1000.

Chapter 9

Speaking

Carol A. Fowler
Dartmouth College and Haskins Laboratories, New Haven, USA

1 INTRODUCTION

Language users can speak understandably about almost anything, and they can do so almost anywhere. Moreover, the sequences of words composing their utterances can be novel in the experience both of the speaker and of the hearer. All that is required for novel utterances to be understood, roughly, is a competent speaker and a competent listener. A theory of the speaker's competence helps to define the problem of speech motor control. Consider, for example, the problem of sequencing that arises from the linguistic requirement, if understandable utterances are allowed sometimes to be novel, that sentences and words have an internal structure. Comprehension of novel utterances by hearers is possible because utterances universally are composed of familiar parts; that is, all utterances are composed of words that, ideally, are in the lexicon of the hearer as well as of the speaker. More than that, the familiar words of a novel utterance are ordered or otherwise marked according to syntactic conventions of the language, and the syntactic conventions allow the hearer to identify the roles of words in the utterance. Thereby they allow the listener to know the roles of the words' referents in the event being talked about – even if the listener has not witnessed the event or anything much like it.

To enable understandable communication about almost anything, languages need to have large, expandable lexicons of familiar words. And for lexicons to be large and expandable, with all of their component words pronounceable, words of the lexicon must have an internal structure. Words of spoken languages universally are composed of a relatively small number (11 to 141 across the languages in a recent survey; Maddieson, 1984) of meaningless parts that I will call variously phonemes, phonological or phonetic segments, or consonants and vowels. These meaningless components of words have attributes called 'features' that relate in some way to vocal tract behaviors or postures of the articulators of the vocal tract.

Hockett (1960) referred to this dual, ruleful layering of language units – that is, the ruleful ordering of meaningful words in sentences, and of meaningless consonants and vowels in words – as 'duality of patterning'. It is universal to human languages and, Hockett speculated, unique to them. Unique or not, it establishes as a major problem for talkers one of realizing units of the language in their proper sequence.

Handbook of Perception and Action, Volume 2
ISBN 0-12-516162-X

Notice that this is not a problem characteristic of every activity we perform. In basic activities, such as walking, breathing and chewing, there is sequencing but the sequencing is not arbitrary with respect to the physical implementing system. Accordingly, there are no misordering errors in these activities (that is, for example, we never inadvertently inhale twice in succession without exhaling in between). We do make action errors other than spoken ones (Norman, 1981; Reason, 1979), but error-prone actions like speech, and unlike breathing, walking and chewing, are those whose sequencing is arbitrary from the perspective of the implementing physical system.

Not only is there an ordering problem in speech but, in addition, the problem is not alleviated very much for the talker by environmental constraints on sequencing. To enable discussion of almost anything almost anywhere, the behaviors of the talker that implement a spoken communication must, to a considerable degree, be unconstrained by the environmental setting. Or, to put it more positively, to a very great degree, constraints on the serial vocal tract activities that realize a syntactically coherent linguistic message must come from a continuously changing speech plan.

The foregoing discussion of the structure of language is quite familiar perhaps, but in this instance familiarity does not imply any considerable depth of understanding. Despite a near consensus that words and phonemes are real constituents of the language, there is no consensus yet as to their essential properties or as to the media (neural, articulatory, acoustic) in which they can be manifest. Likewise, although theorists are well aware of the serial ordering problem for speakers, they have not reached consensus on how the problem is solved by talkers.

In the review that follows, I will focus on both general issues: the nature of units of spoken language and the means by which they are ordered in speech production. I will restrict the review largely to Hockett's lower of the dual tiers of language. That is, I will not consider how talkers choose what to say, or to any great degree how they select words to say or achieve syntactically acceptable utterances. Rather, given an intent by a talker to produce a particular sequence of words, I will consider what the essential properties of a planned string of words may be and how the plan is implemented as a sequentially ordered activity of the vocal tract.

In the next two major sections of the chapter, I will contrast two theoretical proposals concerning the nature of consonants and vowels. The proposals lead to quite different perspectives on speech activity and on the relationship it bears to those units of the language. A fourth section of the review will examine certain prosodic properties of utterances that emerge when sequences of phonological segments are realized as sequentially ordered, dynamical activities of the vocal tract.

I will introduce each section by describing relevant theories of language structure proposed in the field of linguistics. While foci of linguistic investigation are in some ways far removed from those of investigations of speech-motor control, there are important intersections. One occurs because speech-motor theorists have largely guessed that the units of the language identified in linguistic theory – with the essential attributes that linguists have ascribed to them – will turn out to be the units of the language that talkers aspire to utter. (However, see Kelso, Tuller and Saltzman, 1986, and also Moll, Zimmermann and Smith, 1977, who criticize this approach.) Accordingly, linguistic theories – particularly of phonology – are influential in shaping the form that theories of language and speech production have

taken and in shaping the relationship that theorists see between planned units of the language and vocal tract activity. By introducing each section on the nature of linguistic units in this way, I am attempting to place two contrasting approaches to a theory of speech production in the context of theories of phonology that have informed them.

More than that, while Kelso *et al.* and Moll *et al.* may be correct that the units of the language as described by some *particular* linguistic theory do not aptly characterize the structure of activity in the vocal tract, in my view it is highly implausible that the elements of the language that language users know – that is, the ones that linguistic analyses attempt to uncover – are independent of the activities that implement them. For that reason, I consider it a mistake to study speech production and to develop hypotheses as to the larger organizations characteristic of vocal tract activity for speech in ignorance of the findings of linguistic phonological theories.

Linguistic investigations of systematic properties of spoken utterances uncover structure at a scale that direct observation of the detail of vocal tract activity may mask. That is, the systematicities may be characteristic of vocal tract activity, but if investigators do not know what to look for they may be difficult to see in the wealth of detail that direct observation of the vocal tract uncovers. In short, I include a description of linguistic phonological theories both because these perspectives have been influential in shaping speech production theories and because they should be influential.

2 'LINEAR' PHONOLOGIES AND RELATED APPROACHES TO THEORIES OF LANGUAGE AND SPEECH PRODUCTION

2.1 Linear Phonologies

Linguistic analysis reveals that words have an internal structure. They are composed at least of consonants and vowels. In different amounts and arrangements, consonants and vowels can be shown exhaustively to compose the tens of thousands of words of a language. In addition, they participate as individuals in phonological rules of the language, including rules of insertion, deletion and reordering. None of this evidence is obtained by studying words realized by activities of the vocal tract, and so none of it implies necessarily that these units of the language are units of speech activity or even of speech planning by the talker. If they are nonetheless units of the language, what is their status *vis à vis* speakers and hearers? Since the language itself exists only in so far as speakers and hearers know and use it, if consonants and vowels are units of the language, they must at least be units in the language user's *knowledge* of the language. This generally is the status accorded them in virtue of their roles in the lexicons of languages. They are said to be components of a language user's 'competence' – an idealized speaker–hearer's knowledge of the language that 'provides the basis for actual use of the language by a speaker–hearer' (Chomsky, 1965, p. 9).

To the extent that behavioral evidence converges with linguistic evidence to suggest that words composed of consonants and vowels are components of real language users' plans for an utterance, then there is evidence that components of

competence do provide the 'basis for actual use of the language'. However, from this perspective, separate questions concern the units of vocal tract activity – if any – their nature, and their relationship to units of planning and competence. Units of the language users' competence may or may not be analogous in form to units of vocal tract activity; they may or may not stand in 1:1 correspondence to them.

Linear phonologies have been labeled 'linear', after the fact, to reflect the perspective they take on consonants and vowels as elements of language users' competence. In general, until the middle 1970s different theories of phonology that influenced theories of speaking agreed that consonants and vowels are constellations of simultaneous attributes called 'features'. Because the attributes are simultaneous, vertical slices drawn metaphorically through a word serve to isolate the individual feature constellations of individual consonants and vowels of the word. Theories that ascribe to this 'absolute slicing hypothesis' (Goldsmith, 1976) are 'linear' phonologies.

In *The Sound Pattern of English* (Chomsky and Halle, 1968), the consonants and vowels of a word are represented as columns of features. For Chomsky and Halle, features refer largely to articulatory correlates of the consonants and vowels. A partial representation of the features in 'but' follows:

[b]	[ʌ]	[t]
−vocalic	+vocalic	−vocalic
+consonantal	−consonantal	+consonantal
−high	−high	−high
−back	+back	−back
−low	−low	−low
+anterior	−anterior	+anterior
−coronal	−coronal	+coronal
+voice	+voice	−voice
−continuant	+continuant	−continuant

Chomsky and Halle (1968) identify two functions that features serve in the language. One is 'classificatory': features give distinct lexical entries distinct representations, and they do so in a way that reflects natural groupings of phonological segments as determined by the segments' participation in phonological processes of the language. To serve this function, consonants and vowels of words in the lexicon are represented, as above, by a list of binary-valued features. A second function is to reflect the phonetic capabilities of language users – that is, to reflect any distinction between pairs of speech sounds that a careful listener can hear and transcribe in a speech utterance. To serve that function, features are represented as continuous scales rather than as binary-valued.

In linear phonologies, words of the lexicon may have an internal structure superordinate to the phonological segment that will receive occasional mention below. Some words are composed of a stem and one or more affixes (in English, either a prefix or a suffix); other words, called 'compounds', may be composed of a pair of stems (e.g. 'greenhouse'); still others may be composed of just one stem. Another sublexical unit, the syllable, is not identified as a significant unit in the phonology of Chomsky and Halle (1968).

2.2 Approaches to a Theory of Speech Planning and Production from this Perspective

2.2.1 Introduction

Syntactic constructions span several words. Accordingly, building one requires planning, and one major set of questions that needs to be addressed in a theory of speaking concerns the nature of planning processes and the nature of the planned units. *A priori*, one might guess that finding answers to this set of questions will be more difficult than finding answers to questions concerning the units of vocal tract activity itself, because the latter are public while the former are largely covert. However, considerably more progress has been made in sorting out *planning* processes and units than in sorting out *executed* units of speech.

There are several reasons for this unexpected reversal. First, linguistic theory as just outlined ascribes properties to units of speech, phonological segments in particular, that preclude segments' realization as such in vocal tract activity. Also, speech activity itself does not invite segmentation into any obvious units other than, perhaps, the individual motions of individual articulators. When units of the linguistic message failed to show up in vocal tract activity, researchers were at a loss as to what *else* they ought to look for. It took changes in linguistic theory on the one hand, and, on the other hand, new perspectives on vocal tract actions before progress could be made. But both developments are new, and progress is limited. These developments are described in Section 3.

On the other side, significant advances have been made in understanding planning for speech. It turns out that planning is not wholly covert and that the units of planning that reveal themselves in behavior do, in large part, conform to expectations derived from linear phonological theories.

The most important single source of evidence concerning planning is provided by spontaneous errors that talkers make when they speak. This evidence, augmented recently by experimental elicitation of errors (Baars, Motley and MacKay, 1975; Dell, 1980; Shattuck-Hufnagel, 1986, 1987) or by word games (Treiman, 1983) that require manipulation of parts of words, allows a fairly coherent story to be written about units of planning and about planning operations themselves. In the following sections, I will review the recent literature on speech errors and summarize the conclusions it permits concerning processes of planning.

Although the focus of this chapter is on speaking considered as production of words, themselves composed of phonological segments, the following review includes consideration of planning events upstream of that. Full appreciation of the order that phonological errors exhibit requires putting them in the context of other errors produced, apparently, at other phases of planning that are equally ordered, but respect different ordering constraints.

2.2.2 Definition of Errors

Talkers make a variety of mistakes when they speak. They may lose their train of thought, revise a sentence in midstream, or produce a disfluency that a phonetician would have difficulty transcribing. Alternatively, they may chronically pronounce a word in a way that listeners identify as erroneous, or produce a sentence that

listeners consider ungrammatical but that the talker does not. These are not the kinds of error that researchers have used to investigate planning in speaking. Critical errors for that purpose are fluent departures from an utterance as the talker intended to produce it. Sometimes the talker's intentions are obvious; sometimes talkers correct themselves.

The following general kinds of error are most common (for more complete taxonomies, see Dell, 1986, and Shattuck-Hufnagel, 1979; the following examples of errors are from Crompton, 1982; Dell, 1986; Fromkin, 1971; Garrett, 1980a, b; and Shattuck-Hufnagel, 1979, 1983). One segment of an utterance may be replaced by another. Sometimes the replacing elements may come from the intended string (a *contextual substitution*) or else they may come from elsewhere (a *noncontextual substitution*). Three kinds of contextual substitution are common. In anticipations, an element later in the intended string substitutes for an earlier element ('sky in the sky' for 'sun in the sky'; 'leading list' for 'reading list'). Perseverations are complementary. A segment is substituted for a later one ('class will be about discussing the class' for 'class will be about discussing the test'; 'beef needle soup' for 'beef noodle soup'). In some quite remarkable errors, called *exchanges*, segments swap places ('I left the briefcase in my cigar', 'emeny'). Examples of noncontextual substitutions, in which the substituting segment is not the intended string, are 'Pass the salt' for 'Pass the pepper' and 'Bob McGord' for 'Bob McCord'.

In addition to substitution errors, segments can be added to a string, deleted from it or shifted. In a *shift* error, a segment moves from one location in the string to another ('something all to tell you' for 'something to tell you all'; 'point outed' for 'pointed out'). Finally, two words apparently competing for the same slot in a sentence can blend ('clarinola' for 'clarinet' and 'viola').

Although there are complications, generally it appears that the units that participate in speech errors are discrete units of linguistic theory. Most commonly, units are phonemes, words and morphemes. According to Shattuck-Hufnagel (1987), 40% of all errors in the corpus collected by her and by Garrett (1980a, b) are errors involving single whole consonants and vowels. Very rarely among the error types outlined above are there clear instances of feature errors ('glear plue sky') or errors involving linguistically incoherent sequences. Syllables, which are not assigned a role in the phonological theory of Chomsky and Halle, appear as units in speech errors no more commonly in Shattuck-Hufnagel's (1983) corpus than do CV or VC sequences (where C is an oral consonant and V is a vowel) that do not constitute whole syllables. Even so, syllables play a role in speech errors as outlined below.

2.2.3 What can be Learned from the Errors?

The units of language performance that errors reveal are not, necessarily, units of vocal tract activity. Although errors may be called 'slips of the tongue', they cannot arise in vocal tract activity. When a talker says: 'I left the briefcase in my cigar', he or she has anticipated a whole word three slots earlier in the sentence than its intended location. Moreover, the word – a noun – has been moved into the next closest slot in which a noun belongs. Such an error cannot be an error at the level of motor activity. Presumably, it arises in processes that plan – that is, construct – the intended communication prior to utterance.

Some conclusions that errors do warrant are that utterances are planned, there are planned units, and planned units correspond quite closely to units of linguistic competence as outlined by linear phonologies. Components of a speech utterance that can be arranged differently in different sentences, or words that can substitute one for the other in the language, are just the ones that emerge as discrete and autonomous in spontaneous errors of speech production.

Speech errors are more informative still, because they appear to provide information about the nature and ordering of planning processes in speaking. In particular, evidence suggests both a progressive narrowing of the planning domain and a shift from a focus on content and grammatical roles of words to a focus on surface grammatical form and then to a focus on phonological word shape.

Exchanges of whole words and of phonemes are informative in this regard. Whole-word exchanges can occur over fairly long surface distances in a sentence. In particular, they are very likely to cross syntactic phrase boundaries (81% of the time according to Garrett, 1980a). In contrast, words involved in phoneme exchanges span a syntactic phrase boundary just 13% of the time in Garrett's count.

An interpretation of these differences is that planning in which words are selected to serve some grammatical or propositional role spans several syntactic phrases while planning that focuses on words' labels has a considerably narrower domain. Further, word exchanges tend to be between words of the same syntactic class (85% of the time according to Garrett's count, 1980a), while phoneme exchanges occur between words of the same (39%) or different (61%) syntactic classes apparently indifferently. As described more fully below, constraints on phoneme exchanges are largely phonological. This suggests another shift that accompanies the narrowing of the planning domain. The focus of planning shifts from one on words as conveyers of content to one on words as phonological labels.

Evidence that there may be an in-between planning phase is provided by some 'stranding errors' in which two stems exchange, stranding their suffixes (e.g. 'I thought the park was trucked' from Garrett, 1980b). These look like whole-word exchanges in the sense that sequences of consonants and vowels that can stand alone exchange. However, the stranding exchanges in question occur over short distances and frequently involve exchanges of words from different grammatical categories: 70% of stranding exchanges occur within a syntactic phrase and 57% involve stems from different grammatical classes.

In respect to those properties, these stranding exchanges look almost like phoneme errors; however, the conditions under which the two kinds of error occur are quite different. When single phonemes misorder, the substituting and replaced segments are featurally similar. (They are identical with respect to the features 'consonantal' and 'vocalic'.) Consonantal substitutions move to the same location relative to the vowel of its new syllable (before it or after it) that it occupies relative to the vowel in the intended syllable. Moreover, the contexts of the consonants' and vowels' new and intended slots are featurally similar. In short, constraints on phoneme errors are phonological, while those on stranding errors are not to any obvious degree.

That word substitutions appear to occur in two varieties also points at least to two distinct phases of utterance planning. In one kind ('salt' for 'pepper'), substitutions are semantically similar to their targets, but not noticeably similar phonologically. In another kind ('apartment' for 'appointment'), the relationship is just the reverse. (Substitutions in the latter category are called 'malapropisms' by Fay and

Cutler, 1977.) It is as if the talker accesses the lexicon twice: once to search using semantic criteria for words to fill appropriate sentential roles and then later using phonological criteria to find utterable word labels.

A final source of information on the phasing of planning processes is provided by 'accommodations'. Consider the following errors (from Garrett, 1980b):

(1) 'I don't know that I'd hear one if I knew it' (for 'I don't know that I'd know one if I heard it').
(2) 'Even the best team losts' (for 'Even the best teams lost').

In the first error, 'know' and 'hear' exchange, and 'hear' strands its past-tense marker. Interestingly, when 'hear', pronounced /hɝ/ in the intended string, moves, it is pronounced /hiyr/ appropriately for its new location. (Letters enclosed in slashes represent phonological segments. A list of symbols likely to be unfamiliar appears in the Appendix.) Also, 'know' plus the stranded affix together are pronounced, not 'knowed' but 'knew'. The accommodations of 'hear' and 'know' to their new settings are lexically dependent. 'Heard' and 'knew' are irregular pasts; they cannot be undone and done respectively without consulting the lexicon. But now consider the second error. Here, the plural morpheme shifts from 'team' to 'lost.' It, too, shows an accommodation but the accommodation is not lexically dependent. The accommodation that the error does show is in the pronunciation of the plural morpheme. Whereas it is pronounced /z/ in 'teams' it is pronounced /s/ after 'lost'. This follows a rule for pronunciation of plural morphemes following /m/ (and most other voiced sounds) versus /t/ (and most other voiceless sounds), but it is insensitive to the fact that 'losts' is not a word.

One possibility, of course, is that some talkers (or some talkers sometimes) consult the lexicon to monitor for errors while others (or the same talkers other times) do not. A different generalization appears to fit the collection of errors of these two types, however. Stranding exchanges that occur over long distances, and shifts (or stranding exchanges: 'the park was trucked') that occur over short distances, may take place during different phases of sentence planning during which talkers are sensitive to different properties of a planned sentence. As the planning domain progressively narrows, focus shifts from grammatical and semantic properties to phonological ones, and from sensitivity to lexical properties to sensitivity only to fully systematic phonological properties of the language.

2.2.4 Slot and Filler Theories: Garrett and Shattuck-Hufnagel

Garrett (1980a, b) has proposed a model that takes an early representation of an utterance in which words and their sentential roles are specified into another, surface structure, representation in which words are ordered as they will be spoken and are appropriately inflected. Shattuck-Hufnagel (1979, 1983) suggests how words of a sentence are encoded phonologically for utterance. I will briefly review both proposals and then describe an important revision by Dell (1980, 1986, 1988).

Garrett calls the two representations that he proposes talkers construct *functional* and *positional*. A functional representation assigns sentential roles to words (or to their underlying semantic contents). Essentially, it represents the meaning of the utterance that that talker intends to convey in a fully transparent fashion. A

positional representation orders words and assigns affixes to them in accordance with the syntactic constraints of the language.

Two major kinds of error may occur as the functional representation is assembled. As the lexicon is searched, a wrong word may be selected and, because the search is sensitive to content and grammatical form, substitutions will be semantically similar to targets and will be from the same syntactic category (e.g. 'salt' for 'pepper'). A second error may occur as words are assigned to sentential roles in the functional representation. For example, two nouns selected for different roles may be exchanged ('I left the briefcase in my cigar').

As the functional representation is formed, a positional 'frame' is also under construction. For insertion into the surface structure frame, the lexicon is searched once more, this time for word labels – that is, sequences of phonological segments that label the word contents inserted into the functional representation. Word substitutions can occur in this search as well, but these now are promoted by phonological similarity.

The frame itself is composed of affixes ('past tense', 'plural') and function words (closed-class words such as 'of', 'the'). Function words and affixes are considered part of the relatively fixed frame on empirical grounds. Each behaves differently from content words in its participation in speech errors. Function words do not generally participate in speech errors at all, and exchanging words and even phonemes jump over them. As for affixes, they are sometimes 'stranded' ('the park was trucked'). If they move, they shift, and they shift only a short distance ('the best team losts'). Garrett suggests that stranding errors occur as word labels are inserted into the positional frame. Apparently, the insertion process generally proceeds syntactic phrase by syntactic phrase; accordingly, errors occur largely within syntactic phrases. Yet the insertion process is rather insensitive to the form classes of inserted words; word exchanges over these short distances frequently violate form class.

As for shift errors, they are not really shifts of the bound morphemes, according to Garrett; rather, bound morphemes are part of the fixed positional frame. Instead, shifts involve misinsertions into the frame such that, for example, in 'even the best team losts' a pair of content words is inserted into the single slot before the plural morpheme.

Explaining phonological segment errors appears to require a model closely analogous to the content frame account of the 'positional level' representation proposed by Garrett. Recall some characteristics of phonological segment errors. They occur in the same varieties as word errors. That is, they appear as anticipations, perseverations, exchanges, as substitutions with no source in the planned utterance, as additions and as deletions. Conditions that promote phonological segment errors are phonological. Substituting and replaced segments are featurally similar and the slots into which substituting segments move have contexts similar to those of their source slots if any; these two contexts are featurally similar, and may even include some of the same consonants or vowels. In addition, consonants move to the same position in a syllable – before the vowel or after it – as they occupy in their source locations.

Exchange errors (e.g. 'it's *p*ast *f*assing') provide one major source of evidence favoring a content frame (slot filler) account of planning at the level of phonological segments. Like anticipations, they establish that words are planned for production in advance of their articulation. But they have an additional, quite remarkable,

characteristic. In an exchange, if segment *B* substitutes for *A* in *A's* slot, then *A* substitutes for *B* in *B's* slot. How can that be explained except by supposing that words to be produced are specified as a frame of consonant and vowel slots and separately as a pool of phonological fillers for the slots? If *B* moves to *A's* slot and so has been used ahead of its time, instead of a deletion occurring at *B's* location – this never happens according to Shattuck-Hufnagel – the leftover *A* appears there.

Shattuck-Hufnagel (1979, 1983) proposes that phonological planning involves building a frame for a string of words to be produced. The frame specifies a sequence of slots for consonants and vowels. Lexical entries for words to be produced supply a pool of consonants and vowels to serve as slot fillers. The frame specifies featural attributes of the segments to fill the slots and a 'scan-copier' searches the pool of consonants and vowels for appropriate fillers. Fillers are appropriate if they have the right features and they occur in a context of segments with features like those surrounding the slot to be filled.

As it fills a slot with a segment, the scan-copier deletes the used segment from the pool. An exchange error occurs, then, when the scan-copier picks a wrong segment for one slot and eliminates it from the pool. Because it makes its selections based on the featural attributes of the segment to fill the slot, an erroneously picked segment is likely to be featurally similar to the target segment. Having selected a wrong segment for insertion into the frame and having eliminated it from the pool, the scan-copier has little choice but to use the left-over segment to fill the slot left empty by the earlier error. If the scan-copier makes two errors – if it selects a wrong segment for a slot and then neglects to eliminate it from the pool – then an anticipation occurs. Perseverations occur when a segment is used correctly, but is not deleted from the pool; the scan-copier then chooses that segment over the correct one to fill a later slot. Substitutions that have no source in the pool of fillers may reflect copying errors from filler pool to frame.

Shattuck-Hufnagel's account of phonological-segment level planning is attractive in at least two ways. First, it explains errors at this level analogously to Garrett's explanation of similar errors on words. Second, the account does a good job of explaining errors that occur and in explaining why nonoccurring errors fail to occur.

A major puzzle that the model raises, however, concerns the need for serial ordering of phonological segments of words in speech planning. There is an excellent reason why planning for the serial ordering of words must occur in speech production. It is precisely the reason offered at the very beginning of the chapter for why novel utterances are produceable and understandable. People routinely produce sequences of familiar words, ordered and affixed according to the syntactic conventions of the language, that they have never produced before. Presumably language users have very few stored sentences. Essentially all sentences are constructed for production.

Clearly, this is not true of words. Words are the familiar parts of which novel utterances are composed. They are produced over and over again with, of course, the same ordering of their component consonants and vowels. So the question remains: Why should speech production planning include a phase during which the phonological segments of words are ordered for output?

The only answer in the theory to the question why phonological segments must be redundantly ordered in speech planning seems to be that they must be or else speech errors would not appear as they do. Obviously this is unsatisfactory. If the

only function of the scan-copy process were to create phoneme errors of speech, talkers would avoid the process. An alternative perspective on the errors and so on the answer to the question is provided by Dell's important revisions to the model, considered next.

2.2.5 Dell's Model

Appealing aspects of the models of Garrett and Shattuck-Hufnagel are that the units of planning they propose are, with the exceptions noted, those of accepted linguistic theory. Also, to a considerable degree, the models generate errors as a natural product of processes of speech planning. That is, the models are not, for the most part, designed to produce errors: they are designed to produce speech.

Unappealing aspects of the model are just those aspects that appear to be there only to explain error patterns. They include, most notably, the two retrievals of words from the lexicon in Garrett's model, needed to explain why some word substitutions are semantically motivated and others are phonologically motivated, and the reordering of already ordered phonemes in Shattuck-Hufnagel's model.

Slightly subtler empirical problems with the model led Dell (1980, 1986, 1988; Dell and Reich, 1980, 1981; see also, MacKay, 1982, 1987; Stemberger, 1985) to an important change in perspective on speech planning and to a new model. The change in perspective retains many central aspects of the models of Garrett and Shattuck-Hufnagel, but it eliminates (or motivates) unappealing attributes just described, and handles the data somewhat better.

The subtle failures of the slot filler models derive from their strict modularity. At the functional level of representation, the system is blind to phonological properties of words, but is sensitive to content and sentential role. When errors occur, they are exchanges between words of a common syntactic class or semantically similar word substitutions. At the positional level and beyond, the system is blind to content, increasingly blind to syntactic category and increasingly sensitive to phonological form. Accordingly, for example, phoneme errors frequently create nonwords, and misorderings occur between words that may differ in parts of speech. This characterization approximately fits the data, but it fails in detail.

Word targets and their substitution errors that are classified as meaning-based are more similar phonologically than are the same words randomly repaired (Dell, 1980; Dell and Reich, 1981). Although focus on semantic attributes of words may predominate when those word errors occur, the planning system is not wholly blind to the phonological properties of words. On the other side, although phoneme errors do often lead the talker to produce nonwords, there is a significant bias for them to create real words both in spontaneous errors (Dell, 1980; Dell and Reich, 1981) and in laboratory induced errors (Baars, Motley and MacKay, 1975). It seems that during a phase of planning for the serial ordering of phonemes, even though phonological properties predominate in promoting errors, the planning system is not wholly blind to lexicality and content.

How can a system be mostly modular, or modular with leaks? Dell suggests one way. Modularity is realized in part by frames of the sort that Garrett and Shattuck-Hufnagel propose. In Dell's model, there are three frames: surface syntax, words and syllables. Words are inserted into syntactic slots, morphemes into word slots and syllable constituents (currently isomorphic with single phonological segments in the model as simulated) into syllable frames. Selection of contents for the frames takes place in the lexicon, which has a special structure.

Dell (see also McClelland and Rumelhart, 1981) proposes a connectionist system in which the lexicon is a hierarchical network of units. Levels of the hierarchy include words, morphemes, syllable constituents, phonemes and features. The lexicon is a network in the sense that there are bidirectional excitatory linkages between each superordinate unit and its subordinate constituents on the next level down.

When 'activation' spreads through the lexicon, it spreads along the linkages. Accordingly, if a word, say 'swimmer', is activated because a talker intends it to be part of a sentence, activation spreads from the word 'swimmer' to its component morphemes, 'swim' and '-er'. From there, activation spreads upward from the component morphemes to each word of the lexicon having the morphemes as components. (In turn, those activated words activate their component morphemes, and so on.) Activation also spreads from activated morphemes downward in turn to their component syllable constituents, phonemes and features. The same principles of spreading operate everywhere, so that each activated phoneme sends activation downward to its features and upward to each word that contains the phoneme. Each feature activates any phoneme that contains it. Chaos, it seems.

The chaos is constrained or even regulated, however, by the process of inserting contents into frames. Figure 9.1 shows how speech production planning works in the model. The talker intends to say 'some swimmers sink'. Activation from the 'idea' for the message is not modeled except that it activates relevant words. Activation spreads from the words as described.

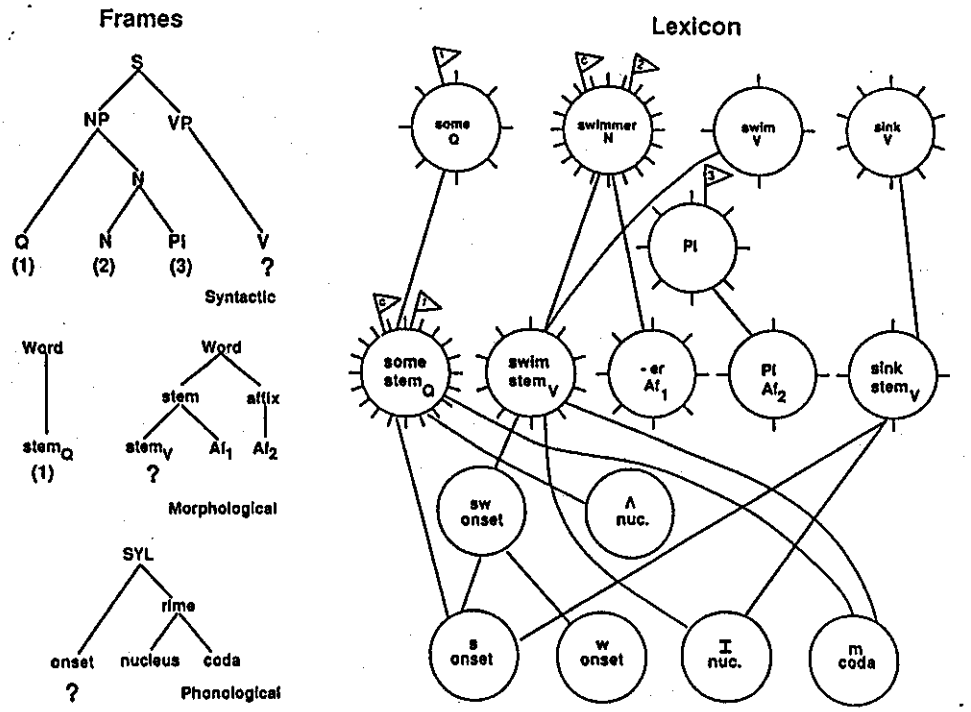


Figure 9.1. *Dell's model of language production including a connectionist lexicon and, to the left, a set of frames for constructing a sentence. See text for further explanation. [Adapted from Dell, 1986.]*

In the course of planning, talkers build a surface structure frame for the utterance. For each slot in the frame to be filled by a word, they build a word frame with slots for each morpheme. Finally, a syllable frame is constructed for each syllable of each morpheme.

The frame for 'some swimmers sink' specifies a quantifier for the first word slot. In the lexicon, the quantifier with the highest level of activation is selected to fill the slot. Most of the time that is the intended quantifier because it received initial activation from the intended message idea. (In the figure, a word in the lexicon is tagged with an order tag when it has been 'inserted' into the frame; accordingly, 'some' has been tagged with the number 1.) When the quantifier is picked, its activation level is set to zero, but it soon rebounds because its similar neighbors are active and will reactivate it.

The frame next calls for a noun and the selection process continues with the noun having the highest activation being selected for insertion. To construct the word frame, the tagged quantifier in the lexicon is designated the 'current node'. By virtue of that, it is given additional activation and its word frame is built. After some period of time to allow spreading to occur (and the amount of time varies with rate of talking), a morpheme of the appropriate sort (stem or affix) for the first available slot is selected; the selected morpheme is the one of the appropriate sort that has the highest activation. Now the next word in the syntactic frame is designated the current node at that level, while the just-selected morpheme is current at its level. This allows its syllable frame(s) to be built.

Insertions take place at three levels in the lexicon: words, morphemes and phonemes (not distinguished from syllable constituents). Errors take place during the insertion process; accordingly, only words, morphemes and phonemes will serves as units in errors.

Errors can occur in just one way. They occur if a unit, appropriate for the frame slot being filled, but not the intended filler, is more highly activated at the time of insertion than is the intended unit. This will happen sometimes because of the spreading process in the lexicon.

At the word level, a substitution error will occur if an unintended word of the appropriate syntactic class for the to-be-filled slot in the syntactic frame has higher activation than the intended word. Words that will be activated include other intended words for the utterance. Accordingly, sometimes exchanges, anticipations and perseverations will occur. On other occasions, a word not in the intended string may be highly activated to the extent that it is semantically and/or phonologically similar to the intended word.

An exchange occurs in the following way. At the time *A*'s slot is to be filled in the frame, a word of the intended utterance, *B*, is more highly activated than *A*, and so *B* is selected to fill *A*'s slot. Having been selected, *B*'s activation level is set to zero, but it soon begins to rebound because its similar neighbors are active and activate it anew. When *B*'s slot is to be filled, *A* may be more active than *B*, because *B*'s activation level has been set to zero. Therefore *A* is selected and the error is an exchange. If *B* has rebounded and surpassed *A*, then *B* is selected and the error is an anticipation. (Notice that this predicts a change in the relative frequencies of exchanges and anticipations as speech rate increases; at fast rates, exchanges will be relatively more likely than at slow rates, because at fast rates *B* will not have time to rebound after being set to zero. Dell (1986) has found the appropriate interaction between speech rate and error type in an experiment in which phoneme errors are induced.)

Stranding errors ('the park was trucked', 'some sinkers swim') occur during insertion of morphemes into word frames, while phoneme errors occur during insertion of phonemes into syllable frames. These latter errors show a lexical bias because of the spread of activation from words through morphemes to phonemes. Substitutions are featurally similar because of the bidirectional spread of activation for phonemes to features. The contexts of misordering phonemes are similar because words that provide similar contexts for different phonemes will activate each other via their shared phonemes and features.

Dell has simulated a small lexicon and insertion process on the computer at the level of phonological selection. In general, it produces phoneme errors of the right sort in the right proportions. In addition, it shows a lexical bias in phoneme errors. More interestingly, perhaps, the model makes some unexpected predictions that are borne out by the simulation.

One, already mentioned, is the change in the ratio of exchange to anticipation errors as speech rate changes. A second is that the lexical bias will be reduced as speech rate increases, because there is less time for spreading of activation through the lexical network. A striking prediction is that low frequency words that are homophones of function words will inherit the resistance of function words to participation in speech errors.

Garrett (1980a) had included function words in the positional frame largely because function words very rarely participate in speech errors. Dell ascribes the resistance of function words to error to their high frequency. In his lexicon, words have resting activation levels that increase with their frequency of production. In effect, a talker is chronically prepared to produce words that are produced frequently. Words with already high activation levels most readily become the most highly activated words when insertion rules select words for insertion into the syntactic frame. Therefore, slips are rare on function words. Low frequency homophones of function words will inherit function words' error resistance to a considerable degree because they share the same syllables, phonemes and features.

Research shows that low frequency words are more likely to participate in speech errors both in spontaneous errors (Stemberger and MacWhinney, 1986) and in experimentally elicited errors (Dell, 1990). However, the frequency difference vanishes for function words as compared to their low frequency homophones (Dell, 1990).

In short, Dell's model preserves the appealing features of Garrett's and Shattuck-Hufnagel's models, and it eliminates (or perhaps motivates) the less appealing aspects of those models. Dell's model does not yet offer a reason why word onsets are so vulnerable to error. (However, Dell (1990) suggests that an adjustment to the model, such that the process of phoneme insertion into the syllable frame is made more serial and less parallel than it is currently, might give rise to an initialness effect.)

2.3 Performance of the Speech Plan

A conclusion from the collection of data and theory on speech errors is that units of competence as suggested by linguistic theories, and of planning as suggested by data from speech errors, are closely convergent. This means that elements that

behave autonomously in language systems serve as autonomous units of speech planning as well. This is certainly good news. Language systems evolve and change during communicative interactions among speakers. They are, at least in part, fossilizations of consistent aspects of those interactions. Accordingly, it seems that the units of those systems should converge with units that talkers use in planning to speak.

The whole picture changes, however, when we turn from speech planning to speech production. Barring evidence to the contrary, one might expect to find a fairly direct relationship between planned phonological segments (the final planning stage in the models above) and activities of the vocal tract. Also, because planning units seem to correspond so closely to linguistic units, one might expect to see the feature columns of Chomsky and Halle's linear phonology realized in some fashion in the vocal tract.

Neither expectation is met, however. There is no neat correspondence between planned units and vocal tract activity and, indeed, correspondence is progressively more difficult to find the lower in the planning hierarchy the unit. That is, it is harder to find phonological segments in vocal tract activity than to find words, even though one might expect it to be the final planning stage that drives motor activity.

Before looking at vocal tract activity and its relationship to units of planning and of language competence, I provide an overview of the physical systems used to produce speech.

2.3.1 The Respiratory System, the Larynx and the Vocal Tract

Four anatomical systems are centrally involved in speech (Figure 9.2); they are: the respiratory system, the larynx, the nasal cavity and the oral cavity. The latter two constitute the *vocal tract*. Articulators in the oral cavity include the velum (the soft palate), the jaw, the upper and lower lips, and the tongue. To a degree, the tongue tip and the dorsum or tongue body are independently controlled.

Generally, speech is produced on an expiratory airflow and the respiratory cycle is modified during speech so that it occupies a considerably greater part of a breathing cycle than it does in vegetative breathing. The larynx, seated on the trachea, houses the vocal folds, which provide the primary sound source in speech. During voiced sounds, the vocal folds are approximated (adducted); while for unvoiced sounds, they are abducted. The adducted folds periodically stop the flow of air from the lungs from passing into the vocal tract. When the vocal folds stop the airstream, pressure builds up beneath them ('subglottal pressure'), and eventually the pressure blows the folds apart. The folds quickly close again, leading to another cycle of subglottal pressure build-up and release. During voiced phonemes, the vocal folds open and close periodically, cyclically releasing puffs of air into the vocal tract. The puffs of air contain energy at the frequency at which the vocal folds open and close (the 'fundamental frequency' or f_0 of voice) and, at successively lower intensities, at harmonics of the fundamental.

Vowels are produced with no obstruction to the airstream. They are distinguished one from the other largely by the positioning of the tongue body in the oral cavity. In high (or 'close') vowels, the tongue body is close to the palate, either toward the front of the mouth (as in /iy/) or toward the back (/uw/). In low ('open') vowels, the jaw is lowered so that the tongue body does not approach a

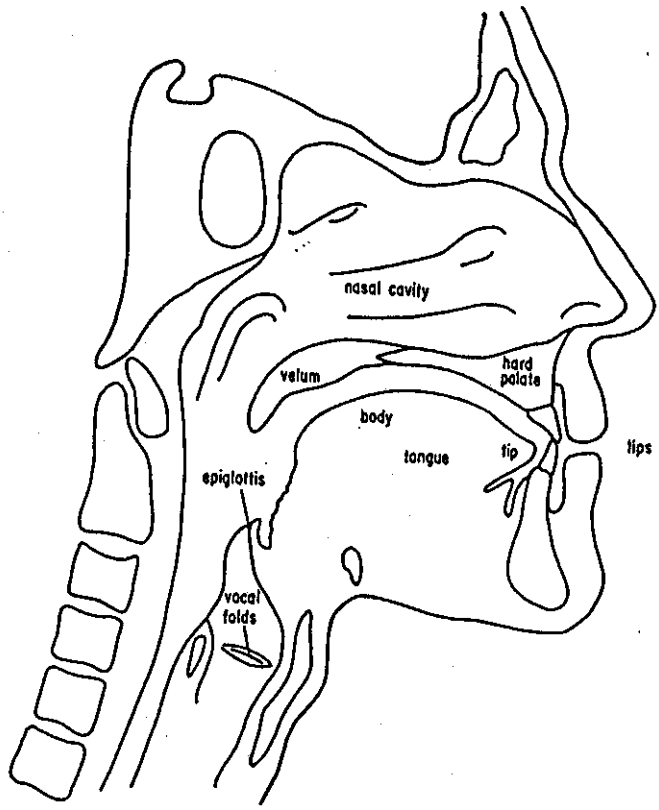


Figure 9.2. *A model of the vocal tract.*

point of constriction with the palate. Some vowels are produced with rounding of the lips (e.g. /uw/ or /ow/ in English); others are produced with lips retracted (/iy/). The movements of the tongue and the rounding or protrusion of the lips create cavities of different sizes and shapes in the vocal tract. In /uw/, a high back vowel, the cavity in front of the tongue body is long, while the back cavity is short; in the production of /iy/, the front cavity is short and the back cavity is long.

For vowels, the vocal tract serves as a filter for the acoustic energy produced at the laryngeal source. The cavities of the vocal tract have characteristic resonances ('formants') that depend on their lengths, and acoustic energy outside of the resonances is attenuated, yielding bands of energy in the vicinity of each formant.

Consonants are produced by obstructions to the airstream, for example at the lips (/b/, /p/, /m/) or at the hard palate using the tongue tip (/t/, /n/, /s/). During stop consonants (in English, /b, p, d, t, g, k/) the airflow is stopped temporarily and noise is produced at the point of constriction as the constriction is released. During fricatives (/s/, /z/, /f/, /v/, for example) an articulator closely approaches a fixed structure of the vocal tract, but leaves a narrow channel through which the airstream passes and becomes turbulent. During nasal segments, the velum lowers to allow air to pass through the nasal cavity, which serves as a resonator of fixed length.

2.3.2 Articulatory Dynamics

How do activities of the vocal tract realize a speech plan? In the last stage of planning, according to models reviewed earlier, phonemes are ordered for production. In linear phonologies, phonemes are bundles of simultaneous featural attributes. Were the relationship between linguistic competence and plan, and between plan and behavior, simple then one might expect to see speakers adopt successive postures of the vocal tract, each posture manifesting the set of featural attributes of a feature column. Short rests between postures might signal a morpheme boundary while longer rests signal a word boundary. Of course, because the vocal tract is a physical system, it cannot shift instantaneously from posture to posture. Accordingly, between postures one should see transitional phases.

However, vocal tract activity for speech looks nothing like that. During fluent speech, there are essentially no intervals in time when postures of the vocal tract are held. Relatedly, vocal tract activity cannot be partitioned into intervals that count as achievement of the features of a phoneme and others that count as transitions between phonemes. During speech, a great many different things go on at once; activities relating to the realization of a single phoneme are not synchronized, and activities for different phonemes overlap. Consider an instance in which it is possible to identify a feature (nasality) with a vocal tract movement (lowering the velum). One might expect the point of maximum lowering – or perhaps the point in lowering when air begins to pass through the nasal cavity (nasal coupling) – to count as achievement of the feature, [+nasal]. Accordingly, that point in time should coincide with achievement of the oral constriction for the nasal consonant if the 'absolute slicing' hypothesis is to hold for speech production. But it need not. Krakow (1988) has found that the point of maximum lowering of the velum for syllable final /m/ considerably precedes achievement of lip closure for /m/. The feature [+nasal] appears to be phase shifted relative to those that lip closure realizes, and it overlaps with gestures for a preceding vowel.

This overlap is known as 'coarticulation', and it is pervasive in speech. Coarticulation ensures that absolute slicing (temporal slicing in speech production) will never serve to isolate all and only movements associated with individual phonemic segments. Nor, for that matter, will 'spatial slicing'. In general, it is not possible to identify a movement of an articulator with one and only one phoneme. For example, closing movements of the jaw for a consonant such as /b/ are less extensive in the neighborhood of an open vowel than in the neighborhood of a close vowel (Sussman, MacNeilage and Hanson, 1973), and jaw height during closure is lower for /b/ before /a/ than for /b/ before /iy/ (cf. Keating, 1990). The consonant and vowel make overlapping demands on the jaw and the movement of the jaw reflects a compromise.

As MacNeilage and Ladefoged (1976) point out, initial reaction in the 1960 s to the discovery of the pervasiveness of context-conditioned variability of vocal tract movements owing to coarticulation was that coarticulation reflects mechanical constraints on vocal tract actions. The speech plan reflects an ideal that the vocal tract cannot realize, because it is a physical system. In fact, however, mechanical constraints cannot explain most of coarticulation, in part because much of it, like velum lowering for a nasal consonant, is *anticipatory* in direction.

Whatever the source of coarticulatory variability, it appears to create a considerable mismatch between, on the one hand, the characterization of phonological

segments in linear phonologies and even in speech plans and, on the other hand, the behavior of the vocal tract. How is the apparent mismatch to be handled?

According to MacNeilage and Ladefoged (1976): 'This has led ... to an increasing realization of the inappropriateness of conceptualizing the dynamic processes of articulation itself into discrete, static, context-free linguistic categories such as "phoneme" and "distinctive feature"' (p. 90).

This does not mean that these linguistic categories should be abandoned entirely, however: 'Instead it seems to require that they be recognized even more than before as too abstract to characterize the actual behavior of the articulators themselves. They are, therefore, at present better confined to primarily characterizing earlier premotor stages of the production process, as revealed by speech errors, and to reflecting regularities at the message level (Fant, 1962) of the structure of language, such as those noted by phonologists' (p. 90).

Other investigators have expressed a similar view. For example, Kelso, Saltzman and Tuller (1986) write that: 'a final implication of the view presented here is that "segments" or phonological units as typically defined by linguists may not be relevant to the speech production system' (p. 46).

If the units apparently *are* relevant to speech planning, but apparently are not relevant to the speech production system, then where are we? One interpretation of this state of affairs is that there is, simply, a distinction between segments in the minds of language users and public behaviors of speakers. Hammarberg (1976) seems to hold this view: 'Segments cannot be objectively observed to exist in the speech signal nor in the flow of articulatory movements ... [T]he concept of segment is brought to bear *a priori* on the study of physical–physiological aspects of language' (p. 355).

This is unfortunate if true, because it considerably complicates the path of a communicative exchange. Speakers plan to produce a message consisting of words, themselves composed of sequences of consonants and vowels. But if consonants and vowels like that are not able to make public appearances in the vocal tract then they cannot structure the acoustic speech signal. Listeners must be like paleontologists (as Neisser, 1967, suggests) – who, finding small bone fragments, reconstruct an entire dinosaur; that is, listeners must use the acoustic signal as a collection of fragmentary hints as to the phonological word labels of a talker's intended linguistic message. Perhaps that is the way it is; perhaps not, however.

A different reason for the mismatch between the phonological segments of linguistic theory and vocal tract actions, as described by production researchers, may be that phonological segments of linear phonologies do not accurately represent real phonological segments of languages. It may be relevant that, while some of their properties are validated by speech errors, not all of them are.

Speech errors seem to reveal the following attributes of phonemes. They are somewhat autonomous, and their featural attributes are cohesive. Phonological segments are featurally distinctive. They are serially ordered in words.

Apparently, there is nothing in this characterization that requires the featural realization of phonemes not only to be cohesive, but also to be simultaneous in the way they are represented in the feature columns of linear phonologies. If cohesion can be achieved in some way other than by simultaneity, then the absence of synchrony in realization of the features of produced consonants and vowels does not separate planned from produced segments. Nor is there anything about planned segments that precludes temporal overlap in their realization as long as

any overlap preserves the autonomy of each phonological segment and its ordering relative to neighbors. If phonemes need not be seen as discrete sequences of simultaneous features in a speech plan, is it possible that they need not be seen that way in the languages user's 'competence' either? That is, is the absolute slicing hypothesis any more required for the work that phonemes do in phonologies than for the work they do in speech plans? Recent developments of 'nonlinear' phonologies (Browman and Goldstein, 1986; Clements, 1985; Goldsmith, 1976) suggest that it is not; indeed, absolute slicing is not even tenable. These new phonologies help to reduce the chasm between phonological segments of linguistic competence and vocal tract behaviors by eliminating the absolute slicing constraint. They are reviewed briefly in Section 3.

Another reason for the mismatch between phonological segments as components of linguistic competence or of speech plans and as vocal tract activity is likely to be that we are not looking at vocal tract activity in a way that best reveals talkers' coordinations and controls. Coarticulation occurs, and so there is considerable overlap of movements associated with different phonological segments and context-conditioned variability of movement. Is there any way of looking at activity like this and recovering a phonological segmental structure?

In 1970, MacNeilage proposed a shift in perspective that looked promising. He proposed that a context-free articulatory correlate of a phonological segment be found in 'spatial targets' achieved in the vocal tract. Even though a target may be approached in different ways depending on its predecessor target, the target itself might be an invariant attribute of a segment. Similar proposals have been offered to explain equifinality in performance of pointing tasks (Bizzi and Polit, 1979; see also Fel'dman and Latash, 1982) and eye–head movements (Bizzi and Polit, 1979). MacNeilage proposed specifically that target muscle lengths are set by the gamma motoneurons of the nervous system that innervate muscle spindles. This should lead to movements toward the target muscle lengths and so toward invariant postures of the vocal tract articulators, independent of their starting positions.

In its specific proposals that invariant spatial targets represent articulated phonemes and that targets are achieved by setting target muscle lengths via the gamma system and muscle spindles, the theory is wrong. When a talker achieves bilabial closure for /b/, for example, the point in space where closure is achieved (relative, say, to the teeth) changes depending on the jaw's contribution to closure. In turn, those different spatial locations imply different lengths of muscles associated with jaw and lip locations.

Although the theory is wrong in its particular form, it is right in one important respect. For at least some attributes of some phonemes, an abstract invariant is achieved in the vocal tract whenever those attributes are planned and executed. When bilabial stops are produced, the lips close; when alveolar stops are produced, the tongue tip achieves contact with the hard palate; when velar stops are produced, the tongue body achieves contact with the soft palate. Possibly, there is a measure of invariance in speech production despite the context-conditioned variability introduced by coarticulation. The invariance is even more abstract and less particular than MacNeilage envisaged in 1970. New theoretical approaches in the domain of speech production attempt to make use of invariants like this. The approaches are described in Section 3.

If we discount, at least for the present, the possibility that there simply are irreconcilable incompatibilities between structure in speech activity and in the

phonological competences and speech plans of language users, we can look for ways to eliminate apparent incompatibilities and so to bring the domains closer together. As MacNeilage and Ladefoged (1976) point out, making revisions to theory should be a two-way street. The phonologies of languages perhaps need developing with attention to the nature of vocal tracts that will realize speech. Theories of speech production need developing with attention to the essential properties of phonological segments (possibly: autonomous, cohesive distinctive attributes, serially ordered) that must be realized somehow in vocal tract activity if phonological segments are to make public appearances. On both sides, theorists must be open to recognizing old conceptualizations as conceptualizations, not as truths about phonological segments, about coarticulation and vocal tract activity and about the possible relationships between the realms of knowing (competence), planning and doing. Section 3 attempts to chart such rapprochement as linguistic theories and theories of speech production have been able to attain to date.

3 NONLINEAR PHONOLOGIES AND ANOTHER LOOK AT SPEECH PRODUCTION

Chomsky and Halle (1968) acknowledge as a major shortcoming of their work its inattention to the 'intrinsic content' of phonetic features (that is, inattention to their physical implementations). In large part, Chomsky and Halle's theory represented phonology as an imposition of a formal system on the vocal tract, the capabilities of which did not shape the formal system. The fact that simultaneous realization of the features of a phoneme is not possible for a vocal tract was not a reason to exclude that manner of representing phonological segments.

As Chomsky and Halle acknowledged, that is not a wholly realistic approach. There are clear indications, both in the segmental inventories of languages and in systematic processes in which the segments participate, that language structure develops with considerable regard for vocal tract capabilities and dispositions.

For example, Locke (1983) reports that the most common consonants in languages of the world are also most commonly transcribed in the babbles of prelinguistic infants – even of deaf infants. Lindblom (1986) has shown that two principles, maximization of perceptual distinctiveness and minimization of articulatory complexity, jointly do a good job of predicting the composition of phonemic–segmental inventories of languages of the world.

As for phonological processes, as MacNeilage and Ladefoged (1976) point out, they sometimes resemble physiological constraints or dispositions. Their example is the tendency in most languages that have been studied (Chen, 1970; Flege, 1988; Mack, 1982; Raphael, 1975) for vowels to be durationally longer before voiced than voiceless consonants. The reason for this is not entirely understood. However, it appears to reflect, in part, the faster closing gestures for voiceless than for voiced consonants (Chen, 1970; Summers, 1987). In turn, this difference may reflect the need to achieve a tighter seal during voiceless closures, which are associated with higher intraoral pressures than voiced closures (Lubker and Parris, 1970). In some languages including English, however, the vowel duration difference is considerably greater than can be explained by the difference in closing velocities (Chen,

1970; Flege, 1988; Flege and Port, 1981; Mack, 1982). In these languages the vowel duration difference, that may be a byproduct of different closing gestures for voiced and voiceless consonants, has been elevated into the phonology of the language as a regular process ('rule'). The rule has apparently in some way been 'triggered' (MacNeilage and Ladefoged, 1976) by the corresponding physiological disposition.

If segment inventories and phonological processes reflect articulatory capabilities, then an approach to phonology that largely ignores the vocal tract is likely to be unrealistic. Next, I briefly describe recent approaches to a theory of phonology that may be more realistic in this regard.

3.1 Nonlinear Phonologies and the Articulatory Phonology of Browman and Goldstein

A problem for representation of phonemes as feature columns arises in instances in which features have a scope smaller or larger than a single column. Consider the feature [+nasal], present in English in /m/, /n/ and /ŋ/. Some languages have so-called 'prenasalized' or 'postnasalized' consonants. Whereas most features of these segments span the whole segment's extent, nasality does not. In prenasalized stops, the first part of the consonant is nasalized, while the remainder is not. Postnasalized stops have the opposite pattern. Because prenasalized and postnasalized consonants have the duration of a single segment, are featurally identical across the segment except in respect to nasality, and have the distribution of single consonants (that is, they occur in contexts where clusters of consonants are otherwise not permitted; see Anderson, 1976), they are identified as single consonants, not as a sequence of two. However, there is no satisfactory way to represent the change in the nasal feature part-way through the segment in a linear phonology.

In other languages, nasality poses a different kind of problem for linear phonologies. These languages exhibit 'nasal harmony' in which a single nasal feature extends over more than one feature column. In the language Gokona (Hyman, 1982), if the first consonant of a word is nasal, all of the segments of the word are nasalized, or if not, but a word-internal vowel is a nasal vowel, every segment after the vowel must be nasalized. In another language, Terena (van der Hulst and Smith, 1982), the first person possessive is signaled by a spreading of nasality that begins at the left edge of a word and spreads until stopped by an obstruent consonant. The obstruent becomes a prenasalized obstruent. That is, the nasality extends half-way into its 'feature column' (so that [owoku] ('his house') becomes [õwõŋ̃gu] ('my house')).

These instances in which a feature's span is either less or more than a whole column are not restricted to cases involving the feature [nasal]. In tone languages, tones may likewise have a span that is less or more than a whole segment (Goldsmith, 1976), as can various vocalic features in languages with diphthongs or vowel harmony.

One implication of these observations is that columns of simultaneous features cannot represent all of the relationships among the featural attributes of the segments of a word. A second is that, when two features have different domains, they are manifesting a measure of independence or autonomy that a phonology will need to capture.

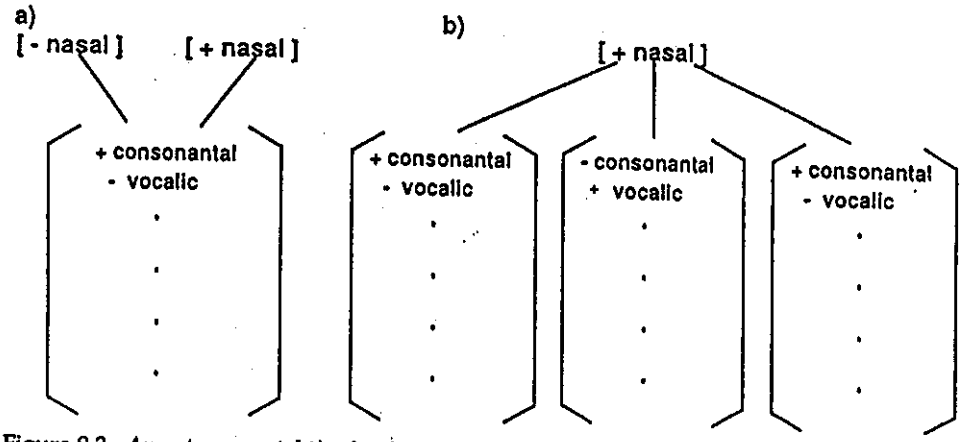


Figure 9.3. *An autosegmental tier for the nasal feature to capture: (a) postnasalized consonants, and (b) nasal harmony.*

A way to express the independence of the feature [+nasal] in languages with postnasalized stops might be as in Figure 9.3a; and in languages with nasal harmony as in Figure 9.3b.

Phonologies that adopt notations like this, in which some features occupy different ('autosegmental') tiers from others, and so may have different domains of influence, are nonlinear or autosegmental phonologies.

It is of interest, however, to go beyond the language-specific representations of Figure 9.3 to develop a 'universal' phonology – that is, a phonological system that expresses the *possibilities* for autonomy among features rather than expressing just the ones that are highlighted in the phonologies of just a particular language being studied.

The development of an 'articulatory phonology' by Browman and Goldstein (1986; 1990; see also, Clements, 1985; Sagey, 1986) reflects an understanding that possibilities for autonomy are set by the character of the vocal tract. By the same token, possible 'features' are determined in part by possible vocal tract actions, and the theory of articulatory phonology also makes an effort to identify its primitives with possible (and actual) vocal tract behaviors in speech.

In Browman and Goldstein's theory, fundamental units are not phonemes or features, but 'gestures'. They are linguistically significant goal-directed actions in the vocal tract defined by: (1) the *end-effector* of the action (for oral constrictions for consonants and vowels, the end-effector is the articulator that makes the constriction); and (2) by values of parameters of a dynamical equation that describes the appropriate action. For oral constriction gestures, the parameter values define a constriction in the appropriate location, and of the appropriate degree (closed for stops, a narrow opening for fricatives, etc.), and a final value ('stiffness' or *k*) regulates rate of approach to the constriction. In addition to gestures for oral constrictions, there are gestures of the velum to open the velar port for nasal consonants or vowels and to raise it for oral ones, and there are gestures of the glottis (that is, vocal fold gestures) to voice or devoice a segment of speech.

Although gestures are identified in part by their end-effectors, the movements that realize oral constrictions are achieved not only by the end-effector itself, but by

supporting articulators as well. Accordingly, gestures to realize values of vocal tract variables, such as the appropriate tongue tip (=end-effector) constriction degree, often imply coordinated involvement of more than one articulator (in the example, the tongue and the jaw).

Consonants and vowels are realized by one or more gestures. For example, /m/ involves an oral constriction gesture and a gesture of the velum; /p/ shares its constriction gesture with /m/, but it lacks the gesture of the velum and includes a devoicing gesture.

The set of gestures possible for the vocal tract during speech has an implicit hierarchical structure that derives from the structure of the vocal tract itself. That structure displayed in Figure 9.4 also implies different degrees of autonomy among gestures and different likelihoods of gestures being grouped in phonological processes or in other classifications of segments.

In place of successive feature columns to represent words, Browman and Goldstein substitute a 'gestural score' such as that in Figure 9.5 for the word 'Tom.'

The association lines between 'root nodes' in the figure and gestural parameters reflect a clustering among the gestures associated with the initial consonant, vowel and final consonant of a word, but also reflect the temporal relationships among the gestures. (Notice, for example, that the glottal devoicing gesture for /t/ does not coincide wholly with the oral constriction gesture. Rather, it is offset to ensure aspiration of the consonant on release.) The wide bracketing of the vowel's gesture reflects its coarticulatory span, while the association lines from the root node to the gesture give its serial order.

Now let us return to the incompatibilities between phonological knowledge as characterized by a linear phonology and vocal tract activity for speech and consider how they have been reduced by this new phonological theory. In fact, they have been considerably reduced. In the new theory, gestures are inherently dynamic not

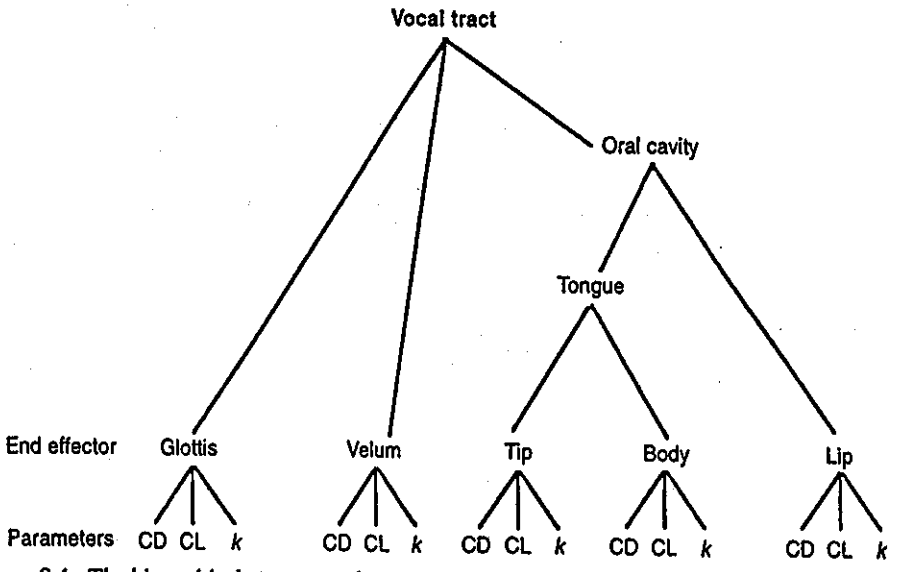


Figure 9.4. *The hierarchical structure of gestures. CD, constriction degree; CL, constriction location; k, stiffness value. [Adapted from Browman and Goldstein,* 1990.*]*

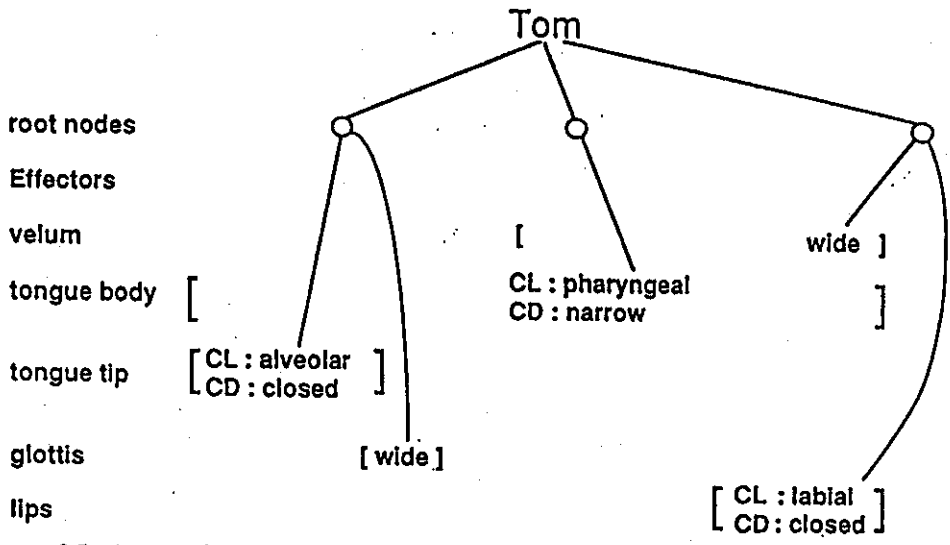


Figure 9.5. *A gestural score for 'Tom'. CD, constriction degree; CL, constriction location.*

static, and they are movements that actually occur in the vocal tract. In addition, the gestures of a consonant or vowel are not necessarily simultaneous and they can overlap with gestures of other segments. In the theory, cohesion among gestures of a consonant or vowel is specified, but not by simultaneity as it is in Chomsky and Halle's theory. The association lines indicate that there is cohesion, but the means by which the vocal tract can achieve it are not specified; that is left to the vocal tract. Likewise, lines to root nodes specify the serial order of segments, but they do not disallow overlap among the ordered segments. Since the oral constriction gesture for a consonant will hide that for a vowel (because it is more extreme), a vocalic gesture's order is signaled by the interval in a word in which it is not hidden.

The new theory of phonological competence is more compatible with speech as produced, as just outlined, but, in being in closer alignment with articulation, is it less compatible with the theories of speech planning outlined in Section 2? Apparently it is not. Were the lexical structure in Figure 9.5 to replace the bottom third of Dell's model in Figure 9.1, apparently no damage would accrue to the theory's ability to handle the facts of speech planning as revealed by errors.

I next consider developments in the theory of speech production that further reduce apparent incompatibilities between speech as known and as produced.

3.2 Speech Production

A problem in relating vocal tract behavior for speech to linguistic units and to planned units is that descriptions of the behaviors yielded by production research tend to be too fine-grained. That is, they are not behaviors at the scale of individual phonemes or even their features. Instead, observations of the vocal tract generally

are of muscle activity or of movements of the individual articulators. However, neither individual muscles nor even individual articulators produce individual phonetic segments, or, for the most part, their features. Moreover, both muscle activity (MacNeilage and DeClerk, 1969) and movements of individual articulators (Sussman, MacNeilage and Hanson, 1973) are context-sensitive, reflecting converging influences of multiple phonetic segments. Thus, not only is the scale wrong, but the specificity of relationship between behavior and message unit is also absent.

Despite that, however, at a different level of description, there is specificity at least between some attributes of consonants and vowels and activity of the vocal tract. For example, as noted earlier, talkers always achieve lip closure for /b/, /p/ and /m/. Even though movements of the three articulators that achieve bilabial closure – the upper lip, the lower lip and the jaw – are context-sensitive, their joint action is invariant.

That there is a lack of invariance at a fine-grained level of analysis, but invariance (for at least some features of some phonetic segments) at a more coarse-grained level implies that somehow goals at the coarser-grained level are constraining finer-grained events (Fowler *et al.*, 1980; cf. Pattee, 1973; Weiss, 1941). It is as if, in bilabial closure, the jaw and lips are being regulated so that they can do whatever else they are disposed to as long as one thing they do is jointly to achieve lip closure.

Constraints like this imply coordination among articulators that respect the constraint. In turn, coordination implies a lack of independence among articulators and so a loss of freedom of movement. Except in producing bilabial obstruents, the jaw and lips are not required to achieve lip closure and they are free to move in ways that will not achieve it. Crucially, when freedom is lost, it is not lost in a hit or miss way. Rather, it is given up precisely to achieve some coarse-grained goal. Loss of independence at a finer level of description of a system brings about coherent functioning at a more macroscopic scale.

Examples like that of bilabial closure imply that there is a real coarse-grained level of organization of the speech system that may allow a closer mapping between vocal tract activity and the units it supposedly realizes. That level is macroscopic enough at least to encompass constraints on groups of articulators to achieve such ends as bilabial closure. Evidence supports the idea that there is such a level of organization.

3.2.1 Vowel Production with Bite Blocks and the Theory of Predictive Simulation

Speakers sometimes talk with a pipe or pencil between their teeth. Although the speech they produce is audibly distinct from speech produced without an obstruction that immobilizes the jaw, it is highly intelligible. Considerations of intelligibility aside, would a speaker choose to speak with a pipe or pencil between the teeth were that to require an entire reorganization of normal means of controlling the articulators? But how can immobilization of the jaw not require a reorganization, given that the jaw is normally quite active in speech? Possibly reorganization is not required because the speaker's means of controlling the articulators for speech include implementing 'constraints' of the sort considered briefly above, and these

constraints build in the means by which a pipe may be compensated for. They build in compensations for jaw immobility not because pipe speech is anticipated, but because coarticulatory influences on the jaw must be handled.

In 1971, Lindblom and Sundberg published findings which suggested that speakers can produce acoustically normal or near-normal vowels with a bite block clenched between the teeth that not only immobilized the jaw but immobilized it in a position uncharacteristic for the vowels being produced. A great many replications and extensions of this experiment have been reported (e.g. Folkins and Zimmermann, 1981; Fowler and Turvey, 1980; Lindblom, Lubker and Gay, 1979; Lindblom *et al.*, 1987). The results are clear: vowels produced with a bite block are near-normal articulatorily (Gay, Lindblom and Lubker, 1981), acoustically (Lindblom, Lubker and Gay, 1979) and perceptually (Lubker, 1979).

Lindblom, Lubker and Gay (1979) offer an explanation for speakers' remarkable abilities to compensate. They suggest that phonetic segments are specified in memory as a set of sensory goals (see also Perkell, 1980). Associated with the goals are motor commands for achieving them in the absence of perturbation, coarticulatory or otherwise. Were the motor system for speech wholly 'open-loop' – that is, uninfluenced by feedback from the periphery – these motor commands would not realize the sensory goals for the vowel most of the time, because coarticulating segments would change the effects of the motor commands at the articulatory periphery. Were the motor system wholly 'closed loop' – that is, sensitive to, and regulated by, feedback – these effects of coarticulation could be undone, but only after they had already had their deleterious effects on vowel production. Predictive simulation was proposed as an improvement on open- and closed-loop modes of motor control.

In the predictive simulation system, effects of stored motor commands on the ongoing movements of articulators are simulated using sensory feedback as information about the current state of the vocal tract. Discrepancies between stored sensory goals and simulated sensory consequences of the motor commands are used to adjust the motor commands so that they will achieve the stored sensory goals for the vowel. A system like this will compensate for coarticulatory influences in the vocal tract and also for a bite block that immobilizes the jaw.

The model makes one prediction that has not yet been tested on bite block vowels, but that has been tested on perturbed consonants: it is that compensation should not be possible for perturbations introduced after movement toward the segment's target has begun. That is, once the revised motor commands have been allowed to affect the vocal tract, the system should behave either in an open-loop or in a closed-loop fashion. Perturbations after that will either go uncorrected or else they will be corrected after errors occur. This prediction is unconfirmed for consonant production.

3.2.2 Consonant Production and Synergies

A perturbation applied to the jaw or lower lip during achievement of bilabial closure that pulls the articulator away from its direction of travel does not prevent lip closure for a bilabial consonant (Abbs and Gracco, 1984; Folkins and Abbs, 1975, 1976; Gracco and Abbs, 1985; Kelso *et al.*, 1984). Rather, short-latency compensatory responses occur in the perturbed articulator as well as in the other articulators that

contribute to closure so that their joint goal (bilabial closure in the example) is achieved. As in the research using a bite block, compensation is not complete; lip closure may be achieved with less compression of the lips than on unperturbed trials (Abbs and Gracco, 1984) and slightly later than on unperturbed trials (by about 25 ms in the research of Folkins and Abbs, 1975). However, the lips do close, and a successful bilabial consonant is produced.

In this research, in contrast to research on bite block vowels, perturbations are transient and they are applied during an ongoing closure movement for a consonant. In terms of the predictive simulation model, the perturbations are applied after motor commands for realizing a sensory goal have been revised and allowed to affect vocal tract movements. Yet compensation occurs and it may begin to occur within 20–30 ms of the perturbation (Kelso *et al.*, 1984). That compensation occurs at all under these conditions appears incompatible with the predictive simulation model. However, even if the model were modified to allow changes at the periphery to bring about further modification to motor commands, 20–30 ms of latency is an implausibly short time to allow for a new simulation and revision to the commands.

Evidence suggests that, even though the origin of compensation must be low level (that is, not too many synapses away from the periphery), the compensations are functional. Shaiman and Abbs (1987) show that when an articulator not involved in a closing movement is perturbed (when the upper lip is perturbed during closing for /f/), the closing response is no different from that on unperturbed trials. Compatibly (Kelso *et al.*, 1984), when an articulator that is involved in closing is perturbed (the jaw in alveolar closure), compensatory movements are not observed in articulators uninvolved with closure (the lips in the example).[1]

What kind of system could implement a constraint spanning more than one articulator that achieves coarse-grained closure goals? A system with these capabilities is a *synergy* (Gelfand *et al.*, 1971) or *coordinative structure* (Easton, 1972). As Lee (1984) points out, these terms have been used either specifically to refer to a system implemented in the peripheral nervous system and musculature that produces a goal-directed action or else more generally and loosely to refer to a system implemented anywhere that explains systematic macroscopic patterning in behavior. Obviously, the first definition is the stronger one and it is the one intended here. It includes classically defined reflexes, but also, more interestingly, neuromuscular systems that are established transiently for a purpose.

Properties of synergies, according to Lee (1984), include activation of the same set of muscles (not always supraliminally) for execution of the same goal-directed action, constrained temporal sequencing or phasing of muscle actions in repeated performances of the goal-directed action and scaling of the properties of the synergy over changes, say, in rate or magnitude of the movements composing the action. I would add one more, and that is short-latency adaptive responses to perturbation, evident in muscle activity and movement.

All of these characteristics are evident in locomotion (for a review, see Grillner, 1981). However, at least some properties may be evident as well in less fundamental, less evolutionarily primitive, actions such as intentional arm movements (scaling: Kelso, Southard and Goodman, 1979; stereotyped muscle action: Cordo

[1]Kelso *et al.* (1984) did observe some extra activity of the orbicularis oris muscle following perturbation of closing during /z/ of /baez/. However, it was not accompanied by kinematic changes in lip activity.

and Nashner, 1982; perturbation: Bizzi and Polit, 1979) or finger movements in typing (Terzuolo and Viviani, 1979).

All have been observed or at least hinted at in speech as well. Muscle actions of the jaw during speech are stereotyped at least in the sense that they are evident in speech produced with a bite block – that is, with the jaw immobilized – as in unconstrained speech (Folkins and Zimmermann, 1981). Actions of the jaw, upper lip and lower lip also exhibit a strict relative temporal ordering in opening gestures (Gracco and Abbs, 1986). Finally, muscle activity (Tuller, Kelso and Harris, 1982) and articulatory movements (Ostry, Keller and Parush, 1983) scale systematically with changes in rate of speech or in stress, as if an invariant system were undergoing simple changes in biomechanical parameters such as equilibrium length or stiffness (Ostry, Keller and Parush, 1983).

The speech system under perturbation also has properties characteristic of synergies. They use lower-level parts flexibly to achieve invariant larger-scale ends in a changing environment and despite perturbation. If the findings from the perturbation research in the speech literature are properly interpreted as revealing use of synergies in speech production, then it is a good guess that synergies are widely used there. The perturbation research shows that constriction gestures for bilabial, labiodental and alveolar obstruents exhibit short-latency compensation for perturbation. It would be surprising if constrictions for other consonants were achieved in some other way. In addition, synergies can account for findings that fostered the predictive simulation account. Possibly, vocalic constrictions, no less than consonantal ones, are achieved by synergies of the vocal tract.

This conclusion is rather momentous in two respects. First, it highlights the fact that there is order in vocal tract behavior at the scale at least of gestural attributes of phonemes. Second, the order in vocal tract behavior corresponds very closely to gestural primitives of Browman and Goldstein's articulatory phonology. Conclusions that there is a fundamental mismatch between elements of linguistic competence and articulatory behavior have been premature. They are considerably weakened by nonlinear phonological theories that attend to the physical characteristics of phonetic segments, and by a more appropriate perspective on vocal tract activity that reveals its macroscopic order.

3.2.3 Saltzman's Task Dynamics Model: Intragestural Coordination

Saltzman and his colleagues (Saltzman, 1986; Kelso, Saltzman and Tuller, 1986; Saltzman and Kelso, 1987; Saltzman and Munhall, 1989) identify synergies in action with nonlinear dissipative dynamical systems more generally (see also Kugler, Kelso and Turvey, 1980; Kugler and Turvey, 1987). Saltzman and Munhall define synergies as 'task-specific and autonomous (time-invariant) dynamical systems that underlie a motion pattern's emergent form as well as its adaptive stability'. In the theory, and in the model of speech production that simulates speaking, the macroscopic invariants (e.g. tract variables such as constriction locations and degrees) of a phonetic gesture serve as 'point attractors' in the vocal tract that cause movement toward the attractor regardless of the current state of the vocal tract and despite perturbation. Dynamic systems are like mass-spring systems, in which a spring, when displaced, approaches its rest position regardless of the direction or extent of displacement.

In speech, movement takes place variably in articulators that compose a synergy for realizing values of those tract variables. Each relevant articulator's contribution to achievement of tract variable values depends on other demands on the articulator imposed by its participation in other synergies that are simultaneously at work in the vocal tract. Saltzman (1986; Saltzman and Kelso, 1987) has successfully modeled achievement of bilabial closure using point attractor dynamics. More importantly, he has modeled effects of an online perturbation that freezes the jaw during closure. Despite the perturbation, the model's 'lips' close without requiring any changes in planned dynamical parameter values. Possibly a weakness in the implementation of the task dynamics model to date is that its 'synergies' are not organizations in the vocal tract neuromusculature *per se* but in a model vocal tract whose actions are seen as driving the vocal tract. From my, perhaps naive, perspective, this may add an intermediary that might be unnecessary were the organization inherent in the peripheral nervous system and musculature – as it apparently is, in the bilabial closure synergy, uncovered by the perturbation experiments described earlier.

3.2.4 Coordination of Gestures for a Consonant or Vowel

Despite the promise of an account of speech production in terms of vocal tract synergies, there remain larger-scale problems of coordination that are not handled by the constriction-producing synergies so far considered. Not all consonants and vowels are exhaustively characterized by their constrictions. Some consonants are unvoiced, and have an associated devoicing gesture of the vocal folds; some phonetic segments are nasalized and have an associated lowering gesture of the velum; some vowels are 'rounded' and have an associated rounding gesture of the lips. There are subtler vocal tract actions as well that I will not consider.

The relationships among the gestures for a consonant or vowel are not well understood. In articulatory phonology, they are represented by association lines to a common 'root node' (see Figure 9.5), and it seems likely that the association is reflected in some way in the vocal tract. Otherwise there would be no way of guaranteeing that, for example, the devoicing gesture of the larynx is timed appropriately with respect to the constriction gesture for an unvoiced consonant. Possibly, synergies for a constriction gesture, and for any other gestures for the same phonetic segment, themselves are subsumed under a larger synergistic organization responsible for maintaining their coordinative relationship. The literature does not yet show this convincingly, however. I briefly review what is known about the relationships among gestures of a segment.

Unvoiced Consonants

Devoicing gestures of the larynx are achieved by abduction then adduction of the vocal folds. In many contexts (for example, generally syllable-initially in English), unvoiced consonants are breathy or aspirated. Producing aspiration requires that the vocal folds be open upon release of the consonantal constriction. Indeed, Löfqvist (1980) reports that the peak opening of the vocal folds occurs very close to release of a voiceless consonant. Furthermore, a linear relationship between onset of closure and peak glottal opening is maintained over variations in rate of speaking and stress that ensures the appropriate time relationship between release

and the devoicing gesture. That a systematic relationship is maintained over variation in stress and rate implies a coordinative relationship between the gestures. This is further suggested by a finding reported by Munhall, Löfqvist and Kelso (1986). Munhall *et al.* applied a perturbation to the lower lip just before oral closure for /p/. The perturbation delayed achievement of oral closure and also delayed the devoicing gesture correspondingly.

Nasalized Consonants

It has been reported (Moll and Daniloff, 1971) that lowering of the velum for a nasal consonant anticipates other gestures for the consonant to a variable extent that depends on the number of vowels preceding the consonant. This observation is taken as evidence for a particular theory of coarticulation known as *feature spreading* (Daniloff and Hammarberg, 1973; see also, Henke, 1966). In a feature-spreading theory, coarticulation is a phonetic or even phonological (Hammarberg, 1976, 1982) process whereby certain features of a segment spread to any neighboring segments that are unspecified for the feature. (While oral consonants are specified [−nasal], English vowels are unspecified for nasality because changing the value of the nasality feature of a vowel in a word never changes the word from one into another.)

A different theory of coarticulation (coarticulation as 'coproduction': Fowler, 1980; or, more specifically, *frame theory*: Bell-Berti and Harris, 1981) holds that coarticulation is not a phonological process at all;[2] instead, it reflects the way that segments are serially ordered in the vocal tract. There is, then, no coarticulatory process by which features of one segment spread to other segments (although, as noted earlier, individual languages may have phonological processes, such as vowel and nasal harmony, in which features spread). Instead, coarticulation is a process of temporally overlapping the articulation of neighboring segments in an utterance. In these theories, if the velum lowers during a vowel before a nasal consonant, it is not because the vowel has acquired a new featural attribute, but because production of the vowel and consonant overlap. According to frame theory (Bell-Berti and Harris, 1981), anticipatory lowering of the nasal consonant retains its affiliation to a nasal consonant and anticipates oral constriction for the consonant by an invariant time interval regardless of the number of vowels preceding the nasal consonant.

How can such a theory be entertained in the light of evidence such as that reported by Moll and Daniloff (1971)? In fact, the data from this and other research are not very clear. Bladon and Al-Bamerni (1982) cite five published reports favoring feature spreading, four favoring frame theory, one with an ambiguous outcome and their own findings which suggest to them that there are two coarticulatory styles that speakers choose between. More importantly, however, Bell-Berti (1980) shows that none of the research up to the time of her review is interpretable. It is well known that vowels are associated with different characteristic postures of the velum depending on vowel height (Moll, 1962) and with lower positions of the velum than occur during oral consonants. When researchers report

[2]This is not to say that the sound inventories of languages have no influence on coarticulatory extent and patterning (Boyce, 1988; Manuel, 1987; Manuel and Krakow, 1984). Languages may constrain the extent to which coarticulation can cause the acoustic consequences of producing a segment to resemble the acoustic product of some other segment of the inventory.

lowering of the velum just after the C in a CVN and a CVVN sequence (where C is an oral consonant, V a vowel and N a nasal consonant), and so an earlier onset of lowering in CVVN, they are almost certainly confusing the lowering gesture from C to the first V with onset of lowering for N. According to frame theory, when those confounding influences are eliminated, lowering of the velum for the nasal consonant begins an invariant interval before oral constriction.

Although I think that frame theory provides a more realistic perspective on the relationshipship between the two gestures for a nasal consonant than does feature spreading theory, its claim of temporal invariance between onsets of the two gestures is almost certainly too strong. It would be surprising if manipulation of rate of speaking and of stress were not to change the temporal interval between the gestures. However, if the interval varies systematically with other temporal intervals associated with nasal consonant production (as the timing of the devoicing gesture does in research by Löfqvist and Yoshioka, 1984), then frame theory's claim of an affiliation between the gestures of a nasal consonant would be supported. There is at least one other source of variability in the relationship between the gestures, however. Krakow (1988, 1989) shows that lowering of the velum begins earlier and reaches maximum lowering earlier for syllable-final than for syllable-initial nasal consonants. Within each syllable position, the timing of maximum velum lowering and the oral constriction gesture is quite stable, however.

Rounded vowels

The same controversy, between feature spreading and frame theory, occurs in the literature on lip rounding. There are reports that lip rounding anticipates rounded vowel onset to an extent that varies with the number of preceding consonants (Benguerel and Cowan, 1974; Sussman and Westbury, 1981), and that it anticipates vowel onset by a fixed interval (Bell-Berti and Harris, 1979). As for nasality, estimates of the anticipation of lip rounding have been contaminated, here by the occurrence of lip muscle activity (Gelfer, Bell-Berti and Harris, 1982) and movement (Boyce, 1988) for consonants preceding the rounded vowel. Gelfer *et al.* find that onset of lip muscle activity (orbicularis oris) anticipates a vowel increasingly the longer the string of consonants before the vowel, whether the vowel in question is or is not rounded! When consonant-related activity is eliminated from sequences in which rounded vowels are produced, the onset of rounding for the vowel is invariant over consonant strings of lengths greater than one.

Here, as in the research on anticipation of nasality for a nasalized consonant, frame theory appears to handle the data better than feature spreading theory. This conclusion is supported as well by findings of a 'trough' in lip rounding movements and in muscle activity during a consonant string between two rounded vowels (Bell-Berti and Harris, 1974; Boyce, 1988; Perkell, 1986). If the feature [+rounding] were to spread to any consonants before a rounded vowel, then there should be no trough.

If there is an invariant timing relationship between rounding and the constriction gesture for a rounded vowel, then there must be some coordinative relationship between the two gestures. Most likely, however, the interval between the gestures will prove not to be invariant over variation in rate of speaking and stress, and an important question will be whether it bears some systematic relationship to other intervals involving the two gestures over these manipulations.

Intergestural Sequencing in Task Dynamics

In Saltzman's model, gestures have associated 'activation coordinates'. These provide values for the influence over time that a gesture exerts on the vocal tract. The time at which a gesture's activation level becomes positive is determined by the gestural score of Browman and Goldstein's model. Saltzman and Munhall do not consider this solution to the sequencing problem satisfactory, and they suggest an alternative that I will consider shortly.

3.2.5 Coarticulation of Constriction Gestures

The foregoing review examined coordination of gestures for a given consonant or vowel. I turn now to sequencing of gestures for neighboring consonants and vowels in an utterance.

Although there is some disagreement as to when movements toward vowels begin in a V_1CV_2 sequence (compare Gay, 1977, and the investigators cited just below), in at least some circumstances, they begin early enough to affect the acoustic speech signal during the closing transition from V_1 to C (Öhman, 1966). In a V_1CdCV_2 sequence, the malleable /d/ vowel may be strongly influenced by both flanking Vs (e.g. Fowler, 1981a, b). Examination of tongue movements during V_1CV_2 sequences suggest that, when the C does not involve the tongue body anyway, the tongue may move smoothly from V_1 to V_2. (These movements may be reduced or blocked when the tongue body is used to produce the consonant; Recasens, 1984.) This has led some investigators to conclude that vowel production is, where possible, continuous during speech (Fowler, 1980; Öhman, 1966; Perkell, 1969).[3] If so, and if movements toward V_2's constriction gesture can occur very early, it may imply that consonants in a sequence bear the major responsibility for preserving the serial order of consonants and vowels in a planned sequence.

Spatial Overlap in Speech

Sometimes overlapping gestures make competing demands on the same articulators. One possible outcome is that multiple influences on a common articulator are wholly independent and they simply sum at the periphery. Some kinds of overlap do look like that. For example, the position for the jaw during /b/ closure is lower if a coarticulating vowel is open than if it is closed (cited in Keating, 1990). Similarly, raising of the velum continues throughout a string of oral obstruents so that the position of the velum is higher during the second /t/ in a /ts#st/ sequence than in a /t#t/ sequence (Bell-Berti, 1980). Finally, Boyce (1988) found additive effects of rounding for vowels and lip movements associated with conson-

[3]Here is another example, possibly, of the phenomenon of 'triggering' discussed by MacNeilage and Ladefoged (1976), in relation to the findings of near-universal differences in vowel duration before voiced and voiceless consonants and of phonologization of the difference in a few languages. Many languages, perhaps all, show vowel-to-vowel coarticulation (see references in the text). However, a few languages (including, for example, Turkish, Hungarian and Yawelmani – an American Indian language of California) have phonological processes known as 'vowel harmony' in which, roughly, the vowels of a word are required to share a phonetic feature(s). For example, in Yawelmani a vowel is rounded ([+back]) if preceded by a vowel in the same word that is rounded and matched in height (Kenstowicz and Kisseberth, 1979).

ants. In fact, however, independence of influences is probably not generally realistic, as Saltzman and Munhall (1989) note. For example, while the jaw is susceptible to vocalic influences during /b/, it is not, or is less so, during /s/ (Keating, 1990). Accordingly, there must be a way to suppress vocalic influences in that context – in the task dynamics model, this is accomplished by 'blending rules' that have gesture-specific and phoneme-specific parameter values. There is, as yet, no systematic investigation of varieties of 'blending' in natural speech.

Serial Ordering of Consonants and Vowels

How does a talker know when to start producing the gestures of a segment? Are initiations timed as if by a clock (cf. 'comb' models of sequencing; Bernstein, 1967; Kozhevnikov and Chistovich, 1965) or are they triggered by information from the periphery signaling that it is time (cf. 'chain' models), or is there some third way? Each extreme model has deficiencies. In the speech literature, one recent proposal is a variant of a chain model.

Kelso and Tuller (1987) suggested that sequencing might depend on a phasing rule. For example, if a vowel's gesture is defined as spanning a 360-degree cycle, then a following consonant might be initiated at some fixed phase angle of the cycle (say after 200 degrees). Initial findings were supportive of the view; however, more recent findings (Nittrouer *et al.*, 1988) are not. Accordingly, the details of sequencing in speech are still not understood.

In Dell's (1986) model, recall that sequencing is achieved by giving a 'current node' extra activation as compared to activation assigned to later elements in a planned sequence. The lowest level of the language at which his model applies this sequencing is that of individual phonemes of a word. Therefore, the model will not give rise to temporally staggered gestures of a phoneme or to coarticulation more generally. With Figure 9.4 above substituted for Dell's lowest level, syllabic frame, gestural sequencing might be achieved by means of the gestural score. Indeed, as briefly noted earlier, that is how gestural sequencing is achieved currently in the task dynamics model.

However, Saltzman and Munhall (1989) do not find that solution satisfying, apparently because, whereas intragestural coordination is a product of the dynamics of the speech system, intergestural sequencing by this account is not. They intend to incorporate a version of Jordan's (1986) model for serial dynamics into the task dynamics model to replace the gestural score.

Briefly, in Jordan's model, learned sequences are trajectories through a 'state space' that, over learning, become 'attractors'. In contrast to the point attractors that cause realization of constrictions for consonants and vowels, however, these are trajectories that attract other less well-learned trajectories to it. A trajectory through a state space, in turn, determines successive values for output units in the model, and these values determine actions of the system. In task dynamics, the succession of positive activation values for gestures will be determined by the succession of positive values for different output units.

This kind of system allows learning of arbitrary orderings of a small inventory of elements and, to a large extent, that is exactly the task for a child acquiring a lexicon of words. In the model, learning a planned sequence occurs as an output sequence is compared to a 'teacher' sequence. Errors in the form of discrepancies between the output and teacher sequences are propagated back through the system and are used to change the weights on linkages between units in the network.

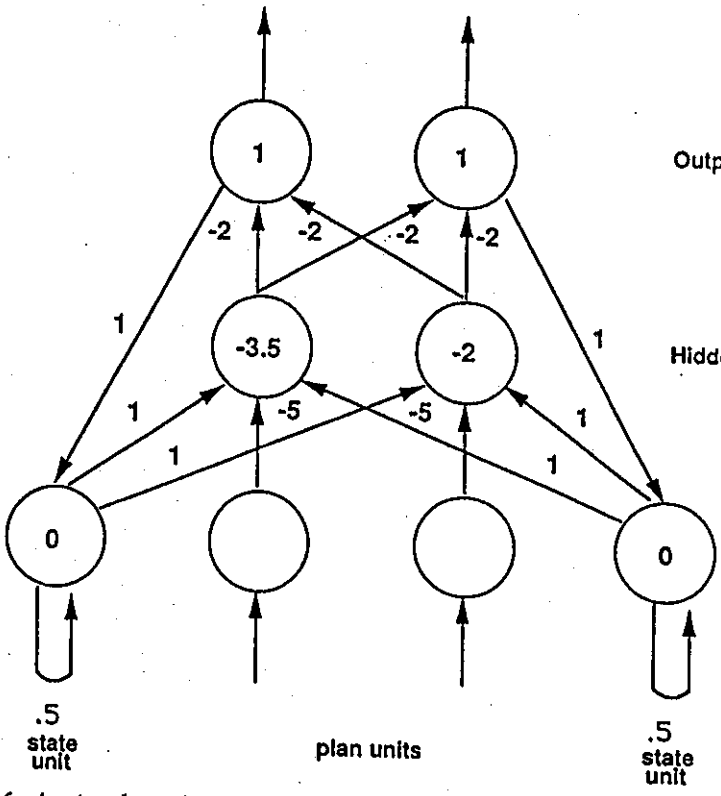


Figure 9.6. *A network consisting of two plan units, two state units, two hidden units and two output units. With the biases (value printed in the circles) and weights (values on the linkages) shown, and with plan units [1,0], the network will produce a sequence AAAB, where A is output [1,1] and B is [0,0]. Outputs are binary, taking on a value 1 if the net input to the unit is positive and 0 if it is negative. If the plan is [0,1], the network produces AB. [From Jordan, 1986.]*

Eventually, in the presence of a plan to perform a sequence, the output sequence matches that of the teacher. (See Figure 9.6 for a sample network that has learned to produce two different sequences in the presence of different plans.)

A remarkable property of Jordan's model is that ordered consonants and vowels of a learned sequence coarticulate without being explicitly instructed to. Coarticulation develops during the learning process as the model learns to match certain specified output values of the teacher. Output values for the segments of a word are determined by a plan for the word (a different one for each word, in which serial ordering of elements is not specified explicitly) and by changing values of state units in the system. Plan units and state units determine values of 'hidden units', which in turn determine values of output units (Figure 9.6). State values change over time, and their next value in a trajectory is determined by their previous value and by the just-computed output values. Therefore, successive values of state units are similar and they produce similar outputs. Accordingly, as learning proceeds, coarticulatory effects of an output tend to spread bidirectionally.

3.3 Concluding Remarks

A major aim of this section was to show that the phonologies of languages may be conceptualized in such a way that the fundamental units of language do not have characteristics that vocal tracts are physically incapable of realizing, and that vocal tract activity can be shown to exhibit macroscopic patterning at a scale commensurate with that of the fundamental units of language. Neither adjustment in point of view threatens the close correspondence uncovered in Section 2 between units of the language and units of a speech plan.

Some researchers have begun to believe that organizations for intentional activity are achieved in ways not unlike self-organizing processes in other dynamic systems. Possibly, then, synergies are examples of dynamic systems as Saltzman and his colleagues suppose. A major unresolved problem is still (cf. Lashley, 1951) to understand how speakers achieve largely accurate ordering of linguistic elements in speech. Jordan's approach to the problem appears quite successful and, according to Saltzman and Munhall, largely compatible with their own dynamic approach to modeling speech production more generally. It remains to be determined to what extent both of these models realistically reflect speaker's manners of implementing synergies and of ordering them sequentially.

There remains a different kind of mismatch between 'competence' and 'performance' that may, in the minds of many investigators, render an effort to equate linguistic units and produced units irrelevant. It is a view that phonological segments are inherently mental categories, where 'mental' means products of and residing in the mind of an intelligent organism.

This point of view is represented in the literature (Hammarberg, 1976, 1982; Repp, 1981). If it has merit, then any attempt to equate units of the language with physical activity of the vocal tract is a sort of category error. In my opinion, however, the view is fundamentally in error.

The claim is that phonological segments as known are mental things and those as uttered are physical things, and so they are things of fundamentally different kinds. This is not quite right, however. If there are phonological categories in the mind, they must have some physical (e.g. neurological) instantiation in the brain, and so they are physical things as well as mental ones. More importantly, however, public activities that communicate linguistic messages to listeners may also be seen as mental (psychological, intelligent) things as well as physical ones. As Ryle points out (1949), intelligent actions have two aspects: they are physical actions, but they are also intelligent. Accordingly:

> 'When a person talks sense aloud, ties knots, feints or sculpts, the actions which we witness are themselves the things which he is intelligently doing, though the concepts in terms of which the physicist or physiologist would describe his actions do not exhaust those which would be used by his pupils or his teachers in appraising their logic, style or technique. He is bodily active and he is mentally active, but he is not being synchronously active in two different "places" or with two different "engines". There is one activity, but it is one susceptible of and requiring of more than one explanatory description.' (p. 51).

A different argument also suggests that linguistic utterances as known and performed are not different kinds of things. When theorists conclude that speech

activity fails to achieve linguistic units, they are taking the view that units of competence are primary and units of performance are derived. That is, what we do when we speak is a pale, not-entirely-true, reflection of what we know about the language and of what we plan to say. But an alternative view is that the relationship is reversed, both in the ontogeny of the individual and in its mature form. We know what units are autonomous in the language because they are used autonomously by native speakers when they talk. That is, competence (what we know) is derived from what we do and what we experience other members of our language community doing. (By analogy, there are such things in the world as chairs that we may come to know by experience. If it is true that we come to have a category corresponding to 'chairs' in memory, it is because there is such a category in the world that we have come to recognize. Chairs in the world are primary; what we know of them is derived.) Because what we know of language is derived from what we do and what we experience other members of a language community doing – indeed, because competence is knowledge of the essentials of language performance – the elements in each domain can, it seems, be the same kinds of things (but see, for example, Chomsky, 1986, for a wholly different point of view).

4 Prosodic Structure in Speech

The acoustic speech signal exhibits considerably more systematicity than the vocal tract activities considered so far will generate. In the domain of fundamental frequency (f_0), talkers produce utterances with an intonational melody. In addition, in at least some styles of speech, f_0 falls gradually throughout a sentence or phrase. Further, each content word of the language has at least one stressed syllable – that is, a syllable apparently produced with greater respiratory and articulatory effort (Lehiste, 1970) than other syllables so that it is longer, more intense, its vowel spectrally less centralized and its f_0 increased over an unstressed syllable consisting of the same phonological segments. In addition, syllables or groups of stressed and unstressed syllables (stress feet) are supposed to be produced rhythmically so that the one constituent or the other (that is, the syllable or the stress foot) is approximately isochronous in syllable- and stress-timed languages, respectively. Whether or not there is such a tendency is controversial (see, for example, Allen, 1975; Classe, 1939; Dauer, 1983; Ohala, 1970). However, it is the case, at least, that in many languages a syllable's vowel is shortened increasingly as consonants are added to the syllable (Fowler, 1983; Lindblom and Rapp, 1973) and a stressed vowel is shortened as unstressed syllables are added to the foot (Fowler, 1981a). Finally, apart from these shortenings (Rakerd, Sennett and Fowler, 1987), languages show lengthenings ('final lengthening'; Klatt, 1976) and pauses at the ends of phrasal constituents that correlate at least approximately with the syntactic depth of the boundary (Cooper and Paccia-Cooper, 1980). Oddly, however, the units that lengthenings and pauses delimit are only approximately syntactic (Gee and Grosjean, 1983).

As Fudge (1969) suggests, there are two kinds of hierarchical constituency in speech. There is the morphosyntactic grouping of phonemes into morphemes, morphemes into words and words into syntactic phrases. In addition, there is a

phonological or prosodic hierarchy of phonemes into syllables, syllables into stress feet and stress feet into larger phrasal groupings.[4] Interestingly, the units of the first hierarchy misorder in speech errors while the units of the latter do not.[5] Nor, for the most part, are the prosodic groupings indicated in writing systems, while morphosyntactic ones often are. Gee and Grosjean's characterization of the prosodic organization as 'performance structures' is probably apt.

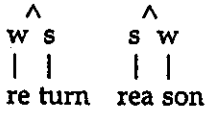
Even though prosodic structures may emerge uniquely in spoken utterances, they are not necessarily automatic consequences of vocal tracts producing speech. For example, intonation contours vary across utterances, and certain 'tunes' are associated with certain speaker attitudes or intended meanings.

In linguistics, metrical phonologies characterize the prosodic patterning best. I will briefly review this kind of phonological theory first as it characterizes stress and then as it has been applied to intonation.

4.1 Metrical Phonology[6]

4.1.1 Stress

The two syllables of any disyllabic word differ phenomenally in prominence or degree of stress. In the word 'return' the second syllable is more prominent than the first; in 'reason' the first is the more prominent. In metrical phonology, that relationship of prominence is expressed as follows (Liberman and Prince, 1977):



where s marks the stronger of the two 'tree branches' and the w the weaker. In longer words, too, one syllable stands out as most prominent or strongest; of the remaining syllables, some may appear very weak ('reduced') while others may be intermediate in prominence between the strongest and the weakest. For example, the third syllable in 'anecdotal' is strongest, the first is next strongest while the

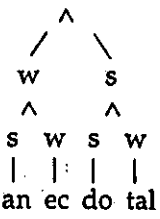
---

[4]In MacKay's (1987) recent theory of speech production, the phonological hierarchy is the bottom part of a single hierarchical strucure that includes the morphological hierarchy as the top part. However, for two reasons, this is an error. First, consider the interface between the hierarchies. In MacKay's model, morphemes are superordinate to syllables. However, while there are some morphemes that can be said to be composed of one or more syllables (e.g. 'dog', 'carpet'), there are also syllables that are composed of morphemes ('grew', 'walks'). Second, and analogously, there are phonological groupings (stress feet, phonological and intonational phrases) larger than the syllable that are not necessarily coextensive with larger morphosyntactic units.

[5]This is not to say that talkers make no suprasegmental errors. They may impose phrasal stress on the wrong word or produce an inappropriate intonation contour (Cutler, 1980). It is only to say that the metrical units themselves do not misorder.

[6]The following is a composite of perspectives offered by Hayes (1982), Liberman and Prince (1977), Selkirk (1980a, b) and van der Hulst and Smith (1982).

other syllables are weak. Its metrical structure is as follows:



Such a tree structure suggests that there are at least three groupings of syllables: one in which pairs of syllables contrast in relative prominence, one in which syllables are grouped into larger constituents (stress feet) which themselves differ in prominence, and one in which feet are grouped into words.

Above the word, analogous rules can assign relative prominence to words of a sentence (see, for example, Selkirk, 1980a, b). In Selkirk's account (but see Selkirk, 1984), there are two hierarchical levels above the word: the 'phonological phrase' and the 'intonational phrase'. In the sentence: 'The absent-minded professor has been avidly reading about the latest biography of Marcel Proust,' there are four phonological phrases (with right edges after 'professor', 'reading', 'biography' and 'Proust'). These are not all syntactic groupings ('has been avidly reading' is not), but they do tend to group words into a structure of which the final word is the 'head' of a syntactic phrase. (The head of a noun phrase is a noun, that of a verb phrase is a verb, etc.) Intonational phrases (the domain of an intonational tune) are groupings of phonological phrases. Selkirk speculates that these groupings may be selected fairly freely by the speaker. Both phonological phrases and intonational phrases assign s to right branches of their trees. This is compatible with a general tendency for speakers to place newer and more foreground information later in a sentence (Gee and Grosjean, 1983); by making right branches prominent, newer or otherwise foreground information will be highlighted by being assigned pitch accents, for example.

Whereas at lower levels of the tree structure, prominence is realized as stress accent, at the higher levels it is realized by accents of an intonation contour (Beckman, 1985).

4.1.2 Intonation

Intonation contours are continuous, during voiced parts of an utterance. However, in the phonological accounts under consideration (Liberman, 1975, Pierrehumbert, 1980), intonational tunes are composed of sequences of discrete tones (low, middle and high in Liberman, 1975; low and high in Pierrehumbert, 1980, and in most subsequent accounts). Pitch accents composed of tones are aligned to prominent syllables; the intonation contour is interpolated and smoothed from prominent syllable to prominent syllable.

Languages may have characteristic tunes – that is, particular sequences of pitch accents and boundary tones that express a speaker's attitude or intended meaning. Possible tunes are generated by rule.

4.2 Why is there a Prosodic Organization Distinct from a Syntactic One?

Stress accents on words ensure that content words receive some measure of prominence for the listener. (Function words tend to be de-stressed.) Larger groupings that elevate the prominence of selected words allow particular words, and so particular parts of the communicative message, to be highlighted. Which words of an utterance are to be highlighted will have to do not only with the syntactic structure of the sentence, but also with the speaker's particular intent in uttering the sentence. The prosodic structure allows a measure of freedom in the talker's performance to augment what can be conveyed by the words in their syntactic groupings.

4.3 Speech Production Once Again: Prosodic Structures

In the final several sections, I will review evidence for four prosodic constituencies in speech production: syllables, stress feet, phonological phrases and intonational phrases. Questions are whether they manifest themselves at all in speech performance and if so how, what kinds of constituents they are and, in cases where proposals have been offered, how they make their way into speech performance.

4.3.1 Syllables

Syllables may have their foundations in the jaw cycle. The jaw is a major articulator for speech that cycles open and closed. Canonically, syllables are closing–opening cycles of the jaw. However, even if that is the ultimate source of syllables in speech, in languages they are more than closing and opening gestures of the jaw. Languages differ in the syllable structures they permit; all languages allow CV syllables, but many disallow complex structures such as the CCCVCC structure of the word 'strength' in English. Even so, the jaw cycle does manifest itself cross-linguistically in a universal tendency for syllables to respect a 'sonority hierarchy' such that consonants in a syllable order themselves so that they increase in vowel-likeness the closer they are to the vowel. (For example, the order of /t/ and /r/ is /tr/ before the vowel and /rt/ after it.) A consequence is that the jaw opens smoothly toward the vowel, rarely reversing direction, and smoothly closes after the vowel (Keating, 1983).

Syllables require mention in theories of phonology for several reasons. In some languages with phonological length distinctions (so that V:, a long vowel otherwise identical to V, counts as a distinct vowel from V, and C: is distinct from C), syllable structure constraints require that a syllable containing V: must be followed by C, not by C: while a syllable containing V is followed by C: (see, for example, Lehiste, 1970). In many instances, syllable structure constrains the application of a phonological process in the language. For example, according to Kahn (1976), in so-called r-less dialects of English, /r/s are dropped unless they are syllable-initial. Therefore, /r/s in 'car' and 'carpet' are dropped, while those in 'carry' and 'rack' are not.

Likewise, phonological theories require that syllables be assigned an internal structure. In 'quantity-sensitive' languages, syllable weight determines whether a

syllable can be metrically strong; weight is determined in most quantity-sensitive languages by the number of consonants in the syllable 'rhyme' (the vowel and final consonants) or by the syllable nucleus (the vowel alone). Rarely is the syllable onset (the prevocalic consonants) relevant to a determination of metrical strength (but see Davis, 1988). For these languages, anyway, this suggests a hierarchical structure of the syllable into onset and rhyme, and of the rhyme into nucleus and final consonants ('coda'). This structure is supported on other grounds as well, such as the relative strength of sequencing restrictions on phonemes within and between these syllable constituents (Fudge, 1969).

I am unaware of any convincing evidence that syllables or syllable constituents are autonomous performance units, however. Although they may constrain the movement of phonemes in errors as described earlier (that is, consonants that misorder in speech errors tend to maintain their position relative to the vowel), they do not misorder in speech errors as phonemes and words do, for example.

Another place in which syllables appear in descriptions of systematic speech behavior is in descriptions of systematic influences on vowel duration. Lindblom, Lyberg and Holmgren (1981) conclude that, in Swedish, syllable structure affects vowel duration in that a vowel shortens progressively as consonants are added to the vowel either before or after it (see also Fowler, 1983, for English). As Lindblom *et al.* point out, this may be seen as a compensatory shortening effect as if the talker were attempting to maintain a constant syllable duration regardless of the number of segments in the syllable. However, as they also point out, as an attempt to maintain syllable isochrony, it is feeble. Shortening of the vowel in the presence of neighboring consonants is far smaller than the duration of the consonants themselves. Syllable isochrony is not maintained. In any case, there is reason to doubt that this is an effect of syllable structure. In the data presented by Lindblom *et al.*, vowels shorten even if the added consonants are in a syllable adjacent to it. Hence the shortening may not index syllable structure at all.

I have suggested that it is an acoustic-durational manifestation of coarticulatory overlap of the vowel by the neighboring consonants (Fowler, 1983). Consistent with that interpretation, Fowler *et al.* (1986) found no evidence of *articulatory* shortening of opening for a vowel as consonants are added to its syllable rhyme; jaw lowering for the vowel has about the same duration and extent regardless of the size of the coda. Instead, Fowler *et al.* found an earlier onset of jaw raising for the consonants in the coda.

Treiman (1983, 1984, 1986) has found considerable evidence that speakers are sensitive to syllable structure. In her research, subjects learn word games that require them to split off parts of syllables and recombine them. If the game requires them to split off the onsets of syllable pairs from their rhymes and to create new syllables consisting of the onset of one syllable attached to the rhyme of the other, they learn the game faster and with fewer errors than if a different parsing is required that violates syllable constituents. For syllable-final, but not syllable-initial, consonants, the difficulty of games that violate their constituency varies with the sonority of the consonants. That is, syllable-final games requiring either V/CC or VC/C partitionings were about equally difficult if the postvocalic consonant was a nasal, intermediate in sonority. Games requiring a V/CC partitioning were considerably easier than those requiring VC/C splits if the postvocalic consonant was an obstruent (a stop or a fricative). Finally, VC/C games were actually easier than V/CC games if the postvocalic consonants were the highly sonorous liquids, /l/ or

/r/. It is not yet evident, however, what makes consonants in the coda more cohesive with the vowel than consonants in the onset, or what makes more sonorous consonants – but possibly only those in the coda – more cohesive with the vowel than less sonorous ones.

In short, although there are hints that the syllable has some reality in speech planning and performance – as a structure that constrains the movement of consonants in speech errors and as a structure otherwise affecting the cohesiveness of consonants with the vowel – its role in speech production is far from clear.

4.3.2 Stress Feet

Syllables may be built on one kind of cycle, the closing and opening of the jaw, while stress feet realize another kind of rhythmicity – an approximate alternation of stressed and unstressed syllables. However, the stress rhythm, if there is one, has a less obvious foundation in dispositions of the vocal tract than the syllabic rhythm.

The stress foot manifests itself in speech in two ways: as an inclination on the part of speakers to alternate stressed and unstressed syllables and as a tendency to shorten stressed syllables when a foot is more than monosyllabic (or else perhaps a tendency to lengthen stressed syllables when a foot is monosyllabic; e.g. Bolinger, 1963, 1981). In fact, these may be joint reflections of a common stress-alternation tendency.

As for the tendency to alternate stressed and unstressed syllables, Kelly and Bock (1988) find that nonsense words with two stressable syllables are likely to be pronounced with trochaic stress (strong–weak) rather than iambic stress (weak–strong) when they are preceded by an unstressed syllable and followed by a stressed one; the tendency to use a trochaic rhythm is significantly reduced (but not reversed) when the preceding stress context is weak and the following one is strong.

Compatibly, disyllabic verbs have been found to be more likely (in spontaneous speech) to be followed by syllabic inflections (that is, inflections such as -ing for verbs or -es, pronounced /Iz/, for nouns) than disyllabic nouns (Kelly, 1988). Moreover, in an experiment in which disyllabic nonwords occurred in contexts where they served as verbs, they were more often produced with iambic stress if they ended in /d/ and took an -ed suffix (pronounced /Id/) than if they were identical except that they ended in a consonant after which -ed is pronounced /d/or /t/.

A different manifestation of the stress foot in speech behavior is the occurrence of stressed syllable shortening when feet are more than monosyllabic. This occurs for words spoken in isolation (Lehiste, 1972) as well as words in context, among at least speakers of Swedish (Lindblom and Rapp, 1973), English (Fowler, 1981a) and Dutch (Nooteboom, 1973). The same pattern is considerably weaker in Italian (Vayra, Avesani and Fowler, 1984) – ostensibly a 'syllable-timed' language, rather than a stress-timed language (even though at least some investigators consider Italian to have a left-dominant stress foot structure; Nespor and Vogel, 1979). In Swedish and English, which have left-dominant feet, comparable shortening of a stressed syllable is not observed when an unstressed neighbor is in a different stress foot – that is, when they precede the stressed syllable (Fowler, 1981a; Lindblom and Rapp, 1973). In English, in at least some contexts, the shortening is correlated with

coarticulatory overlap (Fowler, 1981a); to the extent that an unstressed syllable coarticulates with a stressed syllable, the stressed syllable is shortened by the unstressed syllable. Coarticulation and shortening may be independent indices of the greater cohesion between a stressed syllable and the immediately following unstressed syllable than between a stressed syllable and a preceding unstressed syllable or a final unstressed syllable in a trisyllabic foot. Alternatively, as I suggested above for shortening of vowels by consonants, the acoustic shortening may be a manifestation of coarticulatory overlap.

The shortening of stressed by unstressed syllables has been seen as an indication that talkers may be attempting to maintain isochrony of stress feet ('stress-timing'; for reviews, see Dauer, 1983; Fowler, 1977). However, as many researchers have pointed out, the small amount of shortening of the vowel does not come close to compensating for duration of an unstressed syllable added to a foot (e.g. the duration of 'speedy' is longer than that of the word 'speed' even though the first syllable in 'speedy' is shorter than that in 'speed').

Edwards and Beckman (1988) looked at these findings differently: they considered that the monosyllabic foot lengthened, rather than that the disyllabic foot shortened. They found that the jaw opening phase of a stressed syllable in a monosyllabic (as compared to a disyllabic) foot is lengthened relatively less than the closing phase so that the point of maximum opening of the jaw (and so the prominence peak for the vowel of a prominent syllable) occurs relatively earlier in the monosyllabic foot. They ascribed the lengthening to the presence of a stress clash. That the lengthening accomplishes a relative backward shift of the prominence peak of the vowel suggests that lengthening is a way of shifting the prominence peak of a stressed syllable back away from that of a following stressed syllable and alleviating the stress clash that way rather than by shifting stress off the first syllable altogether.

The foregoing review reveals at least a tendency to alternate strong and weak syllables and in addition, perhaps, a stronger degree of cohesion, in two languages with left-dominant feet, between stressed syllables and following unstressed ones than between stressed syllables and preceding unstressed ones. It does not reveal a reason for the alternations to occur, nor a function for the stress foot.

An entirely different perspective on the stress foot is provided by the work of Sternberg, Monsell and their colleagues (Monsell, 1986; Sternberg *et al.*, 1978, 1980). These investigators designed an experimental procedure to study 'motor programming' in speech and typing. In the procedure as applied to speech, subjects were given a list of words to say as rapidly as they could. However, there was a considerable (several seconds) delay between list presentation and a signal to begin producing the list. The subjects' latencies to begin the utterance and their utterance durations provided the main measures in the experiments.

Because talkers know in advance what they will be saying, their latencies do not reflect a choice between utterance alternatives; the task measures 'simple reaction time'. Moreover, to the extent that it is possible to construct a speech plan in advance of its execution, subjects have the information and the time needed to construct one before the response signal is presented. Even so, the subjects' latencies to begin the utterance vary systematically with the number of things to say: a list of five-digit names is initiated later than a list of three-digit names, for example. Remarkably, the function relating latency to utterance length defined by the number of things to say is linear with a slope that is stable over different

utterance compositions (e.g. digit names, days of the week). However, the function is linear with a stable slope only if the number of things to say is counted in stress feet, not in syllables, words or phrases (Monsell, 1986; Sternberg *et al.*, 1978, 1980). The linearity of the relationship means that each addition of a stress foot to a list of things to say adds a constituent amount of time (around 12 ms) to response latency.

Another remarkable finding of this research that has shaped its interpretation by Sternberg and his colleagues concerns the function relating total utterance duration to number of stress feet in the utterance. The duration function is not a straight line; rather, it is a quadratic function (that is, of the form $y = ax^2 + bx + c$, where y is utterance duration and x is the number of stress feet in the utterance). That the function is a quadratic means that the slope of the function increases by a constant amount as more stress feet are added to the utterance. That is, the more there is to say, the more slowly the components of the utterance are produced. Interestingly, the increase in slope (the coefficient of the x^2 term) is the same as the slope of the latency function (about 12 ms).

How can the collection of findings be explained? An intuitive idea is that the heavier the load on the system, the slower it works. One elaboration of this idea has been tested and ruled out. The elaboration is that the load in question is a 'processing' load in a system with a limited 'processing capacity'. That interpretation was ruled out in an experiment by Sternberg *et al.* (1978), in which talkers were given a list of digits to remember while producing another utterance as rapidly as possible after a response signal. The digit load was meant to require processing capacity and hence to increase latency and utterance duration if limited processing capacity were behind the latency and duration functions. However, the digit load had essentially no effect on the latency and duration functions even though subjects did well recalling the digits.

Sternberg *et al.* offer the following interpretation of their findings. The stress foot is the unit into which consonants and vowels of the words in the utterance list are packaged for execution. Subjects construct a motor program or plan in advance of the response signal consisting of the words of the utterance packaged into stress feet. When the response signal is presented to initiate the utterance, they must *retrieve* the first stress foot, *unpack* it into its parts and execute a *command* process to initiate vocal tract activity. To produce the whole utterance, they successively retrieve, unpack and command production of successive stress feet in the utterance. The retrieval process is sensitive to the number, but not the size, of stress feet, while the other two processes are sensitive to the size of each stress foot, but not the number of stress feet in the utterance. This can explain why, for example, the slope of the latency function is sensitive only to the number of stress feet, but not to their compositions.

The retrieval phase of plan execution is identified as a serial search through a buffer consisting of the stress feet to be produced. Because latency to speak increases with the number of things to be said, it must be supposed that the search does not proceed, say, left to right in a buffer in which stress feet are arrayed in the order, left to right, in which they will be said. Either the search order or the order of items in the buffer, or both, must be different from the to-be-uttered order. Morever, that the slope of the latency function has the same value as the coefficient of the squared term of the duration function implies that the buffer does not shrink as the utterance proceeds. If the talker has produced three of five stress feet in a

sequence, to find the fourth the retrieval mechanism must still search through a buffer containing five stress feet. Otherwise, the latency to produce successive items in the utterance would shrink and overall the coefficient of the squared term of the duration function would be smaller than that of the latency function.

The latency to begin talking will also be affected by the time to unpack the first item in the sequence. The duration of the first stress foot will be determined by the duration of its command process (the more vocal tract gestures to be initiated, the longer the stress foot). The interword interval will be affected by the retrieval time for the next item and its unpacking time. In fact, those variables affect the entire interval consisting of the duration of the second syllable of a disyllabic stress foot and the interword interval. Accordingly, Sternberg *et al.* propose that the final part of a stress foot is allowed to continue during retrieval and unpacking of the next stress foot.

The model involves stages that, intuitively, are parts of motor planning and execution, and it accounts extraordinarily well for the data. However, at least one aspect of the model – its retrieval mechanism – is implausible. Why should the retrieval mechanism search elements in an order different from their to-be-produced order? Why would a talker not order items to be produced in their planned serial order? Dell's (1986) model offers a possible answer to these questions. In that model, the 'order' of to-be-produced items is their location in the lexicon. To signal their planned order in an utterance, they are assigned order tags. It is not difficult to imagine a retrieval mechanism more or less as described by Sternberg *et al.* searching among the order tags to find the appropriate unit to output. However, there are several difficulties with this attempted merger of Dell's lexicon and the motor plan of Sternberg *et al.* One is that sequences such as 'Monday Monday Monday Monday' and its subsets yield the same latency and duration functions as sequences such as 'Monday Friday Wednesday Tuesday' and its subsets. Yet in Dell's model, all the order tags for the 'Monday' sequences would be on the same word node. A second difficulty is that there is an incompatibility of units. Dell's units must be morphosyntactic while those of Sternberg *et al.* have been shown not to be. A third difficulty is that stress feet do not misorder in speech, but a retrieval mechanism of the sort proposed by Sternberg *et al.* seems unlikely to be infallible.

Rosenbaum, Kenny and Derr (1983; see also Rosenbaum, 1985) point out that the qualitative outcome reported by Sternberg *et al.* can be captured in an entirely different way than Sternberg *et al.* propose if elements in the to-be-uttered sequence are ordered and are hierarchically organized into a binary branching tree. In the model proposed by Rosenbaum *et al.*, executing the first element in the sequence to be uttered requires that a pointer traverses a binary branching hierarchy from the top node to the leftmost terminal element. The more nodes in the tree that must be traversed, the longer the traversal time and so the longer the latency to output the first item in the string. In turn, the longer the string, the more nodes in the tree; accordingly, latency will correlate with string length. Outputting the next element after the first requires that the pointer move upward from the leftmost terminal element to the node from which that element and the next one branch and then move down from there to the terminal for the second element. In general, outputting any next element involves moving the pointer from one terminal element to the next by traveling along the tree branches that connect them. The more elements in the string, the more branches and nodes in the tree and so the longer, on average, it will take to get from terminal element to terminal element.

Accordingly, execution of the string will slow with string length. In addition, the model predicts differences within the string in interresponse time, with short times between elements 1 and 2 and between 3 and 4, for example, and the longest times between elements that bisect the string. To my knowledge, these predictions are untested.

This model has an advantage over that of Sternberg *et al.* in not having to suppose that elements to be produced in a given order are unordered from the perspective of the retrieval mechanism. It has the same disadvantages, however, of failing to rationalize the units that make up terminal elements of the tree and of lacking face plausibility. Why are string elements stress feet rather than morpho-syntactic units? Why shift from the syntactic tree (which need not be symmetrical and binary branching) to a binary branching one? What does the tree accomplish other than to slow down the output process (over a process that simply reads out elements of the string left to right)?

Having expressed some skepticism with the way that Sternberg *et al.* explain their data, and over the alternative account of Rosenbaum *et al.*, I have to confess that I do not know a better way. A place to look for an account, however, may be in the direction of the capacity account that Sternberg *et al.* tested and rejected. They rejected an idea that the latency and duration functions might be caused by limitations on central processing capacity by showing that extra demands on processing capacity (a memory load) did not affect response functions. Alternatively, however, perhaps the limitations are downstream of any pool of central processing capacity. Perhaps the limits are on a general pool of energy resources available to produce an utterance. As a first approximation, imagine that an inspiration makes available a pool of resources for producing an utterance on a single breath group. The more there is to say, the more limited the resources available for each unit to be uttered. Effects of unstressed syllables are not noticed because they require negligible expenditures from the resource pool. Reductions in resources affect time to initiate production of stressed syllables. If demands on the respiratory system were an important factor, then manipulating the sizes of inspirations or composing utterances of phonetic segments that deplete the air supply rapidly or slowly (e.g. /f/ versus /m/; see Gelfer, Harris and Baer, 1987) should affect the latency and duration functions where manipulations of memory load would not.

4.3.3 Phonological and Intonational Phrases

Phrasings above the foot are signaled in several ways. Three related ones are a tendency to lengthen syllables at a phrase boundary, a tendency to mark the boundary with a pause, and a tendency for cross-word phonological processes (such as palatalization as in 'did you' becoming 'didja') to be blocked. In addition, intonational phrases are the domains of a coherent intonational melody or tune and of a gradual decrease in fundamental frequency known as 'declination'.

Lengthenings, Pausing and Blocking of Cross-Word Effects

Even when speakers are reading, and so need not decide what to say next, they distribute pauses or other indices of slowing and braking unevenly in their speech utterances. The pauses are not randomly distributed, however.

Cooper and Paccia-Cooper (1980), in an extensive series of experiments, examined the distribution of pauses, lengthening and blocking of cross-word phonological effects in a variety of sentences that were read aloud. They manipulated the syntactic structure of otherwise similar sentences and found a close relationship between surface syntactic structure and the distribution of pausing, lengthening and blocking. In general, the more important the syntactic boundary, the longer the pause, the greater the lengthening of syllables on the 'left' side of the boundary and the greater the likelihood of blocking a cross-word phonological effect. They found no evidence that these three variables patterned differently. Presumably, all are indices of a slowing that serves to break an utterance into phrases.

To predict relative pause duration, amount of lengthening or probability of blocking at each word boundary in a sentence, Cooper and Paccia-Cooper proposed a complicated 14-step algorithm applied to each word boundary in the sentence. They recognized that the algorithm was not a realistic candidate for talkers to use to generate pausing and lengthening, however, and so it remained an unanswered question how and why the durational measures vary as they do in speech.

Gee and Grosjean (1983) tested the descriptive adequacy of the algorithm of Cooper and Paccia-Cooper as well as another proposed by Grosjean, Grosjean and Lane (1979) on a variety of spoken sentences. Although both algorithms explained a considerable proportion of the variation in pause durations, neither came close to explaining all of it, and neither constituted a realistic performance model for talkers.

Gee and Grosjean determined that more of the variance in pausing can be explained if domains between pauses are metrical, not syntactic, phrases. They proposed a new procedure that operates largely left to right in a sentence, producing pauses after each phonological phrase and longer pauses after phonological phrases that end an intonationl phrase. The new algorithm, besides explaining more of the variance than earlier ones, does constitute a more realistic performance model than the others, because it does not require that the talker have a whole sentence planned in order to determine how long to pause at each word boundary.

It is worth asking whether talkers *intend* to mark phonological and intonational phrases with pauses and lengthenings or whether these (and the blocking of cross-word phonological processes) are natural manifestations of occasional slowing of vocal tract activity as talkers pause to plan ahead. My guess is that the answer is 'a little of both'. On the one hand, it is probably not serendipitous that talkers mark phrase edges with *lengthenings*, with lengthenings especially on the left sides of the boundaries and with pauses that block cross-word processes. Many languages have been reported to exhibit final lengthening, while I am aware of no languages reported to show systematic final shortening. On the other hand, the patterns of lengthenings reported by Cooper and Paccia-Cooper and by Gee and Grosjean may be too systematic to reflect brakings to plan ahead, particularly since the talkers in these experiments are reading, not speaking spontaneously. My guess is that patterns of slowing have their origins in talkers' need to pause to plan ahead. Because talkers may plan coherent stretches of speech all at once, they are inclined to pause at phrase boundaries. Accordingly, the pauses and lengthenings provide information to listeners concerning the phrase structure of a sentence. More than that, the pausings tend to occur after phonological and intonational phrases and hence after 'heads' of syntactic phrases. This may help to set off or point to the

heads of phrases for the listener. Because the pauses and lengthenings are informative in these ways, talkers may tend to supply them even when they do not need to pause to plan. This may constitute another example of the 'triggering' phenomenon discussed for the example of vowel length variation and following consonant voicing by MacNeilage and Ladefoged (1976; and see the introduction to Section 3). Systematic variation having a dispositional origin in the vocal tract that is, therefore, common to most languages, may be exaggerated, stylized and incorporated in the phonologies of some languages to serve a communicative function.

Intonation and Declination

Intonational melodies are patternings of the fundamental frequency (f_0) of the talker's voice; f_0 is sensitive to several variables in speech: two important ones are transglottal pressure and the tension of the vocal folds. Transglottal pressure is the difference in air pressure above and below the vocal folds. The larger the pressure difference, the higher f_0, other things being equal. In turn, a major way for talkers to influence transglottal pressure is by increasing or decreasing the pressure below the vocal folds (subglottal pressure or P_s) by pushing more or less air out of the lungs. As for tension of the vocal folds, increasing the tension will increase f_0, other things being equal. A major way to increase vocal fold tension is to contract the cricothyroid muscle of the larynx; a major way to decrease it is to relax that muscle. (In lower frequency ranges, the 'strap' muscles of the larynx may be used to lower f_0 actively.)

In 1967, Lieberman proposed a theory of intonation according to which there are two basic melodies: the 'unmarked breath group' and the 'marked breath group'. In the former, f_0 simply tracks P_s during an expiration. According to Lieberman (see also Ladefoged, 1967), P_s is flat throughout an utterance until a final fall at the end. Therefore, f_0 is flat with a final fall utterance (or phrase) at the end. The marked breath group is similar except that the final fall in f_0 caused by the final fall in P_s is counteracted by an increase in laryngeal tension. Generally, this will cause a final rise in f_0 characteristic of yes/no questions. Lieberman recognized that contours may be more complex than the marked and unmarked breath groups. Accordingly, he proposed a feature, prominence, that could be used to accent a particular word in a sentence that the talker wanted to emphasize. In the theory, prominence is implemented by an increase in subglottal pressure.

Lieberman's theory proved quite controversial, fueling the 'lungs versus larynx' controversy (Ohala, 1978). Most controversial was Lieberman's view that pitch accents in an intonational melody are implemented by an increase in P_s. Currently, the prevailing view is that pitch accents are implemented by tensing and relaxing laryngeal muscles that stretch or shorten the vocal folds.

There is at present no psychological theory of intonational performance, and so no theory explaining how intonational melodies are produced. However, Liberman and Pierrehumbert (1984) suggest that they do not require extensive preplanning; rather, as for pausing and lengthening, they can be implemented left to right as phrases are uttered. In their view, as noted, the intonational melody, between which speakers interpolate, constitute a sequence of discrete pitch accents (approximately, Lieberman's prominence feature).

Despite general disconfirmation of Lieberman's view that intonational accents are imposed by the respiratory system, there is probably a role for systematic

variation in respiratory activity in implementing an f_0 contour. Many researchers, beginning with Pike (1945; see also Breckenridge, 1977; Cohen, Collier and t'Hart, 1982; Cohen and t'Hart, 1965; Maeda, 1976), have noticed a tendency in some styles of speech (but not all; see Lieberman *et al.*, 1985; Umeda, 1982) for f_0 to decline over the course of an interval of speech, probably an intonational phrase. Cohen and t'Hart (1965) coined the word 'declination' to refer to the fall in f_0.

The reason for declination in speech has been controversial. An intuitive reason for the fall is the reduction in lung volume between inspirations. One factor that affects P_s is the elastic recoil force of the expanded lungs. That force diminishes over the course of an expiration as the lungs deflate; other things being equal, so should P_s and f_0. Other things are not equal, of course. Expiratory muscles are increasingly recruited during an expiration to offset the fall in P_s because of lung deflation (Weismer, 1985). As I already noted, Ladefoged (1967) and Lieberman (1967) both report that P_s is flat until the final fall at the end of a breath group. Convinced that a decline in P_s could not account for declination, some researchers (Breckenridge, 1977; Ohala, 1978) concluded that declination is implemented by tensing and then relaxing muscles of the larynx that first stretch the vocal folds and then allow them gradually to shorten.

If declination is implemented by laryngeal action of this sort, then it must not be a dispositional feature of speech, but rather an intentionally implemented one. Cooper and Sorenson (1981) proposed an elaborate model for implementing declination under the assumption that talkers do intentionally impose it on their utterances. In the model, speakers estimate how long a sentence will be in seconds and they estimate when, in seconds from utterance onset, each intonational peak of the sentence will occur. Using these estimates, talkers apply a 'topline rule' to select f_0 values for the accent peaks.

The model has been criticized on a variety of grounds. It does not rationalize declination. That is, it offers no reason why talkers would implement the fall; they appear to engage in considerable computation for no apparent purpose. Moreover, the theory does not offer any insight into why declination occurs so commonly across languages (for a review, see Cooper and Sorenson, 1981). Declination occurs in most languages where it has been sought; I am aware of no languages found to exhibit some other systematic global contour shape. Simon (1980) recommends that dispositional accounts of declination be pursued in favor of this model of declination as an intentional imposition. In any case, the model does not fit the data well (Pierrehumbert and Liberman, 1982) – a fact that was somewhat masked for Cooper and Sorenson, who applied a defective means of estimating the model's fit. In addition, a simpler model, still supposing that declination is intentionally imposed, can fit the data at least as well without having to claim that f_0 contours are preplanned on a second-by-second basis (Liberman and Pierrehumbert, 1984). According to this model, speakers step f_0 down a fixed proportion of its current value at each accent.

Has an account of declination as a dispositional consequence of respiratory changes during an expiration in fact been disconfirmed? In my view, it has not been entirely. Some such account has an advantage over others as well in explaining why declination occurs so commonly across languages.

Early suggestions that P_s is flat over an utterance are not supported by later studies. Collier and Gelfer (1984) and Gelfer (1987; Gelfer, Harris and Baer, 1987) report an exponential fall in P_s over the course of a sentence that the decline in f_0

tracks quite closely. Moreover, in their data, the magnitude of the fall in P_s can explain all of the fall in f_0 except, occasionally, at the very beginning of the contour, where the starting f_0 may be increased sometimes by activity of the cricothyroid muscle (Collier, 1987).

This does not mean that declination is wholly unregulated by talkers. The fall in P_s is not a simple reflection of lung deflation because, as noted, expiratory muscles are recruited increasingly during an utterance to offset the effects of the decline in the recoil force of the lungs. Apparently they often do not offset the reduction of the recoil force entirely. Why not? Possibly, they do not because talkers intend f_0 to decline and that is how they implement declination. Alternatively, they may only offset effects of reduction in the recoil force on P_s enough to ensure sufficient transglottal pressure for phonation out to the end of an utterance (cf. Collier, 1987). Within that constraint, they allow P_s to fall as the lungs deflate, and they allow f_0 to fall with it. The latter account has the advantage of explaining why declination occurs so commonly across languages. The former account may have some validity as well, however.

As noted, talkers may tense the cricothyroid contour initially, possibly to exaggerate the contour for listeners. Second, some languages, including English, may have downstepping intonational melodies in which declination appears in an exaggerated and stylized form. Declination may represent yet another example of 'triggering' whereby a universal, dispositional, behavioral systematicity is elevated in stylized form into the phonologies of some languages, perhaps because it provides useful information to listeners here, in delimiting phrases.

5 CONCLUDING REMARKS

I have proposed that, at the levels on which I have focused, speaking occurs in two major phases: planning and performance. In planning, morphosyntactic units of an utterance – words, morphemes and phonemes – are ordered into syntactic phrases. They make themselves evident as planning units, because they occasionally misorder and they do so in characteristic ways that rule out a hypothesis that the misorderings reflect mistakes in the motor realization of speech.

Recent progress by linguists and speech production researchers has gone a considerable way toward disconfirming a generally held view that linguistic units as components of linguistic competence are not realized or even realizable in the vocal tract, because units of competence have properties, such as discreteness and context independence, that are incompatible with physical systems such as the vocal tract. The work of disconfirming the hypothesis has proceeded in two directions. Linguists have begun to focus on the previously neglected 'intrinsic content' of phonetic segments. Gestural rather than abstract featural primitives have helped to yield linguistic units ostensibly designed to be uttered. As for speech production theorists, they have stepped back from the details of vocal tract activity and found a level of more coarse-grained order. The order is achieved by coordinative structures or synergies that Saltzman and his colleagues identify with dynamic systems more generally. It appears that the smallest synergies at work in the vocal tract during speech implement the smallest, that is gestural, components of a planned utterance.

Despite the fundamental correspondence I claim for units of competence, planning and performance, something new does arise in speech performance. The something new is the grouping of words into metrical phrases – a grouping that is not apparent in speech planning as revealed by speech errors. Ostensibly the new grouping arises in each performance as a talker chooses to highlight certain content words in a sentence. The words are highlighted by pitch accents on them, and the phrases in which they participate are set off by lengthenings and pausing. These highlightings may themselves require some planning but they can, according to current viewpoints, be output largely left to right as the morphosyntactic speech plan is uttered.

APPENDIX

Unfamiliar Symbols for Phonemic and Phonological Segments

Symbol	*Example*
/a/	box
/ɝ/	bird
/I/	bit
/iy/	beed
/ow/	boat
/uw/	boot
/ŋ/	*king*

Symbol	*Interpretation*
~	nasality
V:, C:	length

ACKNOWLEDGEMENTS

The preparation of this manuscript was supported by NICHD Grant HD 01994 and NINCDS Grant NS-13617 to Haskins Laboratories.

REFERENCES

Abbs, J. and Gracco, V. (1984). Control of complex gestures: Orofacial muscle responses to load perturbations of the lip during speech. *Journal of Neurophysiology*, 51, 705–723.

Allen, G. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, 3, 75–86.

Anderson, S. (1976). Nasal consonants and the internal structure of segments. *Language*, 52, 326–345.

Baars, B., Motley, M. and MacKay, D. (1975). Output editing for lexical status from artificially-elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14, 382–391.

Beckman, M. (1985). *Stress and Nonstress Accent*. Dordrecht, The Netherlands: Foris Publications.

Bell-Berti, F. (1980). Velopharyngeal function: A spatio-temporal model. In N. Lass (Ed.), *Speech and Language*, vol. 4 (pp. 291–316). New York: Academic Press.

Bell-Berti, F. and Harris, K. (1974). More on motor organization of speech gestures. Haskins Laboratories Status Reports on Speech Research, No. SR37/38, pp. 9–20.

Bell-Berti, F. and Harris, K. (1979). Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America*, 65, 1268–1270.

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

Benguerel, A.-P. and Cowan, H. (1974). Coarticulation of upper lip protrusion in French. *Phonetica*, 30, 41–55.

Bernstein, N. (1967). *Coordination and Regulation of Movement*. London: Pergamon Press.

Bizzi, E. and Polit, A. (1979). Characteristics of the motor programs underlying visually evoked movements. In R. Talbott and D. Humphrey (Eds), *Posture and Movement* (pp. 169–176). Amsterdam: Mouton.

Bladon, A. and Al-Bamerni, A. (1982). One stage and two stage temporal patterns of coarticulation. *Journal of the Acoustical Society of America*, 72, S104 (Abstract).

Bolinger, D. (1963). Length, vowel, juncture. *Lingustics*, 1, 1–29.

Bolinger, D. (1981). *Two Kinds of Vowels, Two Kinds of Rhythm*. Bloomington, IN: Indiana University Linguistics Club.

Boyce, S. (1988). The influence of phonological structure on articulatory organization in Turkish and English: Vowel harmony and coarticulation. PhD Thesis, Yale University.

Breckenridge, J. (1977). Declination as a phonological process. Bell Laboratories Technological Memorandum, Murray Hill, NJ.

Browman, C. (1978). Tip of the tongue and slip of the ear: Implications for language processing. *UCLA Working Papers in Phonetics*, No. 42.

Browman, C. and Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 2, 219–252.

Browman, C. and Goldstein, L. (1990). Articulatory gestures as phonological units. *Phonology*, 6, 201–251.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 27, 129–158.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

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

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

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*, 2, 225–252.

Cohen, A., Collier, R. and t'Hart, J. (1982). Declination: Construct or intrinsic feature of speech pitch. *Phonetica*, 39, 254–273.

Cohen, A. and t'Hart, J. (1965). Perceptual analysis of intonation patterns. *Proceedings of the Fifth International Congress of Acoustics* (A. 16). Liege, Belgium.

Collier, R. (1987). F_0 declination: The control of its setting, resetting and slope. In T. Baer, C. Sasaki and K. Harris (Eds), *Laryngeal Function in Phonation and Respiration* (pp. 403–421). Boston, MA: College-Hill Press.

Collier, R. and Gelfer, C. (1984). Physiological explanation of f_0 declination. In M. P. R. van den Broecke and A. Cohen (Eds), *Proceedings of the Tenth International Congress of Phonetic Sciences* (pp. 354–360). Dordrecht, The Netherlands: Foris Publications.

Cooper, W. and Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press.

Cooper, W. and Sorenson, J. (1981). *Fundamental Frequency in Sentence Production*. New York: Springer.

Cordo, P. and Nashner, L. (1982). Properties of postural adjustments associated with rapid arm movement. *Journal of Neurophysiology*, 47, 287–302.

Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (Ed.), *Slips of the Tongue and Language Production* (pp. 73–108). Amsterdam: Mouton.

Cutler, A. (1980). Errors of stress and intonation. In V. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand* (pp. 67–80). New York: Academic Press.

Daniloff, R. and Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics*, 1, 239–248.

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

Davis, S. (1988). Syllable onsets as a factor in stress rules. *Phonology*, 5, 1–20.

Dell, G. (1980). Phonological and lexical encoding in speech production: An analysis of naturally occurring and experimentally elicited slips of the tongue. PhD Thesis, University of Toronto.

Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.

Dell, G. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, 27, 124–142.

Dell, G. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5, 313–349.

Dell, G. and Reich, P. (1980). Toward a unified theory of slips of the tongue. In V. Fromkin (Ed.), *Errors in Lingustic Performance: Slips of the Tongue, Ear, Pen and Hand* (pp. 273–286). New York: Academic Press.

Dell, G. and Reich, P. (1981). Stages in sentence production. *Journal of Verbal Learning and Verbal Behavior*, 20, 611–629.

Easton, T. (1972). On the normal use of reflexes. *American Scientist*, 60, 591–599.

Edwards, J. and Beckman, M. (1988). Articulatory timing and the prosodic interpretation of syllable duration. *Phonetica*, 45, 156–174.

Fant, G. (1962). Descriptive analysis of the acoustic aspects of speech. *Logos*, 5, 3–17.

Fay, D. and Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, 8, 505–520.

Fel'dman, A. and Latash, M. (1982). Interaction of afferent and efferent signals underlying joint position sense: Empirical and theoretical approaches. *Journal of Motor Behavior*, 14, 174–193.

Flege, J. (1988). The development of skill in producing word-final English stops: Kinematic parameters. *Journal of the Acoustical Society of America*, 84, 1639–1652.

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

Folkins, J. and Abbs, J. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, 18, 207–220.

Folkins, J. and Abbs, J. (1976). Additional observations on responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, 19, 820–821.

Folkins, J. and Zimmermann, G. (1981). Jaw-muscle activity during speech with the mandible fixed. *Journal of the Acoustical Society of America*, 69, 1441–1445.

Fowler, C. (1977). *Timing Control in Speech Production*. Bloomington, IN: Indiana University Linguistics Club.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8, 113–137.

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

Fowler, C. (1981b). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, 46, 127–139.

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

Fowler, C., Munhall, K., Saltzman, E. and Hawkins, S. (1986). Acoustic and articulatory evidence for consonant–vowel interactions. *Journal of the Acoustical Society of America*, 80, S96 (Abstract).

Fowler, C., Rubin, P., Remez, R. and Turvey, M. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language Production, I: Speech and Talk* (pp. 373–420). London: Academic Press.

Fowler, C. and Turvey, M. (1980). Immediate compensation for bite-block vowels. *Phonetica*, 37, 306–326.

Fromkin, V. (1971).The nonanomalous nature of anomalous utterances. *Language*, 47, 27–52.

Fromkin, V. (Ed.) (1973). *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Fudge, E. (1969). Syllables. *Journal of Linguistics*, 5, 253–286.

Garrett, M. (1980a). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language Production I: Speech and Talk* (pp. 177–220). London: Academic Press.

Garrett, M. (1980b). The limits of accommodation. In V. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand* (pp. 263–271). New York: Academic Press.

Gay, T. (1977). Cinefluorographic and electromyographic studies of articulatory organization. In M. Sawashima and F. Cooper (Eds), *Dynamic Aspects of Speech Production* (pp. 85–102). Tokyo: University of Tokyo Press.

Gay, T., Lindblom, B. and Lubker, J. (1981). Production of bite-block vowels: Acoustic equivalence by selective compensation. *Journal of the Acoustical Society of America*, 69, 802–810.

Gee, P. and Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411–458.

Gelfand, I. M., Gurfinkel, V., Tsetlin, M. and Shik, M. (1971). Some problems in the analysis of movements. In I. Gelfand, V. Gurfinkel, S. Fomin and M. Tsetlin (Eds), *Models of the Structural–Functional Organization of Certain Biological Systems* (pp. 329–345). Cambridge, MA: MIT Press.

Gelfer, C. (1987). A simultaneous physiological and acoustic study of fundamental frequency declination. PhD Thesis, CUNY.

Gelfer, C., Bell-Berti, F. and Harris, K. (1982). Determining the extent of coarticulation: Effects of experimental design. Paper presented at the 103rd meeting of the Acoustical Society of America, Chicago.

Gelfer, C., Harris, K. and Baer, T. (1987). Controlled variables in sentence intonation. In T. Baer, C. Sasaki and K. Harris (Eds), *Laryngeal Function in Phonation and Respiration* (pp. 422–435). Boston, MA: College-Hill Press.

Goldsmith, J. (1976). *Autosegmental Phonology*. Bloomington, IN: Indiana University Linguistics Club.

Gracco, V. and Abbs, J. (1985). Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes. *Journal of Neurophysiology*, 54, 418–432.

Gracco, V. and Abbs, J. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research*, 65, 156–166.

Grillner, S. (1981). Control of locomotion in bipeds, tetrapods and fish. In V. B. Brooks (Ed.), *Handbook of Physiology: Motor Control* (pp. 1179–1236). Baltimore, MD: Williams and Wilkins.

Grosjean, F., Grosjean, L. and Lane, H. (1979).The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11, 58–81.

Hammarberg, R. (1976). The metaphysics of coarticulation. *Journal of Phonetics*, 4, 353–363.

Hammarberg, R. (1982). On redefining coarticulation. *Journal of Phonetics*, 10, 123–137.
Hayes, B. (1981). A metrical theory of stress rules. PhD Thesis, MIT.
Hayes, B. (1982). Extrametricality of English stress. *Linguistic Inquiry*, 13, 227–276.
Henke, W. (1966). Dynamic articulatory modeling of speech production using computer simulation. PhD Thesis, MIT.
Hockett, C. (1960). The origin of speech. *Scientific American*, 203, 89–96.
Hyman, L. (1982). The representation of nasality in Gokona. In H. van der Hulst and N. Smith (Eds), *The Structure of Phonological Representations*, Part I (pp. 111–130). Dordrecht, The Netherlands: Foris Publications.
Jordan, M. (1986). *Serial Order: A Parallel Distributed Processing Approach*. Institute for Cognitive Science, University of California, San Diego.
Kahn, D. (1976). *Syllable-based Generalizations in English Phonology*. Bloomington, IN: Indiana University Linguistics Club.
Keating, P. (1983). Comments on the jaw and syllable structure. *Journal of Phonetics*, 11, 401–406.
Keating, P. (1990). Mechanisms of coarticulation: The window model of coarticulation: Articulatory evidence. In J. Kingston and M. Beckman (Eds), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (pp. 451–470). Cambridge: Cambridge University Press.
Kelly, M. (1988). Rhythmic alternation and lexical stress differences in English. *Cognition*, 29, 107–138.
Kelly, M. and Bock, K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 389–413.
Kelso, J. A. S., Saltzman, E. and Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, 14, 29–56.
Kelso, J. A. S., Southard, D. and Goodman, D. (1979). On the coordinating of two-handed movements. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 229–238.
Kelso, J. A. S. and Tuller, B. (1987). Intrinsic time in speech production: Theory, methodology and preliminary observations. In E. Keller and M. Gopnik (Eds), *Motor and Sensory Processes in Language* (pp. 203–222). Hillsdale, NJ: Erlbaum.
Kelso, J. A. S., Tuller, B. and Saltzman, E. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, 14, 29–60.
Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E. and Fowler, C. (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.
Kenstowicz, M. and Kisseberth, C. (1979). *Generative Phonology*. San Diego, CA: Academic Press.
Klatt, D. (1976). The linguistic uses of segment duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208–1221.
Kozhevnikov, V. and Chistovich, L. (1965). *Speech: Articulation and Perception*. Washington, DC: Joint Publications Research Service, No. 30, p. 543.
Krakow, R. (1988). *Articulatory Organization and the Structure of Words and Syllables*. Meeting of the American Speech and Hearing Association.
Krakow, R. (1989). The articulatory organization of syllables: A kinematic analysis of labial and velar gestures. PhD Thesis, Yale University.
Kugler, P., Kelso, J. A. S. and Turvey, M. (1980). On the concept of coordinative structures as dissipative structures, I. Theoretical lines of convergence. In G. Stelmach and J. Requin (Eds), *Tutorials in Motor Behavior* (pp. 3–47). Amsterdam: North-Holland.
Kugler, P. and Turvey, M. (1987). *Information, Natural Law and the Self-Assembly of Rhythmic Movements*. Hillsdale, NJ: Erlbaum.
Ladefoged, P. (1967). *Three Areas in Experimental Phonetics*. London: Oxford University Press.

Lashley, K. (1951). The problem of serial order in behavior. In L. Jeffress (Ed.), *Cerebral Mechanisms in Behavior* (pp. 112–136). New York: John Wiley.
Lee, W. (1984). Neuromotor synergies as a basis for coordinated intentional action. *Journal of Motor Behavior*, 16, 135–170.
Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, 2018–2024.
Liberman, M. (1975). The intonational system of English. PhD Thesis, MIT.
Liberman, M. and Pierrehumbert, J. (1984). Intonational invariances under changes in pitch range and length. In M. Aronoff and R. Oehrle (Eds), *Language Sound Structure* (pp. 157–233). Cambridge, MA: MIT Press.
Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249–336.
Lieberman, P. (1967). *Intonation, Perception and Language*. Cambridge, MA: MIT Press.
Lieberman, P., Katz, W., Jongman, A., Zimmerman, R. and Miller, M. (1985). Measures of the sentence intonation of spontaneous speech in American English. *Journal of the Acoustical Society of America*, 77, 649–657.
Lindblom, B. (1986). On the origin and purpose of discreteness and invariance in sound patterns. In J. Perkell and D. Klatt (Eds), *Invariance and Variability of Speech Processes* (pp. 493–510). Hillsdale, NJ: Erlbaum.
Lindblom, B., Lubker, J. and Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech-motor programming by predictive simulation. *Journal of Phonetics*, 7, 147–161.
Lindblom, B., Lubker, J., Gay, T., Lyberg, B., Branderud, P. and Holmgren, K. (1987). The concept of target and speech timing. In R. Channon and L. Shockey (Eds), *In Honor of Ilse Lehiste* (pp. 161–181). Providence, RI: Foris Publications.
Lindblom, B., Lyberg, B. and Holmgren, K. (1981). *Durational Patterns of Swedish Phonology: Do They Reflect Short-term Memory Processes?* Bloomington, IN: Indiana University Linguistics Club.
Lindblom, B. and Rapp, K. (1973). *Some Temporal Regularities of Spoken Swedish*. Papers in Linguistics from the University of Stockholm, No. 21, pp. 1–59.
Lindblom, B. and Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw and larynx movement. *Journal of the Acoustical Society of America*, 50, 1166–1179.
Locke, J. (1983). *Phonological Acquisition and Change*. New York: Academic Press.
Löfqvist, A. (1980). Interarticulator programming in stop production. *Journal of Phonetics*, 68, 792–801.
Löfqvist, A. and Yoshioka, H. (1984). Intrasegmental timing: Laryngeal–oral coordination in vowel–consonant production. *Speech Communication*, 3, 279–289.
Lubker, J. (1979). The reorganization times of bite block vowels. *Phonetica*, 36, 273–293.
Lubker, J. and Parris, P. (1970). Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop-consonant cognates. *Journal of the Acoustical Society of America*, 47, 625–633.
Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America*, 71, 173–178.
MacKay, D. M. (1982). The problem of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review*, 84, 483–506.
MacKay, D. M. (1987). *The Origin of Perception and Action*. New York: Springer.
MacNeilage, P. (1970). Motor control of serial ordering of speech. *Psychological Review*, 77, 182–196.
MacNeilage, P. and DeClerk, J. (1969). On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America*, 45, 1217–1233.
MacNeilage, P. and Ladefoged, P. (1976). The production of speech and language. In E. Carterette and M. Friedman (Eds), *Handbook of Perception: Language and Speech* (pp. 75–120). New York: Academic Press.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.
Maeda, S. (1976). A characterization of American English intonation. PhD Thesis, MIT.
Manuel, S. (1987). Acoustic and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages. PhD Thesis, Yale University.
Manuel, S. and Krakow, R. (1984). Universal and language-specific aspects of vowel-to-vowel coarticulation. Haskins Laboratories Status Reports on Speech Research, No. SR 77/78, pp. 69–87.
McClelland, J. and Rumelhart, D. (1981). An interactive activation model of context effects in letter perception. Part I: An account of basic findings. *Psychological Review*, 88, 375–407.
Moll, K. (1962). Velopharyngeal closure on vowels. *Journal of Speech and Hearing Research*, 5, 30–77.
Moll, K. and Daniloff, R. (1971). Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America*, 50, 678–684.
Moll, K., Zimmermann, G. and Smith, A. (1977). The study of speech production as a human neuromotor system. In M. Sawashima and F. S. Cooper (Eds), *Dynamic Aspects of Speech Production* (pp 107–127). Tokyo: University of Tokyo Press.
Monsell, S. (1986). Programming of complex sequences: Evidence from the timing of rapid speech and other productions. In H. Heuer and C. Fromm (Eds), *Experimental Brain Research Series*, vol. 15: *Generation and Modulation of Action Patterns* (pp. 72–86). Berlin: Springer.
Munhall, K., Löfqvist, A. and Kelso, J. A. S. (1986). Phase-dependent sensitivity to perturbation reveals the nature of speech coordinative structures. *Journal of the Acoustical Society of America*, 80, S38.
Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.
Nespor, M. and Vogel, I. (1979). Clash avoidance in Italian. *Linguistic Inquiry*, 10, 467–482.
Nittrouer, S., Munhall, K., Kelso, J. A. S., Tuller, B. and Harris, K. (1988). Patterns of interarticulatory phasing and their relation to linguistic structure. *Journal of the Acoustical Society of America*, 84, 1653–1661.
Nooteboom, S. (1973). The perceptual reality of some prosodic durations. *Journal of Phonetics*, 1, 25–45.
Norman, D. (1981). Categorization of action slips. *Psychological Review*, 88, 1–15.
Ohala, J. (1970). Aspects of the control and production of speech. UCLA Working Papers in Phonetics, No. 15.
Ohala, J. (1978). Production of tone. In V. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 5–39). New York: Academic Press.
Öhman, S. (1966). Coarticulation in VCV utterances. *Journal of the Acoustical Society of America*, 39, 151–168.
Ostry, D., Keller, E. and Parush, A. (1983). Similarities in the control of the speech articulators and the limbs: Kinematics of tongue dorsum movements in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 622–636.
Pattee, H. (1973). The physical basis and origin of hierarchical control. In H. Pattee (Ed.), *Hierarchy Theory: The Challenge of Complex Systems* (pp. 71–108). New York: Braziller.
Perkell, J. (1969). *Physiology of Speech Production: Results and Implications of a Quantitiative Cineradiographic Study*. Cambridge, MA: MIT Press.
Perkell, J. (1980). Phonetic features and the physiology of speech production. In B. Butterworth (Ed.), *Language Production, I: Speech and Talk* (pp. 337–372). London: Academic Press.
Perkell, J. (1986). Coarticulatory strategies: Preliminary implications of a detailed analysis of lower lip protrusion gestures. *Speech Communication*, 5, 47–68.
Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. PhD Thesis, MIT.
Pierrehumbert, J. and Liberman, M. (1982). Modeling the fundamental frequency of the voice (Review of Cooper and Sorenson (1981)). *Contemporary Psychology*, 27, 690–692.

Pike, K. (1945). *The Intonation of American English.* Ann Arbor, MI: University of Michigan Press.

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

Raphael, L. (1975). The physiological control of durational control between vowels preceding voiced and voiceless consonants in English. *Journal of Phonetics,* 3, 25–35.

Reason, J. T. (1979). Actions not as planned. In G. Underwood and R. Stevens (Eds), *Aspects of Consciousness* (vol. 1, 67–90). London: Academic Press.

Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America,* 76, 1624–1635.

Repp, B. (1981). On levels of description in speech research. *Journal of the Acoustical Society of America,* 69, 1462–1464.

Rosenbaum, D. (1985). Motor programming: A review and scheduling theory. In H. Heuer, U. Kleinbeck and K.-H. Schmidt (Eds), *Motor Behavior: Programming, Control and Acquisition* (pp. 1–33). Berlin: Springer.

Rosenbaum, D., Kenny, S. and Derr, M. (1983). Hierarchical and nonhierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance,* 9, 86–102.

Ryle, G. (1949). *The Concept of Mind.* New York: Barnes and Noble.

Sagey, E. (1986). The representation of features and relations in non-linear phonology. PhD Thesis, MIT.

Saltzman, E. (1986). Task-dynamic coordination of the speech articulators: A preliminary model. In H. Heuer and C. Fromm (Eds), *Experimental Brain Research Series,* vol. 15: *Generation and Modulation of Action Patterns* (pp. 129–144). New York: Springer.

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

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

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

Selkirk, E. (1980b). *On Prosodic Structure and its Relation to Syntactic Structure.* Bloomington, IN: Indiana University Linguistics Club.

Selkirk, E. (1984). *Phonology and Syntax: The Relation Between Sound and Structure.* Cambridge, MA: MIT Press.

Shaiman, S. and Abbs, J. (1987). Phonetic task-specific utilization of sensorimotor activity. Paper presented to the American Speech and Hearing Association.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. Cooper and E. Walker (Eds), *Sentence Processing* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. MacNeilage (Ed.), *The Production of Speech* (pp. 109–136). New York: Springer.

Shattuck-Hufnagel, S. (1986). The representation of phonologial information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook,* 3, 117–149.

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech errors. In E. Keller and M. Gopnik (Eds), *Motor and Sensory Processing of Language* (pp. 17–52). Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. and Klatt, D. (1979). Minimal use of features and markedness in speech production. *Journal of Verbal Learning and Verbal Behavior,* 18, 41–55.

Shields, J., McHugh, A. and Martin, J. (1974). Reaction time to phonemic targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology,* 102, 250–255.

Simon, H. (1980). How to win at twenty questions with nature. In R. Cole (Ed.), *Perception and Production of Fluent Speech* (pp. 535–548). Hillsdale, NJ: Erlbaum.

Stemberger, J. (1983). *Speech Errors and Theoretical Phonology: A Review*. Bloomington, IN: Indiana University Linguistics Club.

Stemberger, J. (1985). An interactive activation model of language production. In A. Ellis (Ed.), *Progress in the Psychology of Language*, vol. 1 (pp. 143–186). London: Erlbaum.

Stemberger, J. and MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected words. *Memory and Cognition*, **14**, 17–26.

Sternberg, S., Monsell, S., Knoll, R. and Wright, C. (1978). The latency and duration of rapid movement sequences: Comparison of speech and typing. In G. Stelmach (Ed.), *Information Processing in Motor Control and Learning* (pp. 117–152). New York: Academic Press.

Sternberg, S., Wright, C., Monsell, S. and Knoll, R. (1980). Motor programs in rapid speech: Additional evidence. In R. Cole (Ed.), *Perception and Production of Fluent Speech* (pp. 507–534). Hillsdale, NJ: Erlbaum.

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production. *Journal of the Acoustical Society of America*, **82**, 847–863.

Sussman, H., MacNeilage, P. and Hanson, R. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research*, **16**, 385–396.

Sussman, H. and Westbury, J. (1981). The effects of antagonist gestures in temporal and amplitude parameters of anticipatory labial coarticulation. *Journal of the Acoustical Society of America*, **46**, 16–24.

Terzuolo, C. and Viviani, P. (1979). The central representation of learned motor patterns. In R. Talbott and D. Humphrey (Eds), *Posture and Movement* (pp. 113–121). New York: Raven Press.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition*, **15**, 49–74.

Treiman, R. (1984). On the status of final consonant clusters in English. *Journal of Verbal Learning and Verbal Behavior*, **23**, 343–356.

Treiman, R. (1986). On the status of final consonant clusters in English. *Journal of Memory and Language*, **25**, 476–491.

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

Umeda, N. (1982). 'F_0 declination' is situation dependent. *Journal of Phonetics*, **10**, 279–290.

van der Hulst, H. and Smith, N. (1982). An overview of autosegmental and metrical phonologies. In H. van der Hulst and N. Smith (Eds), *The Structure of Phonological Representations*, Part I (pp. 1–46). Dordrecht, The Netherlands: Foris Publications.

Vayra, M., Avesani, C. and Fowler, C. (1984). Patterns of temporal compression in spoken Italian. In M. P. R. van den Broecke and A. Cohen (Eds), *Proceedings of the Tenth International Congress of Phonetic Sciences* (pp. 541–546). Dordrecht, The Netherlands: Foris Publications.

Weismer, G. (1985). Speech breathing: Contemporary views and findings. In R. Daniloff (Ed.), *Speech Science* (pp. 47–72). San Diego, CA: College-Hill Press.

Weiss, P. (1941). Self-differentiation of the basic pattern of coordination. *Comparative Psychology Monographs*, **17**, 21–96.