination of movement. In MacNeilage (1983). 137–173.

Kelso, J. A. S., B. Tuller, E. Vatikiotis-Bateson & C. A. Fowler (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance* **10**. 812–832.

Kelso, J. A. S., E. Vatikiotis-Bateson, E. L. 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., P. J. Carney & L. R. Severeid (1974). Velar movement and timing: evaluation of a model for binary control. *Journal of Speech and Hearing Research* **17**. 470–488.

Kent, R. D. & K. L. Moll (1969). Vocal-tract characteristics of the stop cognates. *JASA* **46**. 1549–1555.

Kugler, P. N., J. A. S. Kelso & M. T. Turvey (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In Stelmach & Requin (1980). 3–47.

Kuipers, A. H. (1976). Typologically salient features of some Northwest Caucasian languages. *Studia Caucasia* **3**. 101–127.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago: Chicago University Press.

Ladefoged, P. (1980). What are linguistic sounds made of? *Lg* **56**. 485–502.

Lass, R. (1984). *Phonology: an introduction to basic concepts*. Cambridge: Cambridge University Press.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass.: MIT Press.

Liberman, A., F. Cooper, D. Shankweiler & M. Studdert-Kennedy (1967). Perception of the speech code. *Psychological Review* **74**. 431–436.

Liberman, A. M. & I. G. Mattingly (1985). The motor theory of speech perception revised. *Cognition* **21**. 1–36.

Liberman, M. & J. Pierrehumbert (1984). Intonational invariance under changes in pitch range and length. In Aronoff & Oehrle (1984). 157–233.

Liberman, M. & A. Prince (1977). On stress and linguistic rhythm. *LI* **8**. 249–336.

Lindau, M. (1984). Phonetic differences in glottalic consonants. *JPh* **12**. 147–155.

Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. *Quarterly Progress and Status Report, Speech Transmission Laboratory, University of Stockholm* **4**. 1–29.

Lindblom, B. & K. Rapp (1973). Some temporal regularities of spoken Swedish. *Papers from the Institute of Linguistics, University of Stockholm* **21**. 1–58.

Lisker, L. (1974). On time and timing in speech. In T. A. Sebeok (ed.) *Current trends in linguistics* **12**. The Hague: Mouton. 2387–2418.

Lisker, L. & A. S. Abramson (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* **20**. 384–422.

Löfqvist, A. (1980). Interarticulator programming in stop production. *JPh* **8**. 475–490.

Löfqvist, A. & H. Yoshioka (1980a). Laryngeal activity in Swedish obstruent clusters. *JASA* **68**. 792–801.

Löfqvist, A. & H. Yoshioka (1980b). Laryngeal activity in Icelandic obstruent production. *Haskins Laboratories Status Report on Speech Research* **SR-63/64**. 272–292.

Löfqvist, A. & H. Yoshioka (1985). Intrasegmental timing: laryngeal–oral coordination in voiceless consonant production. *Speech Communication* **3**. 279–289.

Lovins, J. B. (1978). 'Nasal reduction' in English syllable codas. *CLS* **14**. 241–253.

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *LI* **12**. 373–418.

McCarthy, J. J. (1984). Prosodic organization in morphology. In Aronoff & Oehrle (1984). 299–317.

MacNeilage, P. F. (ed.) (1983). *The production of speech*. New York: Springer-Verlag.

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.

Mitleb, F. M. (1984). Voicing effect on vowel duration is not an absolute universal. *JPh* **12**. 23–27.

Nurse, D. (1979). *Classification of the Chaga dialects*. Hamburg: Helmut Buske Verlag.

Ohala, J. J. (1974). Experimental historical phonology. In J. M. Anderson & C. Jones (eds.) *Historical linguistics*. Vol. 2. Amsterdam: North-Holland. 353–389.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *JASA* **39**. 151–168.

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

Pétursson, M. (1977). Timing of glottal events in the production of aspiration after [s]. *JPh* 5. 205–212.

Pike, K. L. (1943). *Phonetics.* Ann Arbor: University of Michigan Press.

Polit, A. & E. Bizzi (1978). Processes controlling arm movements in monkeys. *Science* **201.** 1235–1237.

Port, R. F. (1981). Linguistic timing factors in combination. *JASA* **69.** 262–274.

Port, R. F. & M. L. O'Dell (1984). Neutralization of syllable final voicing in German. *Research in Phonetics, Department of Linguistics, Indiana University* 4. 93–133.

Prince, A. S. (1984). Phonology with tiers. In Aronoff & Oehrle (1984). 234–244.

Raphael, L. J., M. F. Dorman, F. Freeman & C. Tobin (1975). Vowel and nasal duration as cues to voicing in word-final stop consonants: spectrographic and perceptual studies. *Journal of Speech and Hearing Research* **18.** 389–400.

Rosen, R. (1970). *Dynamical system theory in biology.* Vol. 1: *Stability theory and its applications.* New York: Wiley-Interscience.

Rubin, P. E., T. Baer & P. Mermelstein (1981). An articulatory synthesizer for perceptual research. *JASA* **70.** 321–328.

Saltzman, E. L. (forthcoming). Task dynamic coordination of the speech articulators: a preliminary model. *Experimental Brain Research Supplementum.*

Saltzman, E. L. & J. A. S. Kelso (1983). Skilled actions: a task dynamic approach. *Haskins Laboratories Status Report on Speech Research* **SR-76.** 3–50.

Sawashima, M. & H. Hirose (1983). Laryngeal gestures in speech production. In MacNeilage (1983). 11–38.

Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *LI* **11.** 563–605.

Stelmach, G. E. & J. Requin (eds.) (1980). *Tutorials in motor behavior.* Amsterdam: North-Holland.

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.

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 regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance* **8.** 460–472.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (eds.) *Perceiving, acting, and knowing.* Hillsdale, New Jersey: Lawrence Erlbaum Associates. 211–265.

Vaissiere, J. (1981). Prediction of articulatory movement from phonetic input. *JASA* **70.** S14.

Vatikiotis-Bateson, E. (1984). The temporal effects of homorganic medial nasal clusters. *Research in Phonetics, Department of Linguistics, Indiana University* 4. 197–233.

Walsh, T. & F. Parker (1982). Consonant cluster abbreviation: an abstract analysis. *JPh* **10.** 423–437.

Weismer, G. & A. M. Fennell (1985). Constancy of (acoustic) relative timing measures in phrase-level utterances. *JASA* **78.** 49–57.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *JASA* **73.** 1322–1326.

Yoshioka, H., A. Löfqvist & H. Hirose (1981). Laryngeal adjustments in the production of consonant clusters and geminates in American English. *JASA* **70.** 1615–1623.