In Ray Daniloff (Ed.) Recent advances in speech, hearing and language, Vol. 4.

CHAPTER 6

Current Perspectives on Language and Speech Production: A Critical Overview

Carol A. Fowler

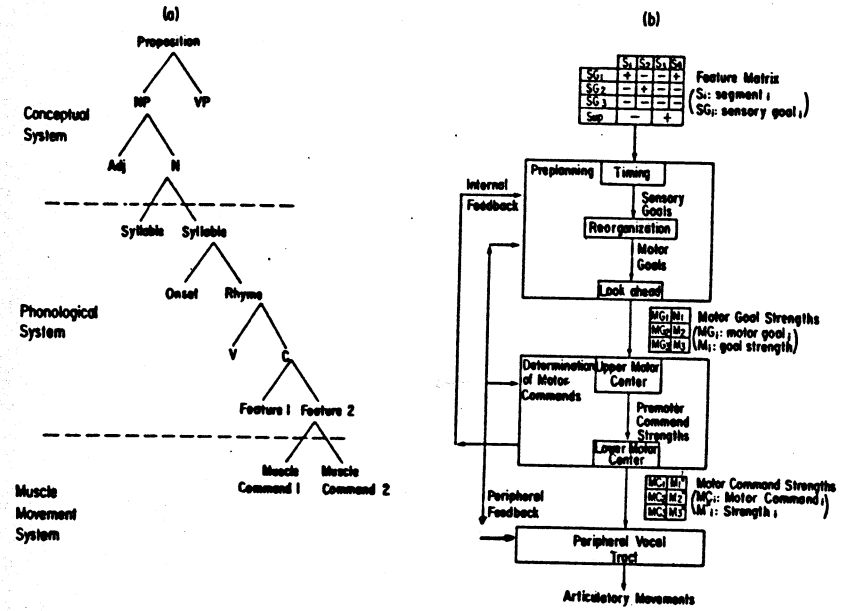
Speech presents two quite distinct aspects to the researcher. On the one hand, it is a linguistic communication consisting of symbols combined by grammatical rules. On the other hand, it is a complex motor skill in which activities of the respiratory system, the larynx, and the supralaryngeal vocal tract are coordinated at several temporal scales.

These different aspects of speech invite theorizing of distinct types. When linguists, psycholinguists, and cognitive psychologists study speech in its linguistic–representational aspect, they focus on language as a "mental" or cognitive capability and attempt to discover and formalize private representational correlates of articulatory performance. When phoneticians, speech scientists, and students of motor performance study speech, they focus on its physical manifestations and ask how the primitive units of a linguistic description can be realized as coordinated gestures of the vocal tract.

Perhaps because the labor of studying speech is divided in this way, we know very little about the relationships between its two aspects. An assumption reflected in many models of language production (used here to mean putative private mental events occurring prior to and during actual speech utterance) and of speech production (used here to refer to aspects of motor control during speaking) is that the different aspects constitute different *phases* in a speech event; in particular, the output of the representational component serves as the input to the motor component. Indeed, in some instances, models of language production and of speech production appear to dovetail relatively well because the one model leaves off where the other begins.

An example is provided in Figure 6-1. Figure 6-1*a* and *b* are adaptations of recently proposed models of language production (MacKay,

Figure 6–1. Recent models of language production (MacKay, 1982) and speech production (Perkell, 1980).



1982) and of speech production (Perkell, 1980). MacKay's model posits three global stages in utterance production: a conceptual stage in which ideas are realized as a grammatical sequence of words, a phonological stage in which words are realized as sequences of phonetic feature specifications for individual phonological constituents of a word, and a motor stage in which feature specifications become motor commands. In contrast, Perkell's model begins with feature specifications and details hypothetical events intervening between those specifications and movements of vocal tract structures. His model also has three global stages, consisting of a preplanning stage, a stage in which motor commands are determined, and one in which movements are produced. Together, then, the two models outline five or six global stages in the process of utterance production. At least the first three of these, and perhaps more, occur in the representational–cognitive phase of language production; the last necessarily is motor.

The assumption that the representational and motor aspects of speech relate sequentially as they do in these models is largely unexamined and should not be accepted uncritically. There are several problems, apparent

or real, associated with a view that at some point in a chain of events underlying speaking, the events cease to be cognitive and representational and become motor instead.

One problem, of course, is the deep philosophical puzzle of how mental and physical events are related (or of whether it is even sensible to address the puzzle couched in this way [Ryle, 1949]). Largely, this difficulty is handled in language and speech production models by a kind of sleight of hand. The final stage in the model of language production in Figure 6–1 and an intermediate stage in the model of speech production is a bridging stage, in which "motor commands" are sent out to the vocal tract. Proposing this bridging stage is a sleight-of-hand maneuver because, whereas "commands" are things that arise in the mental or cognitive domain, the presumed recipients of the commands, motoneurons or muscles, exist in the physical realm. In other usages, motoneurons and muscles are not seen as the kinds of things that obey commands; rather, they respond to release of transmitter substances. The kinds of things that can obey commands are sentient, sapient beings. Therefore, the concept of motor command (or the like) is a metaphor for which there may or may not exist a real-world interpretation.

There are other, less abstruse difficulties with the practice of assuming without study that control of speaking involves a sequence of cognitive–representational events followed by a sequence of motor events. For example, researchers have not typically asked explicitly which issues and which phenomena require explanation in cognitive or linguistic terms and which require explanation in motor process terms. One consequence of neglecting careful consideration of these questions is that some issues and phenomena may be addressed and given incompatible accounts in both realms and others may be left aside. The large number of hypothetical processing stages underlying speaking that emerge from an attempt to interface extant models of speech and language production (as in Figure 6–1) suggests that there may indeed be some redundancy with respect to the production events for which each presumed phase takes responsibility. Other events or issues may be neglected. For example, the second global stage in MacKay's model is the domain of the phonological system. The model is laid out as if this stage stands in a hierarchical relation both to the conceptual stage and the motor stage. That is, it appears as if relatively large units of the syntactic structure of a sentence (clauses, words) are *partitioned* into relatively smaller phonological units (syllables, syllable constituents, phonological segments, and features) that in turn are partitioned into constituent motor commands. But this layout is somewhat misleading. In fact, the units of the "metrical structure" of an utterance are not just syllable constituents and syllables, but in addition are stress

feet (e.g., Liberman and Prince, 1977) and possibly "prosodic words," phonological phrases, and intonational phrases as well (e.g., Selkirk, 1980a, b). These structures emerge as relevant and necessary to attempts to capture systematic differences in the relative prominence of syllables in an utterance and also, perhaps, to rationalize talkers' pausing patterns in speech (Gee and Grosjean, 1983). The metrical structures are not *constituents* of syntactic units. Rather, syllables, stress feet, phonological phrases, and intonational phrases are similar in size, respectively, to morphemes, words, syntactic phrases, and clauses, but they may be different in domain. For example, the utterance "Omaha Nabraska" consists lexically of two words, but metrically of three stress feet, "Ohma," "haNe," and "braska," which violate the lexical structure of the utterance. Instead of standing in a subordinate—superordinate relationship to syntactic units, then, metrical structures appear to relate to syntactic units as one perspective on a Necker cube relates to the other. That is, they emerge from different perspectives on approximately the same grain-size of analysis of the language. (It may still be the case, however, as MacKay intends, that, in a sequence of events preceding articulation, a talker's specification of the phonological [and metrical] structure of an utterance is temporally later than his or her specification of the syntactic structure.)

Metrical structures apparently stand in a closer relation to articulatory performance than do syntactic units. Whereas syntactic units are preserved in written language productions, metrical structures need not be and frequently are not. The metrical structures, then, may have their rationale—and they need one from production theorists—in special characteristics or strategies of the articulatory system that are required when it confronts the task of realizing the syntactic units of language.

The many puzzles to which the existence of these structures gives rise have largely been neglected. Linguists have attempted to capture their systematic properties (Liberman and Prince, 1977; Prince, 1983; Selkirk, 1980a); however, their function in utterance production is unknown. If syntactic units indeed are "packaged" into metrical structures before (or at the same time as) speech is produced, there must be reasons why they are. However, metrical structures have no discernible role in MacKay's model and they do not appear in Perkell's. Something important is missing from the models—possibly because of the meager attention devoted to issues surrounding the relationship between the representational-cognitive and motor aspects of speech.

There is a final problematic consequence of dividing the labor of studying spoken language along the lines cognitive and motor. In particular, the partitioning might exacerbate the difficulty of integrating the domains of speech production if such an attempt were to be made. Because the units of articulation do seem different in some ways from those of language viewed as a representational system, investigators in the representational domain feel free to ignore research on articulation. For example, according to Dell (1980), "What we know of production comes from a rather unique source of data—speech errors or slips of the tongue" (p. 1). Similarly, some investigators in the speech production realm are tempted to ignore the units of language's representational aspect. For example, MacNeilage and Ladefoged (1976) recognize, among speech production researchers,

> an increasing realization of the inappropriateness of conceptualizing the dynamic processes of articulation itself in terms of discrete, static, context-free linguistic categories, such as "phoneme" and "distinctive feature" This development does not mean that these linguistic categories should be abandoned as there is considerable evidence for their behavioral reality (Fromkin, 1971). Instead, it seems to require that they be recognized, even more than before, as too abstract to characterize the actual behavior of 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)

Yet it is clear that the separation between the representational and motor aspects of language is not entirely clean. First, as already noted, some of the systematic properties of the phonologies of language—in particular, their metrical structures—apparently must have, at least in part, a rationale in articulatory terms. Second, other aspects of the phonologies of language—the popularity of certain "natural" rule types (Donegan and Stampe, 1979), the segment inventories of languages (Lindblom, 1971), and aspects of their diachronic changes (Ohala, 1981)—reflect fairly clearly the fundamental bond between language, the vocal tract, and the ear. Languages evolved to be spoken and heard. On the other side, the character of speech as a motor skill is affected by the fact that articulation realizes a language. To a large extent, the sequencing of articulatory events in speech is dictated by the language's grammar. A special property of grammars is that the constraints on symbol combination that they impose are arbitrary with respect to vocal tract movement capabilities. (That is, for example, the vocal tract has no preference for subject-verb-object orderings.) Perhaps this arbitrariness explains why ordering errors are so common in speech but are relatively rare in activities in which the sequencing of events is not arbitrary from the perspective of the implementing system. So, for example, we never mistakenly attempt to inhale twice without exhaling in between; and, while walking, we never inadvertently take two steps with our left foot without taking one with the right foot in between. There is, clearly, a serial ordering problem in

talking that is diminished in activities in which "cognitive" choices are limited (Dell, 1980).

In the present chapter, some recent accounts of language production are reviewed, followed by some of the evidence and proposals concerning regulation of speech production. The review is intended to make two major points in itself and to serve as the basis for speculation concerning the relationships between language and speech.

One major point to make clear is that when proposals concerning the "mental processing" involved in language and speech production are gathered together, and the talker's many hypothesized processing representations are examined, they loom impossibly large. This is perhaps already evident in Figure 6–1*a* and *b*, but the model becomes even more outsized when proposals concerning processing to achieve durational patterns, pausing patterns, and intonation in language production and to achieve durational patterns in speech prooduction are added to the models. Future theoretical efforts will need to be directed toward collapsing and categorizing these different sources of evidence and toward minimizing rather than proliferating hypothetical processing stages.

A second major point, already alluded to, is that despite the size that any integrated model of language and speech would take, it would fail to provide an understanding of the role of speech in language. A realistic model of language production will have to provide that understanding.

LANGUAGE PRODUCTION

In this chapter, issues surrounding the pragmatics of conversation are left aside to focus narrowly on those issues surrounding the production of an utterance by a talker who has already selected its content. That is, only at models that begin more or less where MacKay's model begins are discussed.

An adequate theory of language production, restricting itself to this limited domain, would have to address a variety of issues and account for certain performance measures.

Issues

Two central concerns for language production theories are considered. One is to characterize and explain the capabilities to which linguistic competence gives rise. The other is to rationalize language's multiple levels and kinds of structure.

Capabilities of a Language User. Communications by humans need not, although they can, express emotional states; likewise, they need not, but can, refer to ongoing events. That they need not either express emotional states or refer to events in the here and now probably sets them off from nonlinguistic communications by other species and follows from a great evolutionary discovery of language: the use of rules and representations that are largely conventional in nature, and wholly conventional in function, and that thereby realize a separation at once of function and form and form and substance in language.

The rules and representations of language are conventional in two senses. First, they are largely arbitrary both with respect to the physical system that realizes them (here, the talker, and in particular, his or her vocal tract) and with respect to their significations. This frees utterances from serving only as *signs* either of internal or of external states of affairs. Second, the rules and representations of language are conventional in that they are shared by a linguistic community. Because they are shared, a language user can count on his or her communication being understood even though it consists of symbols rather than signs and of orderings of symbols that are "frozen accidents" (Pattee, 1973).

Two other salient capabilities of language users are those of producing and understanding utterances they have never heard spoken. To theorists, this generativity in language use implies two underpinnings. First, language is "rule governed"; second, spoken utterances are planned. Knowing the rules of a system allows a participant to generate any and only legal instances in the system. However, using the rules of a system—in particular, using the grammatical rules of a language—requires planning. In speech, a plan has to span the domain of a rule, which may be several words in extent. Explaining generativity and planning are central concerns of a theory of language production.

Levels of structure in speech. Another concern of language production theories is to rationalize language's multiple structural aspects. First, as Hockett (1960) has pointed out, languages have "duality of patterning." They consist of meaningless segments and rules for their combination. One reason for duality of patterning in language is fairly uncontroversial. Providing sentences with an internal structure of words makes the class of sentences that can be uttered and understood open rather than closed. Similarly, providing the primitive meaningful units of language—words or morphemes—with an internal structure opens up the lexicon of a language. Were each word or morpheme a unique utterance without internal

structure, we would, perhaps, soon run out of words we could coin and remember.

A different perspective on the multiplicity of structured aspects to language has already been alluded to and becomes salient when the interface between language and speech production is contemplated. Languages consist of units that participate in rules for creating sentences and they consist of other units that do not. Units of the first type—phonological segments, morphemes, words, syntactic phrases, and clauses—tend to exist in all human language systems, spoken and signed, and they tend be preserved in derivative writing systems. Units of the second type—syllable constituents, syllables, stress feet, and possibly larger metrical units—apparently do not exist in sign and are not, largely, preserved in writing systems.

Performance Measures

In addition to the issues just considered, the literature offers several performance measures for a theory of language to address. Oldest, and perhaps most productive of research, are speech errors. Errors of spontaneous language production have long been viewed as providing a window to the mind behind an utterance (Freud, 1958; Merringer and Meyer, 1895). Recently, spontaneous error collections have been supplemented by experimentally induced errors (e.g., Baars, 1980; Dell, 1980; Kupin, 1979) and by simulations of error-producing language systems (Dell, 1980), and researchers have begun seeing the whole array of errors as providing a window, specifically, to the cognitive-representational aspects of utterance production.

Similarly, pausing by talkers (or segmental lengthening) has recently been used as an index of planning (Cooper and Paccia-Cooper, 1980) or execution (Gee and Grosjean, 1983) in production. Measures of fundamental frequency declination (Breckenridge, 1977; Cooper and Sorenson, 1981) apparently provide compatible, or even redundant, information.

Two final performance measures used to study language production are less detectable in spontaneously produced speech than in experimentally provoked utterances. They are latency to begin producing a planned utterance and utterance duration. These measures vary in systematic and interesting ways with certain structural properties of an utterance (Sternberg, Monsell, Knoll, and Wright, 1978) and with practice (MacKay, 1982) and are used as the basis for inferences about speech planning and execution.

Models of Language Production

Models of language production, or more loosely, proposals concerning its underpinnings, suffer somewhat from narrowness of scope. With few exceptions, they have been based largely on the patterning of just one dependent measure. The measure for many researchers is speech errors, for others it is pausing, and for a few, utterance latency and duration. This chapter is organized, therefore, primarily around the different dependent measures and secondarily around model types.

Speech Errors

Logic. An utterance contains a speech error if its producer agrees that it deviates from his or her intended utterance. Excluded from consideration, therefore, are productions that other listeners would consider deviant but that conform to what the talker meant to utter.

To a degree, the strategy for drawing inferences from speech errors is similar to that suggested by the Russian theorist and physiologist Bernstein (1967). Bernstein suggested that the design character of an unknown system can be determined by discovering what classes of tasks the system accomplishes with equal ease and what classes it accomplishes with difficulty or not at all. For example, with a compass it is possible to draw circles of many radii with equal ease, but an ellipse or a rectangle only with difficulty; a compass is designed for drawing circles.

The logic behind studying speech errors is similar in part. By discovering the conditions under which errors occur frequently, researchers learn something of the hidden workings of the system behind the verbal output. For example, phonological segments frequently are anticipated, perseverated, or exchanged as in Examples 1 to 3 below. (The errors reported here are either from a small corpus of the author's, or, where noted, from published examples.)

1. tree lined lane—tree laned
2. sore shoulder—sore soulder
3. face painted—pace fainted

Two striking properties of these errors are that the source and destination contexts of the migrating segment or segments are phonologically very similar, and interacting segments themselves are phonologically similar. In Example 1, for instance, the vowel /ey/ occurs in the context /l/—/n/ in both the source and the destination word. Similarly, in Example 2, the segment /s/ occurs in the context #—V in both the source and the destination word. In Example 3, not only are the contexts similar from

which and to which the exchanging segments move, but also the exchanging segments themselves are similar. Both /f/ and /p/ are voiceless, labial consonants.

These are well-documented characteristics of sound errors, and presumably they reveal that, for the job of ordering phonological segments, similar segments—particularly in similar contexts—make the ordering job difficult. A model of production that includes an ordering mechanism, then, will have to incorporate a like fallibility.

Examples 1 to 3 are seen as providing evidence not only about the processes underlying utterance production, but also about the structures on which the processes work. That phonological segments move frequently in errors implicates them as discrete units in language production. Indeed, speech errors may provide the strongest behavioral evidence currently available that phonological segments (at some, as yet undetermined, level of abstraction) are "psychologically real."

The errors above also provide information about the "window" of speech on which processes work when phonological segments are being ordered. Anticipation errors indicate that the window extends beyond a word in which segments are being selected or ordered. Perseveration errors indicate that the window also extends backward in articulatory time. According to Garrett's findings (1980a), in approximately 87% of exchange errors the window extends to, but not across, a phrase (e.g., noun phrase [NP] or verb phrase [VP]) boundary. Thus, the range over which speech elements can interact may provide information about the domain of a speech plan.

However, these inferences have to be drawn cautiously. The evidence provided by Garrett's percentages are not taken to signal the window size of *the* speech plan. Different kinds of errors may have different domains, and based on this and other evidence, investigators have tended to posit a variety of plans in utterance production, each with its own window and its own job to perform in constructing an utterance.

Speech errors sometimes suggest inferences about orderings of events in utterance production. In Example 4 (from Garrett, 1980a), a third person singular morpheme shifts from one word to another. (This shift is classified as a morpheme shift rather than a phonological segment shift because word-final segments that are morphemes are vastly overrepresented in word-final segment shifts.)

4. It certainly runs out fast—ran outs fast
/s/

In shifting, the phonological realization of the morpheme accommodates to its new context. Whereas the third person singular morpheme would be realized /z/ in "runs," it is realized as /s/ in "outs." This is interpreted by Garrett and others as revealing that the error occurred prior to a stage of language production in which the phonetic form of morphemes is determined.

Garrett (1980a) proposes two additional guidelines for drawing inferences from speech errors. These take the form of plausible assumptions that reduce the ambiguity associated with drawing inferences. One assumption is that if two elements interact in an error they are elements of the same descriptive type. The second is that the set of conditions underlying the occurrence of a particular kind of error will always be conditions of just one descriptive type.

A final assumption made by error collectors is that, despite being labeled "slips of the tongue," speech errors occur "in the head," not "in the mouth." That is, they are errors of language production, not of speech production. Evidence for this view is that speech errors of approximately the same types and in approximately the same proportions are reported by subjects engaging in internal speech as by subjects talking aloud (Dell, 1980). Additionally, Baars, Motley, and MacKay (1975) elicit slips of the tongue by inducing subjects to develop covert and competing utterance plans and then requiring them to select one plan under time pressure. For example, subjects see and prepare to say one at a time the sequence of word pairs: Ball Doze, Bash Door, Bean Deck, and Bell Dark. The sequence of B—D— items leads subjects to expect another. If, instead, Darn Bore is presented for rapid production, subjects often produce Barn Door erroneously.

The Data. This section presents only as much information as is necessary to motivate the language production models described below. More comprehensive reviews are available in a variety of sources (Dell, 1980; Fromkin, 1980; Garrett, 1980a; Shattuck-Hufnagel, 1979; Stemberger, 1982; see also Cutler, 1982).

Most of the errors made by talkers can be classified as one of the following types: anticipation, perseveration, exchange, substitution, addition, deletion, or shift. The first three error types were illustrated in Examples 1 to 3 for sound errors. Examples 5 to 8 illustrate the remaining error types, also for sound errors.

5. collect them—correct them (substitution)
6. back burner—black burner (addition)
7. painstaking—paintaking (deletion)
8. slept soundly—sept sloundly (shift)

Elements of speech that move in errors most commonly are phonological segments, clusters, morphemes, and words. Most researchers agree that errors involving feature movements are rare (but see Stemberger, 1982).

Errors in which syllables move also occur, but they are rare. However, syllables and other metrical structures are involved in the specification of the conditions in which errors—especially sound errors—will occur, and of the ways in which the errors will manifest themselves. For example, consonants that move in an error almost always preserve their original location either before or after a vowel in a syllable. (In addition, vowels interact only with vowels and consonants with consonants.) So, for instance, Examples 9 and 10 (from Garrett) are common error types. Errors such as Example 11 (from Shattuck-Hufnagel, 1979) are rare.

9. Do you know where I can get a clear piece—clear pliece
10. It happened in the first, second, third, and fourth—thirth and fourth
11. Trees—stree

Similarly, interacting sounds tend to occur in metrically similar environments. That is, for the most part stressed segments interact with stressed segments and unstressed segments with unstressed segments. Finally, interacting segments tend to be members of the same phonemic clause (Garrett, 1980a).

Metrical structures do not appear to play a similar role in word or morpheme errors. It is true that, in some cases, the environments for two exchanging words are metrically quite similar (from Garrett, 1980a):

12. You should see the one I kept pinned to the door of my room—to the room of my door

In other cases, however, they are not similar and indeed, the stress patterns and syllable compositions of interacting words may be quite different (from Garrett, 1980a):

13. I left the cigar in my briefcase—I left the briefcase in my cigar
14. Fancy getting your nose remodeled—Fancy getting your model renosed

Words involved in errors do tend to share stress level; that is, a stressed word does not normally exchange with an unstressed word. However, that observation may be a consequence of the confounding of stress level with classification of words as "content" and "function" words.

A major condition on whole-word exchanges is that the interacting words are members of the same grammatical class. This occurs 85% of the time according to Garrett's corpus. Word exchanges tend to occur over longer stretches of speech than sound exchanges. Moreover, in contrast to sound errors, they need not be phonologically very similar or to occur in similar contexts (but see Dell and Reich, 1981, for a qualification).

Garrett, Shattuck-Hufnagel

Garrett (1975, 1980a,b) proposes a fundamental separation in processing type during language production between a "functional" and a "positional" level.

At the functional level, words are selected and their grammatical roles and phrasal memberships are determined. At this level, the phonetic forms of words are unspecified and hence, if an error occurs, it will be unaffected by the phonetic properties of a word. Errors will tend to be substitutions of words or word blends, for example, or exchanges, anticipations, or perseverations of words in which the interactions are among words of the same grammatical form class.

At the positional level, the words selected and organized at the functional level are inserted into a structural frame in which affixes and function words are specified. Stress levels are also part of the structural frame. Errors may occur if lexical items are inserted incorrectly into the frame as in "stranding" errors (from Garrett, 1980b):

15. I went to get a check cashed—cash checked

Alternatively, inflectional affixes may get attached to the wrong content word (from Garrett, 1980b):

16. I'd forgotten about that—I'd forgot abouten that

Notably, errors of both types appear to freely produce ungrammatical forms such as "abouten." The processes appear blind to information about form class or lexicality.

They are not blind to information about phonological realizations of inflections and stems, however. Example 4 reveals that shifts of inflections lead to phonological accommodation.

Sound errors may occur at the positional level. And they may also show accommodation (from Kenstowicz and Kisseberth, 1979):

17. Tail spin ([t^heyl spIn])—pail stin ([p^heyl stIn])

This implies a stage subsequent to the positional level in which the phonetic forms of ordered phonologically specified words are selected. Garrett (1975) includes such a stage in his model followed by one in which instructions to articulators are specified.

An interesting property of Garrett's proposal is its sharp separation of content and frame in sentence production. This separation is an attractive property of the model because it suggests that language production explicitly uses the "great evolutionary discovery" referred to earlier.

This theme is elaborated by Shattuck-Hufnagel (1979). Somewhat in contrast to Garrett's approach, she emphasizes the similarities in the error

types characteristic of different linguistic units, especially phonological segments, morphemes, and words. She suggests that because they all participate in the same five basic error types (exchanges, substitutions, additions, omissions, and shifts), "the most parsimonious model...is one in which a single underlying mechanism accounts for all error types across all [types of linguistic units that participate in the error types]" (p. 303). The reason the errors may manifest themselves in somewhat different ways across the different linguistic units—for example, the reason sound errors have phonologically similar source and destination contexts, but word errors do not—is because the mechanism operates at different *levels* of the sentence representation when words and sounds are being ordered and hence it utilizes different information.

The mechanism Shattuck-Hufnagel has in mind incorporates a version of the distinction between functional and positional representations as proposed by Garrett. She lists five linguistic units that participate in the basic error types. They are features, phonological segments, morphemes, words, and sentence constituents. However, apparently, she does not intend that five levels of representation be assumed independently to undergo the serial ordering processes leading to error. In particular, there is some indication (Shattuck-Hufnagel, 1979, p. 313) that, in her view, the morpheme and word levels are not distinct. In any case, at each relevant level of representation, two independent representations of the utterance-to-be are generated. One includes the "content"—that is, the selected linguistic units—the other includes the frame into which they are to be inserted. At the word level, the frame corresponds more or less to Garrett's positional level. At the phonological level, the frame is provided by sequences of canonical syllable structures into which target phonological segments will be inserted.

Insertion of content into frame is achieved by a "scan copier" that selects each appropriate target unit from the content representation and inserted it into its slot in the frame. Two monitors watch over this process. One checks off target segments once they have been selected to prevent their reselection later to fit a similar slot. A second scans the final product for errors.

Justification for separation of content from frame derives most convincingly from exchange errors such as Example 3 above. In these errors, quite remarkably, not only does a segment (/p/ in the example) appear early in a slot similar to its intended slot, but, in addition, the replaced segment (/f/ in the example) shows up in the slot left "empty" by migration of the first segment. In this way, the structure of the utterance is unaffected by the error. Two segments simply changed places. For this to occur requires that the structure of the utterance be, in some way, detachable from the particular segments that realize it.

Shattuck-Hufnagel justifies the two monitors based on occurrence and nonoccurrence of various error types. First, exchange errors suggest a monitor that marks the anticipated segment (the /p/ in Example 3), as already selected, thereby preventing it from occupying its intended slot. If the monitor fails, the error is an anticipation.

The second proposed monitor scans the result of the scan-copy process looking for errors. It may weed out phonotactically improper sequences (for example, word-initial /tl/), or sometimes it may wrongly weed out sequences that look like the kind of error the scan-copy mechanism would make. For example (from Shattuck-Hufnagel, 1979),

18. That would be behaving—That would be having

Here an apparent perseveration may have been "corrected" by an error monitor.

Activation Models

The speech error literature offers a different kind of model from those of Garrett and Shattuck-Hufnagel. Versions offered recently by Dell and Reich (Dell 1980; Dell and Reich, 1980, 1981) and by Stemberger (1982), differ in one (Stemberger) or two (Dell) major ways from the proposals of Garrett and Shattuck-Hufnagel. First, they adopt an old, but currently popular, associationist approach to characterizing mental events. In these modern forms of association theory, a knowledge system (a lexicon, for example), is instantiated as a network of associated "nodes" (see also MacKay [1982], whose model is depicted in part, in Figure 6-1*a*). In a lexicon, a node may be a concept, a word, a phoneme, or a letter (see, for example, McClelland and Rumelhart, 1981). A node may be activated in comprehension if it is activated directly or indirectly by stimulus input. It may be activated in production if it is associated directly or indirectly with concepts the talker intends to convey. Activation spreads from a "primed" node to any others to which it is associated, and the selection of a word as present in stimulus input or as one to be uttered is based on the relative activation levels of word nodes. "Spreading activation" models instantiate a small number of very general and simple processing assumptions. In relation to the processes discussed by Garrett or Shattuck-Hufnagel (for example, Shattuck-Hufnagel's error monitor), the fundamental processes in these models are low level, lack intelligence, and are very powerful.

Dell's model differs from those of Garrett and Shattuck-Hufnagel not only in respect to its psychological processing assumptions, but also in respect to the linguistic theory that it instantiates. Dell noted the compatibility between the spreading-activation models being developed in

psychology and relational-network theories of language (Lamb, 1966; Lockwood, 1972; Reich, 1970). In Sampson's view (1980), relational-network theory is the "most interesting radical alternative on the contemporary linguistic scene to Chomsky's theory of language" (p. 167).

In a review of the theory, Sampson offers the serious criticism that it apparently cannot handle structure-dependent syntactic processes (that is, processes, such as those involved in relative-clause formation, that depend on the whole structure of a sentence). However, in his view, it is "much more plausible than its rivals as a model of how speakers and hearers actually operate" (p. 177). This is because, in the theory, the speaker-hearer's linguistic competence *is* the means by which he or she both produces and understands sentences. This contrasts favorably with Chomsky's theory in which a grammar enumerates sentences of the language and cannot serve by itself as the means by which sentences are either said or understood. Rather, the grammar is held to be a component in a performance model, whose implementation by performance mechanisms is unspecified.

In relational-network theory, linguistic knowledge consists of two tiered components: a realization network and a set of tactic patterns. The realization network realizes the units of the language—its concepts, words, phonemes, and so forth—essentially as convergences of relations. Figure 6–2*a*, borrowed from Sampson (1980), makes clear this aspect of the theory. The word "under" is nothing other or more than the convergence of a set of concepts (including "lower than"), an ordered sequence of phonemes, and a form class membership enforced by the tactic pattern (see below). Figure 6–2*b* is meant to show that the label "under" for the word node is redundant with the set of relationships and need not be represented in the model; the convergence of relations *is* the word.

The second component is a network of tiered tactic patterns. The tactic patterns express the internal relations among units at a particular level of the language. For example, the pattern in Figure 6–3 (also from Sampson, 1980) is counterpart to the phrase-structure rules of a generative grammar; it is part of a system that expresses possible sequences of words in a sentence.

Units in the tactic patterns are connected to their counterparts in the realization network. The realization network, then, captures what the linguistic units *are* (content) and the tactics, the structures in which they can participate (frame).

Let us consider how Dell's model marries spreading-activation models and relational-network theory to create a model of language production that makes natural errors.

Figure 6–2. Fragments of a realization network adapted from Sampson, 1980.

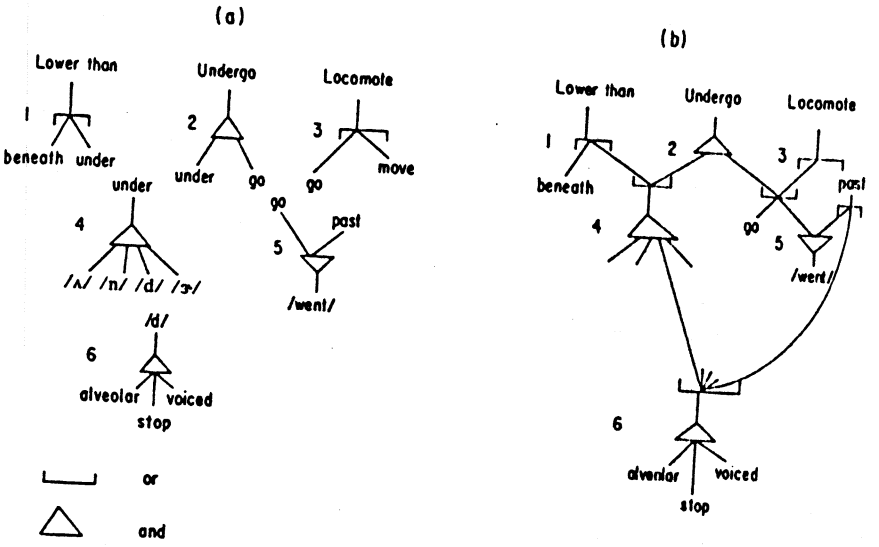
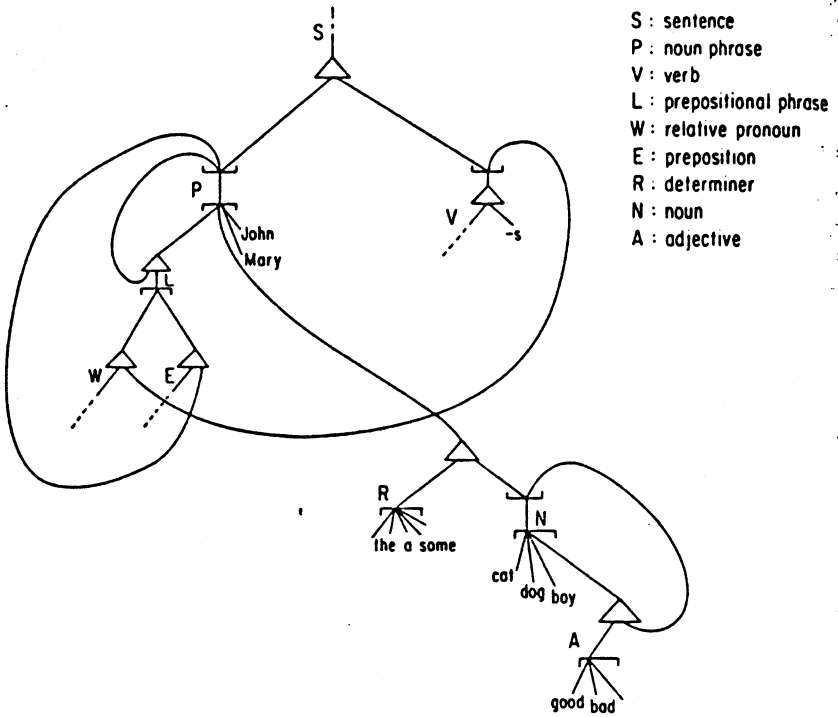


Figure 6–4 (from Dell, 1980) depicts a piece of a realization network. In the network, words are convergences of relations between concepts, phonemes, and tactic patterns. An unusual property of the network is the coding of phonological segments by syllable position. That is, in the model, word-final /t/ in "cat" is a different segment from word-initial /t/ in "tan." As we will see, this property of the network ensures that slipping sounds will retain their intended position within a syllable.

Dell is primarily interested in the patterning of sound errors, and hence, has developed his model largely with reference to utterance realization at the sound level. Sounds are selected for utterance when they are activated in the realization network and selected by the tactic pattern at the sound level, called the "phonotactic selection mechanism."

Dell has realized his model of production as a computer simulation. In Dell's model, a word to be uttered receives some activation; words following it are activated somewhat less. Some proportion of the initial activation then spreads from the word nodes to the concept and phonological segment nodes to which the words are related in the network. In turn, these newly activated nodes send some proportion of their activation to nodes to which they are connected, including the originally activated words themselves.

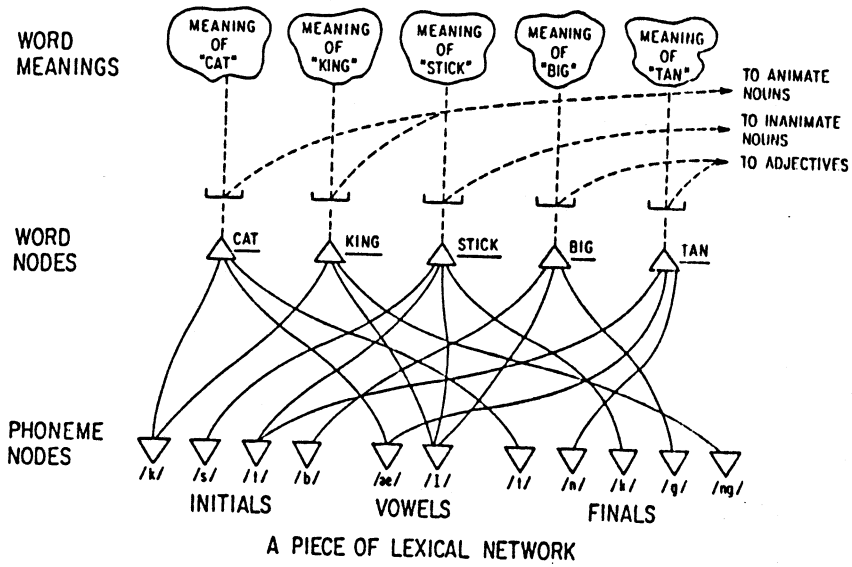
Figure 6-3. A fragment of a tactic pattern adapted from Sampson, 1980.



In this way, activation spreads unevenly through the lexicon, with positive feedback relationships being established between nodes of the originally activated words. Periodically, activation decays to some proportion of its current level. This prevents the eventual activation of every node in the lexicon.

The phonotactic selection mechanism allows activation to spread for some period of time. Then it selects the most highly activated of the initial consonants, or it selects more than one highly activated initial consonant if the consonants are phonotactically compatible. The phonotactics order the segments if more than one is selected. For example, if both /t/ and /s/ are activated in prevocalic position, the phonotactics prescribe the ordering /st/ if the language is English. Next, the phonotactic selection mechanism selects the most highly activated vowel and, last, the most highly activated final consonant or consonants. Once a segment has been selected, its activation level in the network is set to zero. However, because its word node is still activated, it is gradually reactivated.

Figure 6-4. A realization network from Dell's (1980) model of language production.



An error occurs if a target segment is not the most highly activated at the time at which selection occurs. This may occur for a variety of reasons. Consider, for example, the error in Example 1 noted earlier. When "lined" is activated, "lane" will get a lesser degree of activation at the same time. "Lined" will activate its component phonemes, two of which, syllable-initial /l/ and postvocalic /n/, are shared with "lane." These phonemes, substantially activated by "lined," will send activation both to "lined" and to "lane." In turn, "lane" will activate its component phonemes. Because of the similarity between "lined" and "lane," as the one grows in activation level, so does the other. Therefore, /ey/ in "lane" will receive more activation than it would in the context of a word less similar to "lane" than is "lined." The existence of words in the lexicon with word-initial /ley/ (for example, "late") or with the rhyme /eyn/ ("mane") will be activated by the corresponding phonemes in "lined" and "lane" and will further boost the activation level of "lane." This pattern of activation, possibly enhanced by noise in the system that boosts the activation level of words at random, may cause /ey/ to be more highly activated than /ay/ when the phonotactic selection mechanism chooses the syllable's vowel. If /ey/ is erroneously selected, its activation level will be set to zero, thereby promoting the

occurrence of an exchange error. However, /ey/ will continue to be activated because postvocalic /n/ is activated and because the syntactic selection mechanism will select "lane" as the next word to be said. Whether /ay/ or /ey/ is selected for the vowel in "lane" depends on their relative activation levels when the phonotactic selection device chooses a vowel for "lane." If /ay/ is selected, the error is an exchange; if /ey/ is selected, it is an anticipation.

Dell's model reproduces many of the salient characteristics of sound errors. It produces anticipations, perseverations, and exchanges. The erroneously produced segments preserve their position in the syllable because they are coded for syllable position in the lexicon and the phonotactics select only among segments in the appropriate syllable position. Finally, errors are more likely when the source and destination contexts are similar. The simulation not only produces these error types, but, with appropriate settings of the parameters, it produces them in the relative proportions that they occur in spontaneous-error corpora.

Additionally, the model allows a number of novel predictions, some of which have been tested and confirmed. For example, it predicts a lexical bias in sound errors—that is, a tendency for sound errors to produce real words in the language. In the model, the lexical bias occurs because of the positive feedback relationship that is established between word and segment nodes. If a sequence of phonological segments constitutes a word, then the word will be activated by all of the segments and in turn will activate all of them. A sequence of segments that does not constitute a word has no superordinate node to reinforce it and to be itself reinforced by the sequence. A lexical bias occurs in Dell's simulation. It also occurs in experimentally elicited errors (Baars, 1980) and in corpora of spontaneously produced errors (Dell, 1980; Dell and Reich, 1981). Although other investigators (Fromkin 1973; Garrett 1976) have commented on the large number of nonwords generated by sound errors, analyses by Dell of their corpora as well as his own reveal that words are generated disproportionately. This is not expected in Garrett's model unless it is supplemented by an explicit error monitor, similar to one proposed by Shattuck-Hufnagel, which turns nonwords in the speech plan to words. An elegant feature of Dell's model is that it provides the monitoring function without a homunculus.

In common with Wickelgren's proposals (1969, 1976), Dell's simulation permits the prediction that phonemes shared by neighboring words in a planned utterance will promote errors. In Wickelgren's view, this "repeated phonemes" effect occurs because segments are context sensitive. That is, the /ey/ in "lane" is represented as $_{l}ey_{n}$ and /l/ and /n/ as $\#l_{ey}$ and $_{ey}n\#$. Errors occur when a serial ordering mechanism attaches a segment to a context that is compatible or partially compatible with its context specifications, but is not the intended context. This serial ordering procedure implies that only segments adjacent to repeated phonemes will slip. Dell's model, however, implies that segments repeated across neighboring words will promote slippage by any other segments in either word. Dell confirmed this prediction in an analysis of spontaneous errors.

Dell has not simulated selection of words by syntactic processes. However, he points out one counterintuitive prediction that such a simulation would provide. Because of the nature of the lexical realization network, not only will sound errors be lexically biased, but in addition word errors will be disproportionately phonologically similar. Phonological similarity between target and error has been noted for one type of word error—namely, malapropisms (e.g., "equivocal" for "equivalent" from Fay and Cutler, 1977), and these have been ascribed to selection errors from a lexicon organized by sound rather than meaning (Fay and Cutler, 1977). However, Dell's model allows the prediction of disproportionate phonological similarity in all word errors including blends (e.g., "class," "course"—"clourse"), so-called meaning errors (e.g., "knee" for "elbow") and even word misorderings within a sentence. Dell has confirmed this in an analysis of word errors in his spontaneous-error corpus.

Commentary

Speech errors probably constitute the richest source of evidence available concerning language production. Not the least of their advantages is that they can be studied both "in the world" and in the laboratory.

Perhaps because the error data are abundant and, in some respects, clear in their patterning, some conclusions can be drawn from the patterns with confidence. These conclusions are common largely to the two model types just described. Both model types capture the distinction between form and function that lies at the heart of language. Similarly, both treat the units of the linguistic message as distinct from their carriers, the syllables of the language. Finally, the models explicitly realize language's duality of patterning by keeping the different levels of structure and the different ordering processes separate.

Both model types also treat the tiered units of a linguistic message as hierarchically organized. Moreover, in these models, speech planning consists of temporally successive phases which respect the hierarchy. Large units are selected and ordered before their realizations as smaller units are determined. The evidence for this notion of speech planning over time is compelling, too. When a talker produces errors such as those in Examples

19 and 20 (from Garrett [1980b]), it is difficult to avoid concluding that two ordered events conspired to generate the error:

19. I don't know that I'd know one if I heard it—I'd hear one if I knew it
20. Even the best teams lost—Even the best team losts.
/s/

In Example 19, the two words, both verbs, are exchanged. One, "hear," apparently has stranded its tense marker. That marker, attached to "know," is not realized phonologically as /d/ as it would have been in the intended utterance. Rather, remarkably, it is realized as the correct, irregular, past tense of "know." The two ordered events, then, are misordering of words of the same form class and then, *later,* selection of the phonological form of the past tense morpheme.

In Example 20 (see also Example 4 earlier), a morpheme shifts and, in shifting, acquires a new phonological realization. This shift implies two ordered events also, one of which, however, is distinct from the stages inferred in Example 19. When words shift across clauses, as in Example 19, the exchanging words generally share form class. Therefore, when affixes are stranded, they are reattached to a word that can take that affix in some form. In Example 20, however, a bound plural morpheme shifts from a noun to a verb. Verbs do not take the plural {-S} morpheme, and in any case, "lost" can never take any affix realized as an /s/. Despite that, the affix does undergo the voicing assimilation characteristic of the plural morpheme. That is, the plural morpheme, realized as /z/ in "teams" became /s/ in "lost." Two ordered processes are implied here, too, then, and only the second of them may coincide with the two interred from Example 19. In Example 20, a morpheme shifts as if it were blind to form class, and then a phonological voicing assimilation process occurs. One could even argue for four ordered stages based on the errors in Examples 19 and 20. The plural morpheme could not be realized "correctly" as /s/ until "lose" + past (see "know" + past = "knew" in Example 19) became "lost." Hence four ordered events in Examples 19 and 20 are as follows: functional-level word exchange, as in Example 19; past tense realization—still sensitive to lexicality—as in Example 19; affix shift, blind to lexicality and form class, as in Example 20; voicing assimilation of plural morpheme as in Example 20. These kinds of errors strongly imply that speech planning involves some planning events that feed others.

Other properties of the models are less compelling and less attractive. Although appropriately they are models of language production and not of speech errors, nonetheless they have some properties that seem motivated only by the requirement that error patterns be reproduced. Two examples stand out. First, syllables have no role in the models. It is true

that Shattuck-Hufnagel has the ordered phonemes of a lexical form read into canonical syllable slots in an output buffer. However, it is not apparent what function the syllable structures serve except to ensure that sound errors will preserve their syllable position. For its part, Dell's model has no syllables at all, but only segments in the lexicon marked for syllable position. This approach is likely to be incorrect for a variety of reasons. First, it fails to motivate the syllabic coding of segments in the first place. Second, it suggests no closer connection for a language user between the /t/ in "tap" and that in "pat" than between the /t/ in "tap" and the /l/ in "pal." Yet alphabetical writing systems in which both versions of /t/ are written with the same letter suggest a close connection. Perhaps from the perspective of the language researcher, it is the speech production theorist's job to motivate the role of syllable structures in speech.

A second feature of both models also seems unmotivated except as a means to generate appropriate error patterns. In both models, words—which are, in part, ordered sequences of segments—are selected and ordered at one stage of the planning process. Even though this should, by implication, order the segments to be produced, in both models a later stage selects and orders the segments of each selected word. As Shattuck-Hufnagel points out, this is difficult to motivate, but it seems necessary to generate sound-ordering errors.

Such an ordered sequence of selections may seem more natural in a spreading activation network in which it takes time to get from a word node to a segment node. However, it would seem that the naturalness is spurious. In relational-network theory a word is nothing other than a convergence of relations between phonological form, form class membership, and semantic usages in the language. The word "node" is just the point of convergence; it is not a thing in itself. Therefore, it still makes little sense to select a word and *then* its component phonemes; the word is, in part, its phonemes. Conceivably, the ordering is not in respect to selection of units of various sizes, but rather in respect to the talker's attention to units of various sizes.

Duration, Pausing, Coarticulation, f0: Introduction

The literature on errors supports some fairly strong conclusions concerning language production. In particular, it supports a view of language as a tiered structure, apparently constructed over time by sequences of processes each sensitive to linguistic structures at the level on which it is working but relatively blind to those on other tiers.

In this next section, additional theoretical proposals and models of language production are reviewed. These new proposals have been based

on a collection of measures other than speech errors. The measures—pausing, segment and word duration, coarticulation and its blocking, and fundamental frequency—provide mutually redundant information and hence are grouped in the review.

For a variety of reasons, proposals concerning language production based on the patterning of these measures are difficult to integrate with those based on speech errors. For one thing, they are proposals of different types than those based on speech errors. Where researchers in the field of speech errors have built models of fallible language production rather than of speech errors per se, researchers studying duration, pausing, and the other measures under review here have modeled the measures themselves. To the extent that they succeed, the models (called "algorithms" by their creators) generate the natural patterning of the measures in an utterance. However, they provide little insight into the means of their generation (or, for that matter, of the wherefores of their generation).

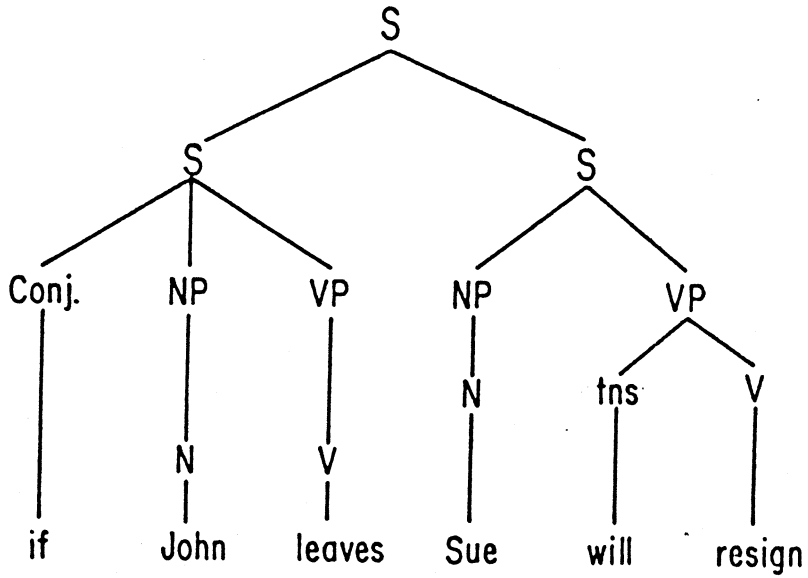
A second difficulty in integrating the findings and proposals based on these new measures with those from the literature on speech errors is that the findings appear to disagree at least with the tenor of those latter proposals. In contrast to the models of Garrett and Dell, in which phonological processes are relatively blind to the syntactic structure of an utterance, the algorithms that generate pausing, duration patterns, and so on, are strongly sensitive to it. Yet these are measures of language *performance* and presumably reflect processes occurring no earlier than those in which the phonological structure of a sentence is realized.

First an overview of the findings based on these measures is provided, then how they may be viewed in relation to theorizing based on speech errors and on the latency measures, which are reviewed next, are discussed.

Duration, Pausing and Coarticulation Blocking: Cooper and Paccia-Cooper and Gee and Grosjean

The basic findings from studies of language production are straightforward (Cooper and Paccia-Cooper, 1980; Gee and Grosjean, 1983). Other things being equal, the duration of a word (or of its final segments), the probability of pausing after a word, the duration of any pause that may occur, and the probability that perseveratory coarticulation will be blocked across the word boundary are all correlated with the "strength" of the boundary following the word. In Cooper and Paccia-Cooper's view, boundary strength is determined by syntactic structure. In Figure 6-5, "leaves" and "Sue" border a major boundary; "will" and "resign" flank a minor boundary. Hence "leaves" is more likely to exhibit

Figure 6-5. A tree-structure representation of a sentence, indicating differences among interword boundaries in "boundary strength" (adapted from Cooper & Paccia-Cooper, 1980).



word-final lengthening, a following pause, and blockage of perseveratory coarticulation than is "will."

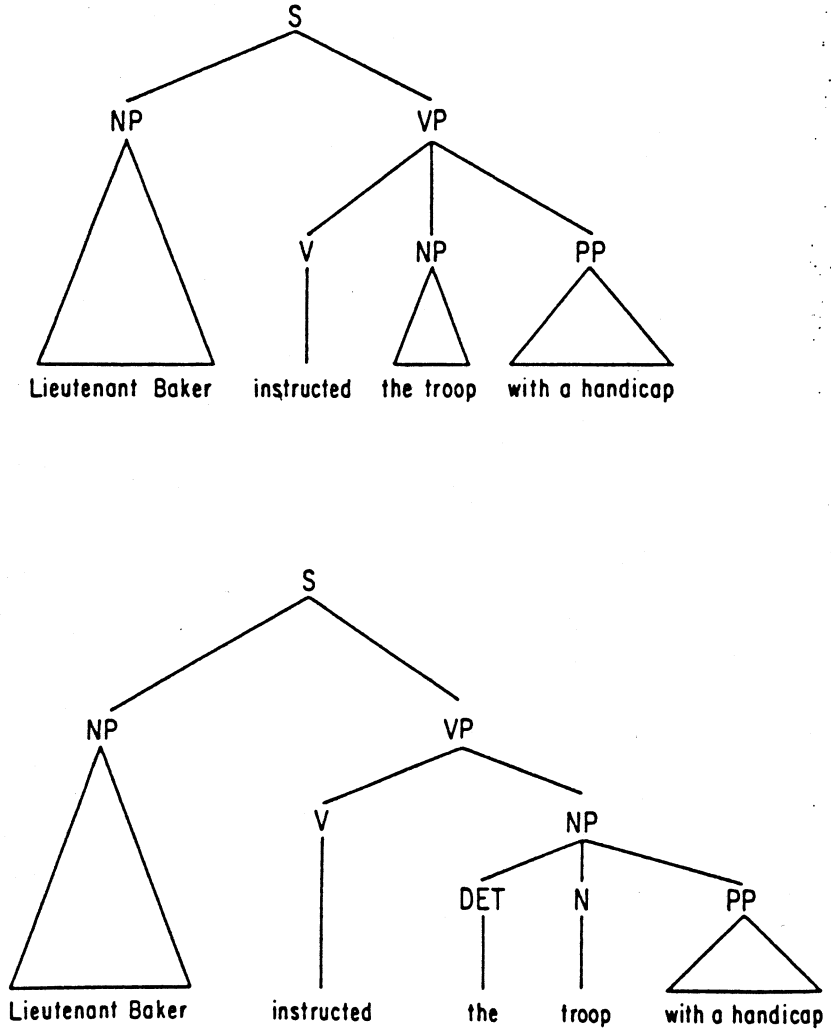
Measures of pausing and of durational lengthening appear to be quite sensitive to syntactic structure, distinguishing, for example, the two readings of Example 23, at least among readers who are aware of the ambiguity:

23. Lieutenant Baker instructed the troop with a handicap.

The durations of "troop" and of the pause following it are both longer if Lieutenant Baker has a handicap than if the troop has one. This difference between the sentences in duration and pausing mirrors a difference in their syntactic structures, as revealed in Figure 6-6 (from Cooper and Paccia-Cooper, 1980). When the troop is handicapped, "with a handicap" bears a closer syntactic relationship to "troop" than if Lieutenant Baker has the handicap.

In long sentences, the patterning of durational differences and of pauses becomes extremely rich, yet still predictable. Table 6-1 gives most of the steps in an algorithm proposed by Cooper and Paccia-Cooper to reproduce the lengthening and pause structure of sentences.

Figure 6–6. Two readings of an ambiguous sentence represented as differences in phrase-structure representation (adapted from Cooper & Paccia-Cooper, 1980).



Cooper and Paccia-Cooper point out that they devised the algorithm to generate the patterning of the dependent measures in their experiments and not to mimic, necessarily, the means of its generation by talkers. Indeed, the algorithm has several properties that eliminate it as a candidate

Table 6–1.
Algorithm for Determining Pause Durations.

1. For each critical boundary in a tree-structure representation of a sentence, locate the dominating phrase-structure nodes. These are the highest nodes that do not dominate both the words flanking the critical boundary, but that dominate one or the other.
2. Count the nodes between the nodes just identified and those immediately dominating the words flanking the critical boundary. Discount nodes referring to conjunctions, determiners, and nonlexical prepositions and, on the left side of the boundary, any nonbranching, nonterminal nodes.
3. Increment the count for any branching S node.
4. Multiply by 2 the count for the left side of the boundary.
5. Sum the counts for the two sides of the boundary.
6. Bisection (from Grosjean et al., 1979). Count the number of major grammatical words in the largest constituent being analyzed. Divide by 2. Count the number of major category words from either end of the constituent to the boundary (whichever is less). Divide by the bisection point.
7. Multiply the outputs of steps 5 and 6 for each word boundary. The largest product identifies the major constituent break. Any such boundary retains its product.
8. If the two constituents demarcated by the major break found in the last step contain more than seven major category words, repeat steps 6 and 7.
9. Constituent length—if the largest constituent boundary precedes or follows a constituent including seven or more words belonging to major grammatical classes, add or subtract a percentage amount to the output of the last step for that boundary.
10. Other steps adjust for speaking rate or other prosodic effects.

Adapted from Cooper, W., and Pascia-Cooper, J. (1980). *Speech and syntax.* Cambridge, MA: Harvard University Press.

psychological model of language production. First, the tree structure for an entire sentence is involved in computing measures of word duration, probability of pausing and pause duration, and probability that coarticulation will be blocked at a boundary. Yet is highly implausible that this amount of detail is available before the first word of a sentence is uttered. Second, the algorithm itself is complicated enough to render it implausible that talkers would run through all of the steps before producing an utterance. Third, the algorithm implies that word duration and segment duration are explicitly computed. Yet a large part of the complexity of the algorithm that computes these durations concerns reproducing the complex microstructure of the durational patterns the talker's rationale for which, if any, is obscure. In particular, the full complexity of the durational patterns probably is unappreciated by listeners. Listeners may well hear and use the durational lengthening and pausing at major boundaries, but they are not likely to use, or even to hear, many

of the finer aspects of the timing pattern. Nor do the steps in the algorithm typically have an obvious rationale in terms of production. For example, why are left nodes of the tree multiplied by 2? Why do sentences have a bisection point? In short, it is not clear why talkers would bother with the steps of the algorithm if reproducing the microstructure of the durational pattern were an end in itself as the algorithm suggests.

These considerations make clear that, although the algorithm may work to generate the timing pattern of a sentence (in fact, however, it accounts for only 56% of the variance in sentence pauses examined by Gee and Grosjean, 1983), at best it fails to reveal the source and rationale for the pattern, and, at worst, it may obscure it.

Gee and Grosjean (1983) propose a new model designed to overcome the shortcomings of the Cooper and Paccia-Cooper model and similar shortcomings of an earlier model by Grosjean, Grosjean, and Lane (1979). They suggest that two general factors govern the durational structure of a sentence: its syntactic structure and the distinction between content and function words. The latter distinction they see as having two relevent aspects: the different information contents of the two types of words and their different stress levels in production. They suggest, then, that syntactic structure, information load, and stress level all contribute to the durational pattern of a sentence.

To generate the durational patterns of sentences in ways sensitive to the syntactic, informational, and stress patterns of the language, Gee and Grosjean borrowed from the growing literature on prosodic (or metrical) structures in language (Liberman and Prince, 1977; Selkirk, 1980a, b). This literature attempts to characterize and reveal the metrical patterns of language—in particular, the patternings of strong and weak segments and syllables in spoken language. The literature provides perhaps the strongest evidence that sentences have metrical structures—syllables, feet, phonological phrases, and intonation phrases—that are distinct from syntactic units, but are sensitive to them.

Gee and Grosjean use the metrical structures with the largest domain—phonological and intonational phrases—to generate durational patterns in fluent speech. Their algorithm moves left to right through an utterance's surface structure, and determines durational patterns of early sentence constitutents in the absence of a fully specified surface structure for the sentence. In the algorithm, words up to and including the head of syntactic phrase constitue a phonological phrase and are organized together metrically (see also Selkirk, 1980b). Consequently, pausing boundaries within a phonological phrase will be minor boundaries; those across a phonological phrase are more important. Phonological phrases themselves are organized into intonational phrases subject to a few contraints. Two phonological phrases subsumed by the same syntactic phrase (excepting VPs) are organized into a common intonational phrase. A phonological phrase headed by a verb is organized with the following phonological phrase unless the preceding one is less complex. Remaining phonological and intonation phrases are organized left to right in the sentence.

With certain adjustments (for example, nonlexical weak monosyllables are not counted as being followed by a boundary, complex words are flanked by relatively longer pauses than their metrical location would suggest, and sentence-final words are lengthened) this new algorithm can account for 92% of the variance in the pausing data collected by Grosjean. Moreover, it has the advantage over earlier models that it works left to right.

Commentary

There remain three major difficulties with this account of durational lengthening and pausing. It does not explain how the metrical structure gets realized as durational lengthening and pausing. The idea that word and phrase durations are computed and assigned is unattractive because the task of computing and realizing the values would be a big job with no apparent purpose. A more attractive proposal, in my view, would be one in which the metrical structure characterizes some relatively "low-level" organization of the talker's vocal tract achieved to realize the surface structure of the sentence in speech. But we are still far from understanding how such a motor organization would allow or facilitate utterance production.

A related difficulty is that the proposal provides no real insight into why the surface structure of a sentence is realized as a metrical structure. According to Selkirk (1980b):

> [T]his review of prosodic structure has shown quite clearly that prosodic structure is not syntactic structure, nor is it isomorphic to it. The two are quite distinct as formal objects. A mapping from one to the other can, and must, be defined, however, for the prosodic structure reflects syntactic structure in certain ways. We would suggest that the mapping from syntactic to phonological representation of a generative grammar is precisely this mapping—the mapping between syntactic and prosodic structure.
>
> Thinking now in terms of speech production and perception, we would hypothesize that the units of prosodic structure we have discussed here in linguistic terms are indeed the appropriate units in production and perception models, that the effect of syntactic phrasing in reproduction, or access to that phrasing in perception, are crucially mediated by these units of the prosodic hierarchy. (p.29)

But this is not entirely satisfactory. Why cannot a surface structure be produced without introducing metrical structure? Again, perhaps, the rationale may have to be provided by speech production theories.

A final difficulty with the algorithm of durational lengthening and pausing proposed by Gee and Grosjean has been alluded to earlier. It concerns any attempt to integrate the model with those of language production motivated by analyses of speech errors. Why, if the metrical structure is part of the phonological representation of an utterance, as it is presumed to be (for example, in the quotation from Selkirk just cited), is it sensitive to the syntactic structure of a sentence, whereas errors at the phonological level of utterance production (and even "earlier" at Garrett's positional level) are not?

The problem probably cannot be resolved by supposing that the metrical structure is not part of the phonological system and that instead sentence constituents are organized into larger metrical structures early in sentence production—say, at Garrett's functional level of processing. As observed earlier, word exchanges often are not metrically similar. (That is, they differ in number of syllables, as in Example 14, or in stress pattern, as in Example 13.) A careful analysis, analogous to Dell's , revealing lexical bias in sound errors and phonological bias in word errors would possibly reveal a metrical bias in word errors. However, Dell would not expect this because it would signify that metrical structures are embedded somehow in the realization network. Yet if the apparent absence of metrical effects at the word level localizes metrical structures within the phonological phase of sentence production, we are left without an understanding of why they, and not sound errors, are so sensitive to syntactic structure.

Fundamental Frequency

The fundamental frequency (f_0) contour in a speech utterance constitutes a rich source of information for a listener. It provides information about the sex, age, height, and weight (e.g., Lass and Davis, 1976) of a talker and about his or her emotional state (Sherer, 1981, 1982; Tarttar, 1980; Williams and Stevens, 1972). In addition, it provides at least two sources of linguistic information. The global intonation contour distinguishes questions, statements, commands, and more one from the other. In addition, the intonation contour is superimposed on a gradual downdrift in fundamental frequency ("declination") that extends over the course of a major syntactic unit (often a sentence). The downdrift in f_0, then, delimits major syntactic units in an utterance for a sensitive listener.

A theory of language production will have to explain how talkers provide all of the foregoing information in the f_0 contour that is controlled.

So far, however, only the last source of information—f_0 downdrift—has been studied with a view to explaining its regulation. Consequently only declination will be covered in this review.

Declination. Utterances show f_0 peaks and troughs. The peaks correspond largely to prominent syllables in the utterance, and the troughs to less prominent syllables. A curve drawn peak to peak (the "topline" of the declination curve) tends to be negatively accelerated, but linear on a log-linear plot of frequency and time (Cohen, Collier, and t'Hart, 1982). One drawn trough to trough (the "bottomline") is linear with a negative slope. Declination refers to either or both downdrifting tendencies.

Languages typically (but perhaps not universally; see Cooper and Sorenson, 1981, for a brief review) exhibit downdrift. The most studied languages, however, are Dutch (Cohen et al., 1982, and references therein), English (Breckenridge, 1977; Cooper and Sorenson, 1981) and Swedish, (Garding, 1979). In these languages, the fundamental frequency at sentence offset is nearly invariant, while the starting frequency may (Cooper and Sorenson, 1981; Cohen et al., 1982) or may not (Maeda, 1976; Breckenridge, 1977) covary with sentence duration. In any case, the slope of the declination lines decrease with an increase in sentence duration.

Although some investigators have proposed physiological accounts of declination (see Breckenridge for a review), the most comprehensive investigation to date on declination in English (Cooper and Sorenson, 1981) follows conclusions by Breckenridge that the declination topline does not automatically "fall out" of the respiratory and laryngeal events underlying sentence production; rather, it is "programmed" by a talker. Sorenson and Cooper (1980; see also Cooper and Sorenson, 1981) model the declination topeline mathematically by fitting a line segment to all f_0 peaks except the first, which lies above the fitted line. The line has the folllowing form (where P_i is the ith f_0 peak; T_i is its time of occurrence relative to sentence onset, and P_n and T_n refer to the final peak and its time of occurrence, respectively):

$$F0 = P_n + 2/3\,(P_i - P_n/T_i - T_n) \bullet (t - T_n)$$

Using this equation for a line, the "topline rule," an investigator or talker can determine f_0 for a peak occurring at any time "t" in the utterance. The declination curve has as its domain a sentence, or, if it is sufficiently long, a major clause.

Sorenson and Cooper propose that the talker uses the topline rule in the following way:

> How does the speaker program an f_0 declination in fluent speech?...At the beginning of an utterance, the speaker's look-ahead mechanism informs him

> of approximate sentence length, which is somehow used to generate the appropriate f_0 value of the first peak. The approximate value of the last peak is also known to the speaker as evidenced by its constancy across sentences of different length.... Once the speaker begins talking, feedback (auditory or otherwise) informs him of the value of the first peak, which together with the value of the last peak and approximate estimated sentence duration can be used to generate the Topline Rule. The speaker then endeavors to produce those peak values of f_0 based on the rule. (p. 419)

The declination line typically is "reset" at the end of a sentence. However, it may be partially reset at the end of a sentence-internal major clause. In addition, finer marking of syntactic boundaries, compatible with the markings by pausing and word lengthening, is achieved by "fall-rise" patterns—that is, a fall in f_0 at the end of a syntactic unit and a rise at the beginning of the next. These fall-rise patterns do not seem to affect the declination lines and therefore are not identified as resetting.

Commentary. The topline rule and Sorenson and Cooper's proposed psychological implementation of it obviously will generate an f_0 contour that drifts downward over the course of an utterance in approximately the same fashion that the topline drifts downward. It is implausible, however, as a psychological procedure (see also Simon, 1980). In particular, it is another example similar to that of the pausing algorithms in which theorists account for aspects of the superficial form of an utterance by proposing a mechanism to reproduce the aspects explicitly. In essence, the account is that speech exhibits declination in the form that it takes because talkers put it there in that form. But it is not always justified to infer that a particular variable is explicitly controlled in a skilled activity just because it has regular properties, and it may not be wise to assume explicit control as a first hypothesis. Some variables exhibiting regular properties are not the kinds of things that actors *can* regulate. That is, some regularities are not regulated at all. They are in a sense "emergent" in the activity (Kugler, Kelso, and Turvey, 1980), or they are byproducts of other things that the actor is doing. On the surface, the declination lines themselves suggest an account of this sort. As already noted, the vast majority of languages show declination. Moreover, across tone languages, falling tones are more frequent than rising tones. Similarly, (untrained) singers can achieve a fall in f_0 more quickly than a rise (Sundberg, 1979). This suggests that downdrifts are relatively easy and natural to achieve—perhaps because speech is produced on an expiratory airflow. As lung volume decreases, other things equal, subglottal pressure will decrease and f_0 will decline.

Why, then, propose a computational model of declination as Sorenson and Cooper do? There are three reasons. First, it has been argued (Breckenridge, 1977) that physiological accounts cannot handle the

magnitude of declination that occurs. These arguments have recently been disputed, however (Cohen et al., 1982). Second, Breckenridge and Sorenson and Cooper appear to assume that declination must be *either* physiologically determined *or* linguistically determined, but cannot have both characteristics. However, declination may be a candidate instance where the form of a linguistic device is explainable in physiological terms (and without reference to linguistic terms), but the deployment of the device, and hence its function, is linguistic.

Alternatively, perhaps even the deployment of declination itself is an automatic more than a "programmed" behavior. In the previous section, research by Gee and Grosjean was described that suggests that speech is packaged for output into metrical rather than syntactic structures. Research has not yet been designed to ask whether these structures rather than syntactic units per se are in fact also the domain of f_0 declination. However, as Cooper and Sorenson point out, a comparison of the pausing data in Cooper and Paccia-Cooper and the f_0 data in Cooper and Sorenson reveals a degree of redundancy in the patterning of major pause boundaries and declination resetting or f_0 fall–rise patterns. This implies that whatever structures, syntactic or metrical, best characterize the packaging of speech involved in generation of pausing patterns, these same structures will characterize that over which f_0 patterns occur. If so, conceivably, declination is not, largely, a contour that a talker programs over a grammatically coherent stretch of speech. It may be a difficult-to-avoid consequence of uttering speech, packaged into metrical structures, and produced on an expiratory airflow.

A third, and perhaps, the main reason why Sorenson and Cooper propose a computational model for declination is that, as a general rule, computational models are the only kinds of models that cognitive psychologists, including psycholinguists, entertain to explain phenomena that they find interesting. The concomitant disinclination to consider accounts whereby the regularities are not explicitly programmed has, possibly, promoted the explosion of variables under review here, which language production theorists, collectively, propose are under programming control.

Latency and Duration

The last model to be considered in this review of language production (Sternberg, Monsell, Knoll, and Wright, 1978; Sternberg, Wright, Knoll, and Monsell, 1980) is distinguished from its predecessors on several counts. First, it is in fact a model of word-string production, not of language

production, in that it makes no attempt to explain how grammatical utterances are constructed. Rather it begins with a planned string of words to be produced and it is concerned with their retrieval from a hypothetical output buffer and their execution. (This model in fact might have been included below in the review of speech production theories because it largely concerns itself with execution of a "motor program." It is included here because its explanatory concepts are in the cognitive–representational domain, in contrast, largely, with explanatory concepts invoked by speech production theorists.) Second, the data underlying the model's construction are obtained from utterances that are not naturally produced. Rather, the utterances are produced with as brief a latency as possible after a signal to begin talking and at as fast a rate as possible. This manipulation is meant to force the characters of the retrieval and execution processes to reveal themselves in the durational measures as the processes operate at their upper limits.

In the procedure used by Sternberg and colleagues (1978, 1980), talkers are given an utterance to say. Across trials, the utterances may differ in length (in number of words) or in complexity (in syllables per word or in heterogeneity among words). They are always well-known word strings; they might be the digits from 1 to 3, for example or the five weekdays, or the word "Monday" repeated four times. The talker knows what he or she is to say well in advance of the cue to respond. The cue, then, only signals *when* the talker is to begin talking. The talker is instructed to say the sequence as quickly as possible following the cue to respond. Sternberg and co-workers measure the latency to begin talking and utterance duration, both as a function of sequence length.

The latency and duration data exhibit regular changes with sequence length. In particular, latency is a linear function of utterance length (in number of words), with a slope between 10 and 15 ms. Utterance duration is an accelerating function of length; that is, as utterance length increases, the average word-to-word interval increases. The latency function loses its regularity at utterance lengths of six or more words. Complexity of the words produced (whether the words are monosyllables or disyllables, for example) affects the intercept of the latency function, but not its slope. Comparable words and nonwords have identical latency and duration functions.

Sternberg and associates propose a model of word retrieval and execution to explain these data. In the model, an output buffer for speech holds the motor program for a to-be-produced utterance. Because words and nonwords generate the same latency and duration functions, the buffer is not presumed to have any special connection to the lexicon of the language. The motor program consists of subprograms, one for each production unit in the utterance. To produce an utterance, each unit is retrieved in turn; following retrieval it is "unpacked" into its constituent articulatory units (perhaps syllables or individual articulatory gestures), and finally a command phase executes the unpacked unit.

Latency to begin talking is presumed to include the retrieval and unpacking times for the first unit in the utterance. (The execution interval is supposed to start with vocal tract movement onset.) Latency increases with utterance length because retrieval of a unit is more time consuming the more items there are in the buffer. (This implies that the first thing to be produced does not occupy a slot in the buffer that is first accessed by the retrieval mechanism.) That the slope of the latency function does not increase when disyllables are produced rather than monosyllables, or even when sequences of stressed and unstressed word pairs are produced, implies that the production unit is not the syllable or the word. Sternberg and associates propose, tentatively, that it is the stress foot. That the intercept is larger for disyllables or word pairs than for (stressed) monosyllables suggests that "unpacking" time does increase with complexity of the production unit.

The duration function is modeled as a quadratic function of the number of word-to-word intervals ($n - 1$, where n is the number of words in the utterance):

$$D_n = a + b(n - 1) + c(n - 1)^2$$

The parameter b, but not c, is affected by word complexity (monosyllable or disyllable); c, the rate at which the duration curve accelerates, has a value very similar to the slope of the latency function. Sternberg and co-workers propose that both c and the latency-function slope reflect retrieval time. If they do, then the value of c supports the conclusion that the production unit is the stress group not the syllable or word because, as noted, it is unaffected by word complexity. In addition it suggests that talkers require more time to produce each word in a long string of words than in a short string because, throughout the string, retrieval time is longer in a large than in a small buffer. Finally, because c has the *same* value as the latency slope, and not half of its value, Sternberg and colleagues conclude that the output buffer does not shrink as stress groups are ouput. They propose, then, that retrieval is a serial, self-terminating search through a nonshrinking buffer.

Sternberg and associates (1980) provide two other kinds of information about utterances produced in their experiment. First, the utterances show declination that is very similar in form to declination described for naturally produced grammatical sentences. In particular, across word strings of length two to five, the final f_0 is invariant. The starting f_0 is also invariant;

consequently, as others have observed with natural speech, the slope of the declination function decreases with increases in utterance complexity.

Second, Sternberg and co-authors (1980) provide some information about the durational microstructure of their utterances. Intervals between word onset and the onset of the second syllable in a disyllable ("within-word" duration) increase with serial position in an utterance, but not with utterance length in words. Intervals between onset of the second syllable in a disyllable and onset of the following word ("across-word" durations) show the opposite pattern. This dissociation is not consistent with the idea that serial position and length effects on duration both reflect the operation of the retrieval mechanism.

Commentary. The model of retrieval and execution accounts well for the data on which it was based, with the exception of the within- and across-word durational patterns. Confronted with other data and other models based on them, a number of disagreements arise. These will be described below under "Puzzles and Inconsistencies." One will be considered here.

An unattractive aspect of the model concerns the relationship between the retrieval mechanism and the output buffer. That the mechanism takes longer to select a stress group for output the more stress groups there are in the buffer suggests an unordered collection of stress groups in the buffer—unordered, at least with respect to the behavior of the retrieval mechanism. This is counterintuitive, at least as a general model for buffering syntactically coherent strings. And it is difficult to believe as an explanation for the durational patterns in such sequences as "Monday Monday Monday Monday," in which the order would not seem to matter. In addition, however, the proposal implies that the stress groups themselves are mutually independent. But stress groups are metrical units that participate, allegedly, in larger metrical units. They are not mutually independent in utterances in which there is more than one of them. That they are not independent is revealed in several ways in naturally produced speech: The relative prominences of their component syllables are affected by participation of their stress feet in larger units; the patterns of pausing are affected and so are word and segment durations (e.g., Lindblom and Rapp, 1973).

It is of some interest that both duration algorithms we have examined—those of Cooper and Paccia-Cooper and of Gee and Grosjean—seem to predict that, other things equal, average durations of pauses will grow with utterance length because on the average, more nodes will dominate words flanking a boundary; therefore, boundary strengths can be larger. Perhaps the increases in latency and across-word duration found by Sternberg and colleagues occur because stress groups participate in superordinate metrical units and not because they are harder to find in the output buffer. To reproduce the durational patterning that Sternberg and colleagues report, the word strings would have to have a structure something like the structure in Figure 6–7, with boundary strength increasing left to right.

Language Production: An Attempt at Evaluation, Integration and Reduction

If we try to integrate all of the foregoing proposals, what does the result look like? In particular, how many and which kinds of variables are talkers presumed to control independently? Under the heading "Integration" an attempt is made to suggest one that is as comprehensive as possible. That integration will suggest some puzzles and some disagreements among the foregoing models that research and theorizing will have to resolve. It will also reveal that theorists have collectively given the talker a gargantuan job of planning and regulation. Under "Puzzles and Inconsistencies" are catalogued some of the problems that the attempt at integration suggests. Under "Streamlining a Theory of Language Production," some ways of reducing the talker's hypothetical problem of control are suggested.

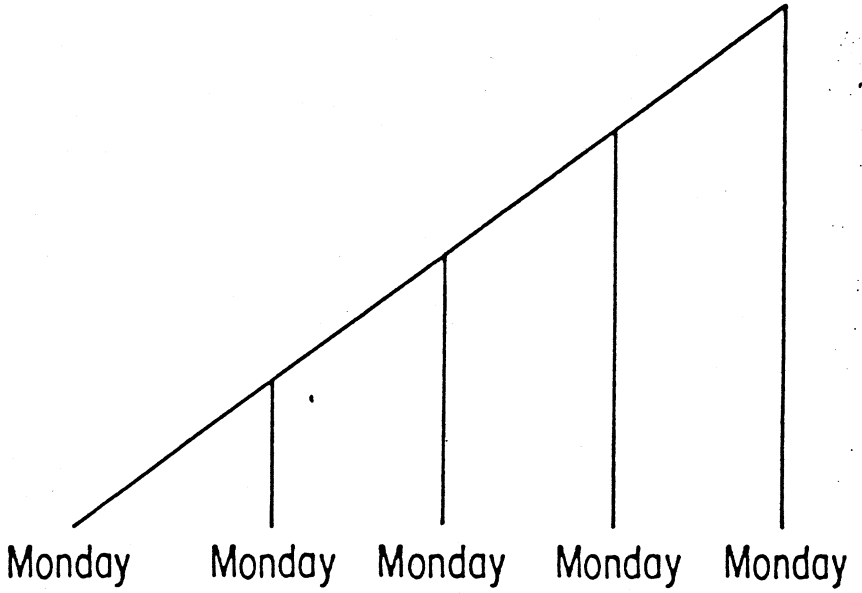
Integration

Having decided what message to convey, the talker is required to construct a surface structure of a sentence. This is accomplished, according to Garrett, in two broad phases. One constructs a functional representation of the message and the other inserts this content into a positional frame. Events at the functional level are (relatively) blind to the phonological forms of words. Consequently, Garrett proposes that phonological processes are applied later in the production process. For their part, phonological processes are (relatively) blind to the syntactic form of the utterance and to lexicality. They are assigned to a third stage of sentence generation. A fourth is involved in speech output.

While this sequence of levels captures many properties of speech errors, it does not capture the fact that the functional phase of language production is only *relatively* blind to phonology and the phonological phase only relatively blind to lexicality. Therefore, an instantiation of the phases more or less along the lines simulated by Dell may be required.

Following phonological selection procedures, the talker has, in an output buffer, a sequence of phonemes arranged in canonical syllable structures (Shattuck-Hufnagel), syllabically coded phonemes (Dell), or an array of stress feet (Sternberg et al.). To output the speech (just having

Figure 6–7. A possible tree structure for "Monday Monday Monday Monday Monday Monday" that would give the pattern of increasing duration reported by Sternberg and colleagues (1978).



placed them *into* the buffer), the talker retrieves them in foot-sized units using a self-terminating, serial search through an apparently unordered, nonshrinking buffer. (Just having packaged them *into* metrical structures), he or she unpacks them, and executes them as speech. Explaining how the symbols of the linguistic representation are translated into activities that a vocal tract can carry out is the responsibility of a speech production theory.

As yet unordered with respect to the foregoing sequence of events and with each other is the regulation of two other sets of variables. At some time (perhaps as speech is uttered; see the quotation from Sorenson and Cooper [1980] above), the talker applies the topline rule. This involves a close attention to the duration of the utterance on a moment-by-moment basis. If an f_0 peak is to be produced (for reasons not covered in this review, but often corresponding to production of a stressed word), according to the topline rule, the talker needs to know when, in seconds after utterance onset, the peak is occurring and how long in seconds the total utterance duration will be. The result of the computation then has to be translated into effects on subglottal pressure or laryngeal tension, in a way, again, to be explained by a speech production theorist.

In addition, at some time, the talker has to code f_0 fall-rise patterns at syntactic boundaries. Presumably, this happens at the same time that the talker codes the utterance's patterns of pausing, durational lengthening, and coarticulation blocking using the algorithm of Cooper and Paccia-Cooper or that of Gee and Grosjean.

To summarize, then, talkers have independent control over variables of at least five different types. They determine the *syntax* of the sentence and which *words* will be fit into the syntactic structure. Having done that, they select and order the *phonological segments* that will realize the words. Some *metrical structures*, in particular syllables and stress feet, are involved somehow in the packaging of words for output, and other metrical structures—in particular, phonological phrases and intonational phrases—are marked, perhaps, by *acoustic* (and *articulatory) variables* including f_0, pausing, and durational lengthening.

If the separate "mental processes" in which a talker is proposed to engage were counted, the result would be far greater than Figure 6–1*a* would suggest. In Garrett's model there are four broad stages, in Sternberg's three, in Cooper and Paccia-Cooper's up to 14 (*per boundary*; there are somewhat fewer in Gee and Grosjean's), and in Cooper and Sorenson's topline rule there is one per peak.

We could, at this point, sit back and marvel at how wonderful the mind is to be able to keep track of all of these things in the course of language production (all of *these* things and then all of those *other* things that are depicted in Fig 6–1*b* and that get the utterance actually uttered). But instead it seems that something must be wrong with this picture because it gives the talker too much to do. Moreover, some of what the talker has to do (for example, computing the durational microstructure of every interword boundary, or computing the precise form of the declination curve) has no apparent purpose. And some of what the talker does is done in ways that appear somewhat perverse (for example, selecting and ordering words and then later selecting and ordering their [already ordered] phonological segments with the result that ordering errors are made; retrieving components of the output buffer with a mechanism—a self-terminating search through a nonshrinking buffer—that appears blind to the ordering of units in the buffer).

Before examining in detail the problems and puzzles to which the present attempt at integration points, we should briefly consider whether the integration accounts for the performance measures specified earlier as the domain of a language production theory and whether it addresses the issues outlined earlier that a model of language production should address.

Obviously, in a fashion, the integrated model generates the performance measures in its domain. It is worth repeating, however, that it does so in two distinct ways. Its components that are responsible for speech errors and its retrieval and unpacking mechanisms are accounts of language production (or word production) that explain how, in the natural course of production, certain errors or durational patterns occur. In contrast, the components of the model responsible for declination, f_0 fall-rise patterns, durational lengthening, and pausing are devices for producing the measures explicitly. An alternative formulation would be one in which, as a byproduct of the talker's planning and execution of speech, certain f_0 and durational patterns arise.

Three aspects of language production were specified earlier that a theory should address. They were the separation of function and form (and of form and substance) in language and speech, generativity, and the role of metrical structures in language. The first two aspects are realized—potentially, anyway—in the modeling supplied by speech-error researchers. The third is not really addressed by the model.

Puzzles and Inconsistencies

The Output Buffer: an Inconsistency. Based on speech errors, Dell and Shattuck-Haufnagel agree that phonemes are, in some sense, organized syllabically before they are produced. Based on latency and duration data, Sternberg and colleagues (1980) propose an organization in terms of stress feet.

In itself, this does not appear to be a major disagreement, because both conclusions could be true at the same time. Moreover, the observation that sounds that are interacting in an error tend to share level of stress may signify that these errors preserve their intended position not only in the syllable, but also in the foot.

However, there is a more subtle, but real difference between the proposal of Sternberg and co-workers and those of Dell and Shattuck-Hufnagel in terms of how the proposed metrical structure, syllable, or foot, is conceived. For Sternberg and colleagues, the foot is a production unit, or, more specifically, a subprogram of the motor program for an utterance. Stress feet are the things that the retrieval mechanism retrieves. For Shattuck-Hufnagel and Dell, however, the units in the output buffer are phonemes; they are embedded in a syllabic frame, but the syllables are not units themselves.

That there is a distinction between "unit" and "frame" (or at least between conventional linguistic units and metrical structures) is strongly implied by ordering errors in speech. Phonemes and words are misordered in speech; syllables and stress feet rarely are. Rather, syllables and perhaps stress feet provide a frame that constrains how sounds can misorder.

In a short-term memory buffer, as it is conventionally studied, misordering errors are common. Two letters, say, may be recalled correctly, but in the wrong order. It would seem, then, that at least some retrieval mechanisms do make ordering errors and that Sternberg and associates should expect misorderings of the units stored in their output buffer—that is, the stress feet.

That misorderings of these units do not occur suggests that stress feet are not units in an output buffer and hence that the latency and duration data require some other interpretaion. Either the production unit is not a foot or the latency and duration functions are not generated by a process of retrieving and unpacking production units. The first option is possible—Sternberg and co-workers drew their conclusion that the unit is the stress foot only tentatively. However, their data are incompatible with the conclusions that the unit is the phonological segment or that it is the word—the two units that do transpose frequently in speech errors. The second alternative—that the latency and duration functions do not reflect the operations of a retrieval process—is also possible. One hypothesis concerning what they might reflect instead has already been suggested. Perhaps sequences to be produced are metrically structured and have structure similar to that in Figure 6–7. The increasing duration and pausing that occur left-to-right in the data of Sternberg and co-authors, then, reflect that structure. (Although the explanation is contrived, applying the algorithm in Table 6–1 to the structure in Figure 6–7 may even generate the latency function—essentially as a 'pause' before the first word in the utterance. It does so because utterance onset has a high boundary strength owing to the large number of nodes under the node dominating the right side of the boundary—that is, the S node. Moreover, as more words are added to the utterance, the stronger the sentence-initial boundary and hence the longer the latency.)

Unpacking: A Puzzle. Sternberg and associates propose that "unpacking" occurs after a stress foot has been retrieved. The basis for their proposal is that the intercept of the latency function, but not the slope, is affected by the complexity of the production unit. In some respects, this proposal has intuitive appeal. The talker does have to generate the full complement of articulatory gestures that the utterance requires, and articulatory gestures have not yet made an appearance in any of the models of language production under consideration.

What is puzzling is that the phonological segments have just been packaged *into* metrical structures for storage in the retrieval buffer. Why

is the structure immediately undone? Put differently, why would a talker have an output buffer requiring a structuring of the language units in a way that they are not structured in early phases of production if the structures have no role to play in articulation?

Durations in Speech: Inconsistencies. Three sources of evidence on durational patterns in speech appear to mutually disagree. Data of Sternberg and colleagues and probably those of Gee and Grosjean and Cooper and Paccia-Cooper suggest a general increase in durations as utterances grow. However, the growth patterns are not alike. This may not be a real inconsistency in view of the fact that Gee and Grosjean and Cooper and Paccia-Cooper ascribe the patterns of durations to syntactic structure (or, in the case of Gee and Grosjean, to syntactic effects on metrical structures). In apparent contrast to these observations, however, are others (see Lehiste, 1980, for a review) in which durations *shrink* as more is said. The bulk of the shrinkage occurs, it is true, as stressed syllables are followed by increasing numbers of unstressed syllables. But shrinkage also occurs as words are added to a sentence. For example, Huggins (1978) provides data showing a 40 ms shortening in the name "Joe" in a five- as compared to a two-word sentence. In the sentences, "Joe took father's shoe bench out," and "Joe called," the key word, Joe, is in both cases at the end of an NP. The boundary strength of "Joe" should be *greater* in the longer sentence, however, and hence "Joe" should be longer in the long than in the short sentence, but it is not.

Streamlining a Theory of Language Production

Clearly, the integrated model—even if all of its parts were compatible—would require streamlining. This will not be attempted here, but two strategies for streamlining are suggested. One is to ask what *minimally* an actor must be assumed to control explicitly in order to explain the dependent measures he reproduces, rather than asking what scheme, however complex, will reproduce the data. This has two corollaries. The first is to look for relationships and dependencies among various measures—that is, to suspect that if four or five measures all exhibit a similar patterning then in some direct or less direct way they are not independently regulated and should not be modeled as if they were five things rather than just one. Prime candidates for reduction in this way are the patternings of f_0, duration, pausing, and blocking of coarticulation, all of which suggest a relaxation of vocal gestures at boundaries (see also Cooper and Sorenson, 1981). A second corollary is that models should be avoided that generate the superficial form of a dependent measure itself

without serious attention either to what, if anything, might motivate the talker's imposing the patterning in the measure, or to how little explicit control needs to be assumed to explain the data.

A second strategy is simply to allow the measures from other research domains to limit and guide theorizing based on a particular measure.

SPEECH PRODUCTION

As Studdert-Kennedy (1980) reminds us in his review of *The Signs of Language* (Klima and Bellugi, 1979), language is form, not substance. Quite properly, then, the models of language planning just reviewed focus on providing an account of how language forms are organized to convey an intended message. From that perspective, the job for a speech production theorist seems to be one of explaining how the plan gets translated into action. Indeed, speech production researchers see that task as their job as well; it is the one outlined in Figure 6–8*a*.

As Kent (1983) points out, however, in the translation from planned message to articulation, "a difficult gap has to be bridged" (p. 59) because the things on the two sides of the translation are things of different types. Thinking of a translation like this taking place leads investigators to ask whether it can really be supposed that language forms are *in* vocal gestures (MacNeilage and Ladefoged, 1976). Indeed, some researchers are quite certain that they cannot be (Hammarberg, 1982; Repp, 1981) on grounds that language forms are irretrievably cognitive or mental and *therefore* can exist in the privacy of the minds of talkers and listeners, but can never make a public appearance.

A different way of conceptualizing the relationship between plan and actualization, which to me is more preferable, is schematized in Figure 6–8*b*. This conceptualization makes use of the observation that language forms, whether planned or actualized, are always realized in some physical medium. When they are planned, the medium (that is, the "physical instantiation" of Figure 6–8*b*) is at least neural; when they are uttered it is at least articulatory. In this conceptualization, there is no translation from mental to physical; indeed, there is no change at all at the "mental" or cognitive level of the linguistic message. There is ony a replication of the message across two media, brought about in whatever way that neural activity brings about motor activity.

Few would argue with most of Figure 6–8*b*, although some (Hammarberg 1976, 1982; Repp, 1981) would argue that the articulatory gestures only hint at the linguistic units, but do not in fact realize them.

Figure 6-8. Two views of the relationship between a plan and an utterance. (See the text for elaboration.)

The inclusion of neural underpinnings for the premotor language representation is untendentious, however. Nevertheless, even if it is uncontroversial, investigators still write as if the translation is from the planned message to the activity (that is, along the dashed diagonal of Figure 6-8*b*), not from one physical instantiation to another, and as if the problem of speech production is to explain that translation. (For example, Perkell [1980] states, "With an 'input' in the form of a feature-specified underlying representation, our hypothetical overview must account for the translation into articulatory movements" p. 347.) In Figure 6-8*b*, however, the problem is one of understanding how linguistic messages can be instantiated in physical media, both neural and articulatory, not how they can be translated *into* physical activity; they are already physically realized, and understanding how linguistic messages can be realized in physical activity is no less problematic for the plan than it is for the utterance.

What are the problems in respect to articulation? Consider again the passage from MacNeilage and Ladefoged (1976) cited earlier:

> [Researchers have] an increasing realization of the inappropriateness of conceptualizing the dynamic processes of articulation itself in terms of discrete, static, context-free lin[illegible]istic categories, such as "phoneme" and "distinctive feature." This development does not mean that these linguistic categories should be abandoned—as there is considerable evidence for their behavioral reality (Fromkin, 1971). Instead, it seems to require that they be recognized, even more than before, as too abstract to characterize the actual behavior of 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 (Fiint, 1962) of the structure of the language, such as those noted by phonologists. (p. 90)

MacNeilage and Ladefoged are responding to apparent contrasts between the properties of linguistic units at the message level ("discrete," "static," and "context-free") and those of their realizations or approximations in articulation (coarticulated, dynamic, and context sensitive). These contrasts probably are not unique to articulated language. Presumably, access to the *neural* activity going on as the message plan is constructed would offer no clearer picture of the critical properties of linguistic units than does access to vocal activity. Yet having witnessed the confusing neural activity, we would not conclude that "linguistic units should be recognized as too abstract to characterize the actual behavior of [populations of neurons] themselves" or that linguistic units are "better confined to primarily characterizing the earlier [preneural] stages of the production process." The buck has to stop somewhere.

As an alternative to assuming that phonemes, for example, are not in articulation (or in neural activity) because we cannot see them in it, we need to consider the possibilities that we cannot see them in these media both because we do not know how to look at the media in revealing ways

and perhaps also because we are looking for the wrong correlates of linguistic units.

In respect to the latter possibility, we can ask whether the properties of linguistic units that articulated speech does not have—in particular, the properties of being "static," "discrete," and "context-free"—*are*, in fact, properties of linguistic units. For example, it can be argued that linguistic units are not static (Fowler, Rubin, Remez, and Turvey, 1980). The dimension of static–dynamic is irrelevant to linguistic units at the message level. At that level, phonemes are symbols engaged in linguistic functions. The critical constraint that linguistic units such as phonemes place on any physical medium that realizes them is not that it realize them as static or as dynamic but that it realize them in *some* way that enables them to perform their linguistic functions.

To serve their functions, phonemes have to be separate one from the other and they have to be serially ordered when they participate in larger linguistic units. To achieve separation and serial ordering, however, phonemes need not be *discrete* if discrete implies nonoverlapping. The separation and ordering of phonemes is preserved in articulation not by discreteness but rather in the order in which each phonological segment predominates over others being coproduced with it, both in the vocal tract and in the acoustic signal (see "Coarticulation as Coproduction" later in this chapter).

Another constraint on a physical realization of linguistic forms is that tokens of a type (e.g., phones of a phoneme) be identifiable as such. The obvious way of ensuring this is to give the tokens of a type some context-free property or properties. Perhaps this is the way the vocal tract realizes this constraint (e.g., Stevens and Blumstein, 1981). We do not yet know, and other ways are possible (Smith and Medin, 1981).

A decision that phonological segments as articulated fail to preserve essential properties of linguistic units commits the theorist to more complicated theories of speech production (because a mind-to-body translation has to be confronted, as in Figure 6–8*a*) and of perception (because the objects of perception are not in the signal) than a decision that the message units are replicated intact across media. It seems to me best to pursue the more straightforward course first. Therefore, although it is a controversial matter, this chapter is written as if the fact that the *linguistic functions* of phonological segments and other linguistic units are preserved in articulation (in that perceivers do extract linguistic messages from acoustic signals) signifies that phonological segments are uttered intact in speech production.

The critical questions on which this focuses are as follows: What are articulated linguistic segments, how are they (here, phonological segments) realized in articulation, and how are vocal structures regulated to realiz them?

In contrast to the review of language production in the first part o this chapter, the remainder is not organized around theories of (parts of speech production because there are relatively few of them. Instead answer to the foregoing questions that have been offered in the literature ar presented. Following this, Perkell's (1980) model, which tries to incorporat answers to these questions, is described. Finally, which aspects of the mode will need revision or elaboration by future research and theorizing are considered.

What are Articulated Phonological Segments?

MacNeilage (1970) proposed identifying phonological segments with the achievement of spatially defined vocal tract targets. MacNeilage's proposal was based in part on the observation of "motor equivalence" (see later discussion). Talkers produce phonological segments that they and others agree are tokens of the "same" segment in different ways, depending on the context in which the token is produced. To take one of MacNeilage's examples, /t/ after/i/ requires a slight elevation of the tongue tip so that it contacts the alveolar ridge. Producing /t/ after /a/, however, requires closing the jaw and perhaps raising the tongue body in addition to tongue tip movement. All that appears invariant across these productions of /t/ is the spatial target reached by the tongue tip.

Several researchers, including MacNeilage, have pointed out that the proposal that the targets are spatial may be too constraining. For example, when Folkins and Abbs (1975, 1976) apply resistive loading to the jaw during a closing gesture for a bilabial stop, lip closure is achieved, but in a different spatial location from unperturbed closure. Specifically, it is achieved with a more open jaw and with a more extensive excursion of the upper lip than in unperturbed utterances. Possibly something similar happens in normal speech when bilabial closures following or preceding low or high vowels are achieved with relatively lower or higher jaw positions (Sussman, MacNeilage, and Hanson, 1973). In either case, the target is not spatially invariant.

Perkell (1980) has proposed that targets may be "orosensory goals." For example, achievement of bilabial closure, wherever the closure may be absolutely, has similar tactile consequences. Perkell's proposal more generally (see Figure 6–1*b* and later discussion) is that segments are specified as features with both auditory and production correlates. Their production correlates are orosensory goals; they may be proprioceptive or tactile, or

they may specify intended air pressure states or airflow characteristics of the vocal tract.

The view of articulated segments as achievements of targets, however defined, captures a central characteristic of the realization of many segments—namely, that of equifinality. However, there is a salient characteristic of articulated segments that the characterizations in terms of targets do not capture well, although with minor modification they could. If segments are only *achievements* of vocal tract states—whether the states are defined spatially or as orosensory goals—then most of the talking process involves getting to segments and less of it is involved in actually producing them. Elsewhere it was proposed (Fowler, 1977) that the phonological segment be seen as the collection of gestures that occur on the way to achieving the equifinal state charcteristic of the segment. In addition to allowing movement to be seen as more than transitional in nature, it allows a conclusion, for example, that vowels that are heard in speech really have occurred even if their "targets" have not been reached.

How are Linguistic Units Realized: Some Measures

One way to examine how linguistic units are realized is to look at the "traces" they leave in various measures investigators take of articulation or of the acoustic signal. Two measures are examined, duration and coarticulation.

Duration

Making durational measurements requires choosing a solution, however temporary or pragmatic, to the "segmentation problem." The segmentation problem in speech is the theorist's problem to extract separate, serially ordered phonological or phonetic segments from an acoustic signal in which acoustic correlates of different segments are interwoven. In studies of duration, typically, the solution is to measure phonological segments as if they were discrete. That is, the signal is segmented by drawing segmentation lines perpendicular to the time axis that serve at once as the "right" edge of one segment and the "left" edge of another. This is not the only way to segment the signal, and as is pointed out later, it implies a particular view of coarticulation that is not the only one possible either. The procedure does, however, reveal some systematic variation in durations of measured segments.

Figure 6–9 shows one of those systematic effects (see also Lindblom and Rapp, 1973). A vowel shortens as preceding or following consonants

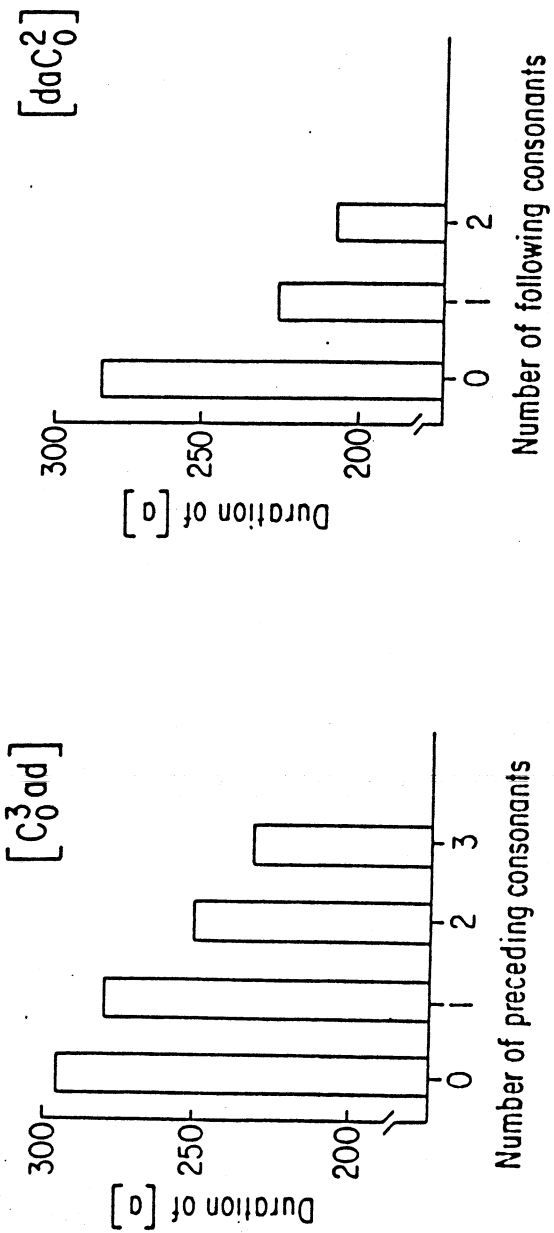


Figure 6–9. Shortening of a vowel as consonants are added to a syllable.

are added to a syllable. Transsyllabic consonants may also shorten a vowel (Lindblom and Rapp, 1973). In the present data, shortening by following intrasyllabic consonants exceeds that by preceding consonants; however, it is not known whether the asymmetry is reliable.

Figure 6–10 shows analogous effects of unstressed vowels on preceding and following stressed vowels. The left side of Figure 6–10*a* shows data from real-word sentences (Fowler, 1977); the data on the right side of Figure 6–10*b* are from reiterant-speech sentences (Fowler, 1981a). In data such as these (see also Huggins, 1975, 1978; Lindblom and Rapp, 1973), effects of preceding syllables are weak and sometimes absent; effects of following syllables are more substantial.

A final, systematic shortening effect occurs on the last stressed vowel of a sentence; this vowel is shortened by the number of stressed words that precede it in the sentence (Lindblom, Lyberg, and Holmgren, 1981; Lyberg, 1981). Analogous effects of preceding and following stressed vowels on nonfinal stressed vowels have been reported using reiterant speech (Lindblom and Rapp, 1973), but not always using real speech (Lyberg, 1981; however, see Huggins, 1978).

As Figures 6–9 and 6–10 reveal, the relationship between duration and number of relevant neighbors is negatively accelerated. It can be described formally (Lindblom, Lyberg, and Holmgren, 1981; see also Klatt, 1976) as follows:

$$D_0 = (D_I - D_{min}) \cdot \alpha^A \cdot \beta^B + D_{min}$$

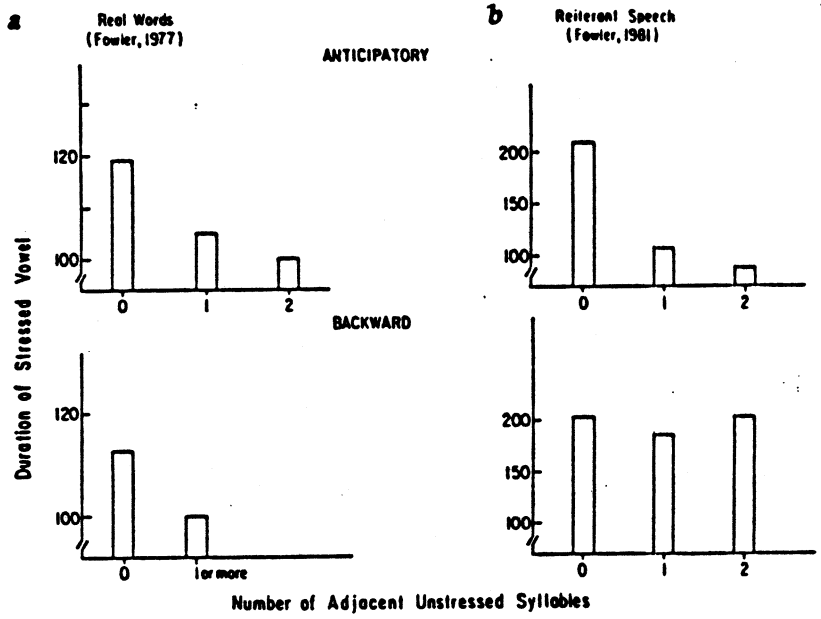
where D_0 is the measured duration of the vowel, D_I is the "inherent duration" of the segment, D_{min} is the vowel's incompressible duration, α and β are parameters of shortening and A and B are numbers of relevant following and preceding neighbors, respectively.

By itself, of course, the equation describes but does not explain the shortening pattern. Several accounts have been proposed.

One is to relate the shortening effects to alleged syllable and stress timing tendencies in languages (e.g., Abercrombie, 1964). Syllable timing is a tendency for speakers of a language to maintain approximately isochronous syllables; stress timing is a tendency to produce isochronous stress feet.

There are several objections to this account. First, it is incomplete as an explanation. To propose that talkers shorten segments in a syllable as more segments are added because they are trying to produce isochronous syllables leaves unexplained why they are trying to do that. Without some independent motivation for syllable or stress timing, the "explanation" is no more than a crude description of the measurements themselves. Second, the description is very crude. Shortening effects of a syllable or segment are small compared to the added duration of the syllable or segment itself.

Figure 6–10. Shortening effects of unstressed syllables on preceding and following stressed syllables. Data on the left is natural speech from Fowler (1977). Data on the right is reiterant speech from Fowler (1981a).



Third, languages are not supposed to be *both* syllable and stress timed, but English and Swedish show both effects.

Lindblom, Lyberg, and Holmgren propose an account in terms of a buffer in which a planned word sequence is stored for outputting. The buffer is seen as analogous to a box with elastic sides. Phonological segments are inelastic, deformable objects that can occupy the box. Shortening economizes on space in the buffer. To explain negative acceleration, Lindblom and colleagues propose that for every added unit in the buffer, a fixed percentage of the segment's compressible duration is given up.

There is one major difficulty with this explanation and one puzzle. The difficulty is in grounding the metaphor of the elastic box. A buffer is an entity at the message level of description of language (and hence, for many researchers, the "premotor" level); consequently, it is difficult to see what the analogue of the elastic walls would be or why segments would occupy space proportional to their uttered duration (that is, why, for example, a 300 ms /a/ would take up more space than a 200 ms /a/). The puzzle is that Lindblom and co-workers invoke an output buffer to

explain the shortening they observe, whereas Sternberg and associates invoke one to explain the lengthening they observe. The data bases from which the two sets of investigators make their proposals appear contradictory; this will have to be resolved. However, the disagreement between the proposals reveals something else as well. It is that, if the concept of "output buffer" can handle shortening and lengthening effects just as readily, it has little explanatory value in reference to the physical realizations of utterances.

Other investigators (Bolinger, 1963; Lyberg, 1979, 1981) have explained some of the durational effects in terms of fundamental frequency. In particular, they propose that long stressed vowels—for example, main stressed vowels or vowels in final position—are long to accommodate F_0 excursions that mark main stress or sentence finality. Although this can explain durational effects at the phrase level, it does not seem to address effects at the syllable and foot levels. These latter effects may be distinct in pattern from phrase-level effects (Lyberg, 1981), but they are remarkably similar to each other, as comparison of Figures 6–9 and 6–10 reveals. Presumably, then, they require analogous explanations. It is my belief that they are, in effect, measurement artifacts and reflect coarticulatory patterns (Fowler, 1981a). This viewpoint will be elaborated on after a look at those patterns themselves.

Coarticulation

Coarticulation is an "influence of a phonetic context on a given segment" (Daniloff and Hammarberg, 1973, p. 239) or it is overlapping production of two or more phonetic segments (Fowler, 1980, 1981a, b). First let us examine some of the data giving rise to the view that speech is coarticulated, and then explanations attempting to account for the data.

The Lips and the Velum. Coarticulatory activity by the lips and velum may be special in one or two ways. First, it can range over very long extents; second, it exhibits a large asymmetry in favor of anticipatory coarticulation.

As for the lips, Benguerel and Cowan (1974) report evidence of lip rounding for a rounded vowel among speakers of French as many as six segments in advance of the measured onset of the vowel itself. In particular, they report that lip rounding for a forthcoming rounded vowel begins immediately following a preceding unrounded vowel no matter how many segments intervene between the two vowels.

Bell-Berti and Harris (1979, 1981) provide a different characterization of the extent of rounding anticipation. They report a fixed *temporal* extent of anticipatory coarticulation of lip rounding. In their data, rounding precedes measured /u/ onset by about 200 ms regardless of the number of consonants preceding the vowel either within or across a word boundary.

These different characterizations are, in part, different ways of looking at the same phenomenon, but, in addition, they may rest on contradictory data bases. Benguerel and Cowan do not report the temporal extent of lip rounding in their data, but the implication is there that it increases with the number of consonants preceding the rounded vowel. This finding is present for three speakers of English in research reported by Sussman and Westbury (1981). In contrast, however, Bell-Berti and Harris find a fixed temporal extent. The disagreement in the data is important to resolve because the two descriptions have suggested quite different characterizations of segment production in speech. One has promoted a view of coarticulation as assimilation (e.g., Hammarberg, 1982)—that is, as a spreading of features from one segment to another—while the other suggests that the rounded vowel consists of a temporally extended sequence of gestures, at least one of which overlaps with gestures for other segments.

The observational differences may be resolvable, at least among the English speakers that have been studied. Bell-Berti and Harris do find a shorter anticipation of rounding in a vowel-consonant-/u/ (VCu) context than in a context of two or more consonants preceding the /u/. In their view, this occurs because the talker avoids rounding a preceding phonologically unrounded vowel and would not avoid it with one preceding consonant if the extent of anticipation were as great as it is with two or more preceding consonants. For their part, Sussman and Westbury only compare two contexts for lip rounding: one preceding consonant or three.

As for the velum, lowering the velum for a nasal segment may precede measured onset of a nasal consonant by one segment or two if the segments are vowels (Kent, Carney, and Severeid, 1974; Moll and Daniloff, 1971). Ohala (1971) finds that velum lowering begins as soon as a closed port is no longer needed for production of any obstruent that may precede the nasal consonant. In addition, he finds more extensive anticipatory than carry-over coarticulation.

Bladon and Al-Bamerni (1982a) report two components to many velum-lowering gestures for a nasal consonant. One component increases its anticipatory extent with the number and duration of vowels preceding a nasal consonant. The other, present in about half of the productions, is an additional higher velocity gesture, time-locked to the nasal consonant's oral articulation. The investigators suggest that the second stage may represent onset of palatoglossus activity brought in occasionally to augment the velum-lowering effects of relaxation of the levator palatini.

They provide indirect evidence for this idea in a later study (1982b) in which velar movement is observed in phonological sequences that do not necessarily include a nasal. Utterances, produced by a speaker of Kurdish and a speaker of Arabic, contained vowels surrounding a nasal consonant, or a pharyngeal or glottal fricative. Anticipatory coarticulatory

effects of the nasal consonant replicated those of earlier studies, showing earlier anticipation the longer the string of pre-consonantal vowels. Like the nasal consonant, the pharyngeal and glottal consonants were produced with a lowered velum, but in these latter cases, anticipatory velum lowering was time-locked to production of the consonants. Bladon and Al-Bamerni suggest that, in production of pharyngeal consonants at least, the palatoglossus muscle is activated along with other faucial muscles to achieve pharyngeal construction. A byproduct (in this case) of palatoglossus activity is velum lowering. Because activation of faucial muscles achieves the construction that Bladon and Al-Bamerni identify as onset of the fricative, velar lowering here is time-locked to measured consonant onset.

This distinction between types of anticipatory coarticulation seems to imply that some coarticulatory effects are sensitive to the segmental composition of a segment's preceding (and, perhaps following) context (that is, they are context sensitive) while others are involved directly in the realization of the given segment itself (and, relatively speaking, are context-free). If the distinction is real, it may clarify the disagreement between Benguerel and Cowan and Bell-Berti and Harris. The one pair of investigators sees anticipatory lip rounding as a context-sensitive gesture, and the other sees it as a relatively context-free component of rounding a vowel.

Glottal and pharyngeal consonants are not the only oral phonological segments to have characteristic postures of the velum. Velar height in vowels is correlated positively with vowel height, while, among vowels and consonants, obstruents have the most closed velum position. Bell-Berti (1980) has shown that the characteristic velar positions for vowels coarticulate. In particular, she finds that a sequence of obstruents has a higher peak velar position if it is preceded or followed by /i/ than if it is preceded or followed by /a/. Similar to velum lowering for the nasals, there is more anticipatory than carry-over coarticulation of vowel-associated velum height.

Velar gestures show other context effects as well. In a sequence of up to five obstruents, each of which individually has a high position of the velum, Bell-Berti finds that the velum rises throughout the sequence. Therefore, the peak velar position is higher in /ist#ta/ than in /it#ta/ and in /its#sta/ than in /ist#ta/. Bell-Berti suggests that the velum specification for each segment may be a movement relative to the velum's current position rather than a spatial target.

Taken together, these systematic activities of the velum give rise to two related questions. The questions to be asked are to what extent the characteristic velar positions or gestures for oral segments in fact need to be controlled by the talker and to what extent, instead (as in Bladon and Al-Bamerni's proposal for pharyngeal consonants), they "fall out" of other things that the talker regulates. The second, more general question asks whether phonological segments to be uttered are given values on all possible articulatory dimensions (including velum height for vowels), or whether they are unspecified on certain apparently irrelevant dimensions.

As to the first, more specific, question—whether the velar positions for each oral segment and their anticipatory and carry-over effects are regulated directly—different answers are suggested by different sources of evidence. If it were possible to extrapolate from the findings of Bladon and Al-Bamerni to the vowels and obstruents studied by Bell-Berti by discovering some muscular or mechanical coupling that affects port size, the variability in velar height among vowels and obstruents could be seen as a byproduct of more primary articulatory gestures for the segments. Fujimura (1980) does find that tongue position and velar height are sometimes correlated. In his example, he observes both a higher posterior tongue position and a higher velum in production of the coda in "pence" than in "pens," and suggests that whatever the causal direction of the effect may be, the coupling itself is probably mechanical and perhaps muscular.

However, Bell-Berti (1980) argues that the different velar positions for different oral segments are regulated directly. She cites research showing that velar elevation in oral segments is correlated with activity of the levator palatini whose major action is to raise the velum. She hypothesizes that the port adjustments during vowels are made to prevent nasal coupling. Coupling is less likely at a range of openings during articulation of open than of closed vowels.

The second question becomes important when theories of coarticulation are devised or evaluated (see Explanations of Coarticulation later in this chapter). In some extant theories, that of Henke, for example (1966), anticipatory coarticulation of a "feature" is allowed to range over any segments that are unspecified for that feature. Vowels become nasalized in the context of a nasal consonant because (in English) vowels are unspecified for nasality; chameleon-like, they acquire the properties of their neighbors. However, Bell-Berti interprets her data as evidence that vowels are specified for velum height and she speculates that the specified height is a minimal one for each vowel that will *prevent* nasal coupling. The height specification for a vowel, then, is antagonistic to the nasal feature, and so, according to Henke's theory, it should prevent spread of the nasal feature in vowels. Coarticulatory spreading does occur, however, and, in Bell-Berti's data, it results in something like a vector summation of the different velar heights of neighbors.

Tongue and Jaw. Jaw and tongue movements for a consonant or vowel segment, like the lip and velar movements just described, exhibit both

anticipatory and carrry-over coarticulatory spread. The anticipations, however, are less marked than for lip and velum. Indeed, investigators have rarely reported marked asymmetries in direction of tongue and jaw coarticulation, and when they have reported an asymmetry—in particular, for effects of stressed vowels on preceding and following unstressed vowels (e.g., Bell-Berti and Harris, 1976; Fowler, 1977, 1981a, b)—the asymmetry is opposite to that reported for lips and velum: carry-over effects are more substantial and extensive than anticipatory effects.

Sussman and Westbury (1981) suggest that anticipatory coarticulation of tongue gestures is less extensive than lip and velum gestures "for the obvious reason that the tongue is intrinsically involved in all speech segments (except those articulated solely at the glottis)" (p. 16). This conclusion is a little sweeping; the investigators do not defend their implication that the tongue is intrinsically involved in bilabial consonants and labiodentals. More importantly, at least one reading of their proposal does not accurately reflect the full range of coarticulatory interactions in which the tongue does participate. Interactions do occur between tongue gestures of adjacent segments when the tongue, apparently, is "intrinsically" involved in the production of both segments. For example, Perkell's (1969) cineradiographic data show concurrent constricting gestures of the tongue body for /k/ and fronting for /ɛ/ in production of /həkɛ/. During closure for the /k/, the tongue body makes a forward sliding gesture along the palate toward the more front positioning it will take for the /ɛ/. In addition, Kent (1983) reports influences of a /k/ on tongue gestures for a preceding diphthong. Hence, involvement of the tongue in producing a segment does not prevent its being influenced by tongue gestures for neighboring segments.

Sussman and Westbury may have had something else in mind, however. The tongue and jaw are "primary articulators" for many segments, at least in the restricted sense that their movements achieve acoustic consequences that guide researchers' segmentation of the acoustic signal into phonetic segments. That is, investigators do not measure onset of /n/, for example, at the point where nasal coupling for the /n/ is first evident. Instead, they measure its onset from the point where the tongue first makes contact with the palate, thereby initiating the segment's closure phase. Similarly, the onset of /u/, conventionally, is not measured at the point where effects of rounding are first evident in the signal; rather, its onset is located at a point where the jaw and tongue have opened the vocal tract sufficiently to create the voiced formant patterns characteristic of vowels and other sonorants.

Conceivably, then, coarticulation of tongue and jaw gestures is less extensive anticipatorily than lip rounding and nasalization for the same reason that Bladon and Al-Bamerni's high-velocity velar gesture is time-locked to pharyngeal consonant onset. These are gestures that *achieve* the acoustic consequences identified as segment onset. If they were anticipated more, the segment itself would be anticipated.

As for the types of interaction that are achieved, MacNeilage and DeClerk (1969) report less extensive coarticulatory effects of consonants on vowels than the reverse influences in stop-consonant–vowel–stop-consonant (CVC) syllables. In an anticipatory direction, they found the tongue body configuration during the first consonant of a CVC to vary with the identity of the following vowel. Carry-over effects into the second consonant could be detected in the pharyngeal region; the pharynx was narrower following a low than a high vowel.

Many investigators have reported noticeable anticipatory and carry-over effects of vowels on consonants in vowel–consonant–vowel (VCV) productions (Barry and Kuenzel, 1975; Butcher and Weiher, 1976; Carney and Moll, 1971; Kent and Moll, 1972; Ohman, 1966). Ohman characterized the coarticulatory effects in VCV utterances as diphthongal vowel-to-vowel gestures on which consonant gestures are superimposed. Kent and Moll provide limited supportive evidence in a comparison of tongue movement from /i/ to /a/ in the utterances "he honored" and "he monitored!" In these productions, the timing and extents of the /i/ to /a/ gestures were identical even though in the one case a consonant intervened and in the other none did.

Carney and Moll (1971) extend Ohman's observations on stop consonants in VCV utterances to fricatives, and again find in cineradiographic vocal tract cross sections, clear evidence of vowel-to-vowel movements of the tongue body during closure for labiodental and alveolar fricatives. In their study, in which both vowels in the VCV utterances were stressed, Carney and Moll found no influence of the second vowel on the steady state configuration of the tongue for the first vowel. They do not discuss carry-over effects on the vowel's steady state. Evidence of vowel-to-vowel coarticulation *is* found when the influencing vowel is stressed and the influenced vowel unstressed (Bell-Berti and Harris, 1976; Fowler, 1977, 1981a, b). These coarticulatory effects are asymmetrical with carry-over effects more extensive and substantial than anticipatory effects.

Explanations for Coarticulation

Several critical reviews of explanations for coarticulation are available in the literature (e.g., Daniloff and Hammarberg, 1971; Kent, 1983; Kent and Minifie, 1977). Only a selective review, therefore, is provided here, focused on three kinds of explanation offered in the literature.

There is no Coarticulation. Wickelgren's proposal (1969; 1976) is that there is no coarticulation. Rather, segments are context sensitive because talkers and listeners store a large inventory of versions of each segment, called "context-sensitive allophones." Talkers pick a version to be uttered according to the context of segments in which it will appear. The stored versions are each adjusted to a unique segmental context of the form *X–Y* where *X* and *Y* are segments that might surround the phoneme in an utterance.

This proposal has not been well accepted in the speech literature. However, as Norman (1980) points out, it has resurfaced in certain machine-based speech-recognition schemes (e.g., Klatt, 1980). In Norman's words:

> I find it somewhat amusing to see this suggestion resurfacing, to see it being taken seriously, and to find that is is perhaps correct. (p. 389)

It is almost certainly not correct, however, and perhaps it is worthwhile to point out some reasons—old and new—why it fails. Criticism of the view has focused largely on the number of allophones that would be required to cover all the ways that a phoneme is produced. When the full range of anticipatory and perseveratory coarticulatory influences is considered as well as the variations in stress and rate of speech that affect segment production, the number of context-sensitive allophones needed to be stored could be very large indeed.

But, of course, so is our lexicon very large and no one is disturbed by the idea of storing tens of thousands of words. What is wrong with the theory of context-sensitive allophones is more obliquely related to the issue of the number of allophones that would be required.

In my view, Wickelgren's theory is falsified by evidence that segment production is generative—that is, it is falsified by the same kind of evidence that falsifies a theory that all possible sentences are stored. Generativity in segment production, as in sentence production, implies procedures for generating appropriate instances of segments in any novel context in which they might appear.

Talkers can produce acceptable versions of phonological segments in ways they never have before. For example, talkers produce normal or near-normal vowels with a bite block clenched between their teeth that fixes the postiion of the jaw (e.g., Lindblom, Lubker, and Gay, 1979). Vowels produced in this way show only minor effects of practice; they are nearly as close to normal as they ever will be from the first pitch pulse of the first vowel produced under bite-block conditions. This observation holds even when the vowels are produced under fairly severe time pressure (Fowler and Turvey, 1980). The limited effects of practice suggest that the small departures from normality of early productions are largely due to consequences of the bite block that physically *cannot* be compensated for.

Even more impressive, perhaps, are talkers' immediate ("on-line") compensations for unpredictable perturbations in movement of an articulator (e.g.) Folkins and Abbs, 1975, 1976; Kelso, Tuller, and Fowler, 1982). Talkers achieve bilabial closure for a /p/ or /b/ when the jaw is unpredictably loaded during its closing gesture. Closure is in part achieved by a short-latency compensatory activation of the upper lip (e.g., Folkins and Abbs, 1976). Under these novel conditions, a talker would be helpless if he or she had only a lexicon of context-sensitive allophones from which to select an allophone.

Speech researchers have been negatively inclined toward Wickelgren's theory for other reasons too, however. One reason is, again, obliquely related to the number of context-sensitive allophones required to cover the different contexts affecting segment production: by assigning the context sensitivity to the stored segments individually rather than to context-sensitizing procedures, the theory of context-sensitive allophones obscures the general character of coarticulation that may explain the *raisons d'être* of context sensitivity. For example (Kent and Minifie, 1977), in English all vowels are nasalized before a nasal consonant. In Wickelgren's descriptive system, this generalization is missed; it is as if coincidental that in the context $_xV_n$ (where "*x*" stands for any preceding context and "*n*" stands for any nasal consonant), all Vs are $\tilde{V}$s.

At the other extreme, in some ways the scheme is insufficiently detailed. In particular, the subscripts surrounding each context-sensitive segment do not really specify the kinds of context effects that will be found. Obviously, not all properties of a segment are shared with neighbors. For example, in the context of a nasal consonant, vowels become nasalized, but they do not become obstruents or take on the nasal's place of articulation. A theory has to specify which properties of a segment will spread, and to what degree. The job of doing so is complicated by the fact that a given segment may share different properties with different neighbors. For example, whereas vowels become nasalized before nasal consonants, oral consonants do not. There is no way to represent that difference in Wickelgren's allophone scheme because vowels and oral consonants before a nasal have the same subscript.

In short, even if the theory of context-sensitive allophones were not falsified by evidence of generativity in segment production, it would need to be supplemented in two ways: (1) by a listing of the general principles of coarticulatory spreading (for example, the principle that nasalization spreads anticipatorily to vocalic segments), and (2) a specification for each kind of segment in the inventory, which of its properties will spread in which contexts. But these supplements to the theory of context-sensitive allophones themselves would constitute a set of procedures for generating coarticulatory effects and would, then, obviate the inventory of allophones.

Coarticulation as Feature Spreading. The next account of coarticulation (Daniloff and Hammarberg, 1973; Hammarberg, 1982; Henke, 1966) constitutes a sort of compromise between Wickelgren's view that there is no coarticulation and the view that is considered last, of coarticulation as coproduction. In this second view, talkers store phonemes (or perhaps a few extrinsic allophones) rather than context-sensitive allophones. Each stored segment is specified as a bundle of features constituting the segment's "canonical form" (Daniloff and Hammarberg, 1973). Canonical segments influence one another by sharing features. In one account (Daniloff and Hammarberg) assimilation by feature is considered sometimes necessary to avoid production of transitional sounds between two planned phonemes.

Some coarticulatory effects—in particular, those involving anticipatory lip rounding and nasalization—range too far for the explanation in terms of transitional sounds to be plausible. Henke (1966) proposed that features tend to spread in an anticipatory direction so long as segments preceding the segment from which the features originate are unspecified for them. So, for example, a rounding feature can spread from a rounded vowel to any preceding and following consonants, because consonants are not specified on a dimension of lip rounding. Similarly, in English, the nasal feature can spread from a nasal consonant to any vowel because vowels are unspecified for nasality. Spread of the nasal feature will be halted by any oral consonant. This proposal handles gross characteristics of the data on rounding reported by Benguerel and Cowan and on nasalization by Moll and Daniloff (1971), but it fails on closer inspection.

First, segments with incompatible feature specifications do coarticulate. For example, Benguerel and Cowan find a small amount of rounding of /i/ by an upcoming rounded vowel. But /i/ contrasts with the vowel /y/ in French and hence must be specified [—round]. Second, Sussman and Westbury (1981) find earlier rounding activity of the orbicularis oris muscle in an iC}u context than in an aC} context. This is unexpected in Henke's scheme because /i/ is positively specified for a spreading feature antagonistic to rounding, whereas /a/ is simply [—round]. At best, there should be no difference in onset of rounding in these two contexts, according to Henke's theory, because both vowels have rounding specifications. In Sussman and Westbury's view the theory should predict later onset of rounding in the context of /i/ because /i/'s spreading gesture is more antagonistic to rounding than is /a/'s absence of a rounding gesture. Presumably, the earlier orbicularis oris muscle activity occurred precisely because spreading has to be overcome in the context of /i/ and not in the context of /a/.

A second possible difficulty was alluded to earlier. Bell-Berti's data on velar height in oral consonants and vowels suggest the possibility that segments are not really "unspecified" on some feature dimensions (or at least on certain articulatory dimensions) that are not defining for the segments. Rather, in some cases, effects of different segments on the same articulator may combine in some way (see also Perkell, 1980). This does seem to describe what happens in respect to tongue body movement during CV syllables in which the consonant is velar (e.g., Kent, 1983; Perkell, 1969).

Other data and considerations are also hostile to the more general view that coarticulation is feature spreading. Carney and Moll (1971) provide cineradiographic evidence already described of tongue body configurations during closure for a fricative. The evidence shows configurations intermediate between the steady state configuration for the vowel preceding the fricative and that for the following vowel. Evidently, the tongue body moves smoothly from its shape for the first vowel to its shape for the second. An account of this phenomenon as feature spreading is strained, and, in any case, appears to overlook what is really going on.

An account in terms of feature spreading would have to allow the *different* feature specifications for the two vowels' front–back and height dimensions all to spread to the intervening consonant and then for the specifications from the different vowels on each dimension be averaged in a way that weights the first vowel more than the second at the beginning of the consonant with second vowel weighted increasingly throughout the consonant. However, clearly, that is not what is going on; rather, vowel-to-vowel articulatory gestures occur during closure for the fricative.

A second data-based objection to the feature-spreading account is related to the first. It is that coarticulatory gestures are not as neatly timed as the feature-spreading view seems to suggest. Sussman and Westbury point out that, in contrast to Bell-Berti's data on anticipatory lip rounding, their own data show no time-locking of rounding to the measured onset of /u/. Instead, lip rounding anticipates the measured /u/ onset more as the number of preceding consonants increases. Their data also show that onset of lip rounding is time-locked to *no* acoustic marker (see Sussman and Westbury, 1981, Figure 2). In some cases rounding begins well within the initial vowel in the VC}V sequence. But if lip rounding were a feature that had been spread to the first vowel, it should have been realized along with the other features of the vowel and should have appeared at vowel onset. Alternatively, if lip rounding had spread only as far as the first consonant in the intervocalic sequence of consonants, rounding should not have appeared within the first vowel at all.

A difficulty with the theory of feature spreading brought out by these data is its proposal that a segment is *intended* to occupy a discrete interval

of time during which its features are realized. It is my belief this idea derives from an overinterpretation of linguistic notational devices (see also Fowler, 1980; Fowler, Rubin, Remez, and Turvey, 1980).

When a linguist chooses a notational device, such as one of representing sequences of segments to be uttered as a left-to-right array of discrete feature columns, his or her choice reflects a convergence of considerations and constraints. One constraint is that the device must reveal clearly the critical properties of the phenomena being represented—here, the distinctive attributes of phonological segments, the coherence of the features of a given segment, the separation of phonological segments in a sequence one from the other, and their ordering in words. Other constraints, however, are more mundane ones—for example, that the device be visible and be conveniently reproducible on paper. These latter constraints, as well as the theoretically interesting ones, guide the selection of a column to represent the coherence of the distinctive attributes of a phonological segment and the selection of a left-to-right array of discrete columns to represent the separation and ordering of phonological segments in a sequence.

As researchers attempting to use the fruits of linguistic analyses, we have to distinguish the essential properties of notational devices such as this one from the properties that are either accidental (for example, that the columns, meant to reveal the grouping of features into segments, make it look—if the order axis is confused with the time axis—as if segments should be temporally discrete and have simultaneous onset of all of their features after spreading) or are promoted by the special physical medium in which the device is realized. A speech plan will not turn out literally to be a buffer consisting of an array of feature columns unless nervous systems and vocal tracts share critical properties with marks that pens make on pieces of paper. The means of realizing the speech plan and the utterance will be a means that preserves the critical linguistic properties of phonological segments in the medium in which they are instantiated.

Coarticulation as Coproduction. In a view of coarticulation as feature spreading, segments ideally are temporally extended in a simple way. Features within a column are turned on at some point in time, and then later are turned off. A segment extends in time from activation to deactivation of its features. In this way, segments are discrete one from the other.

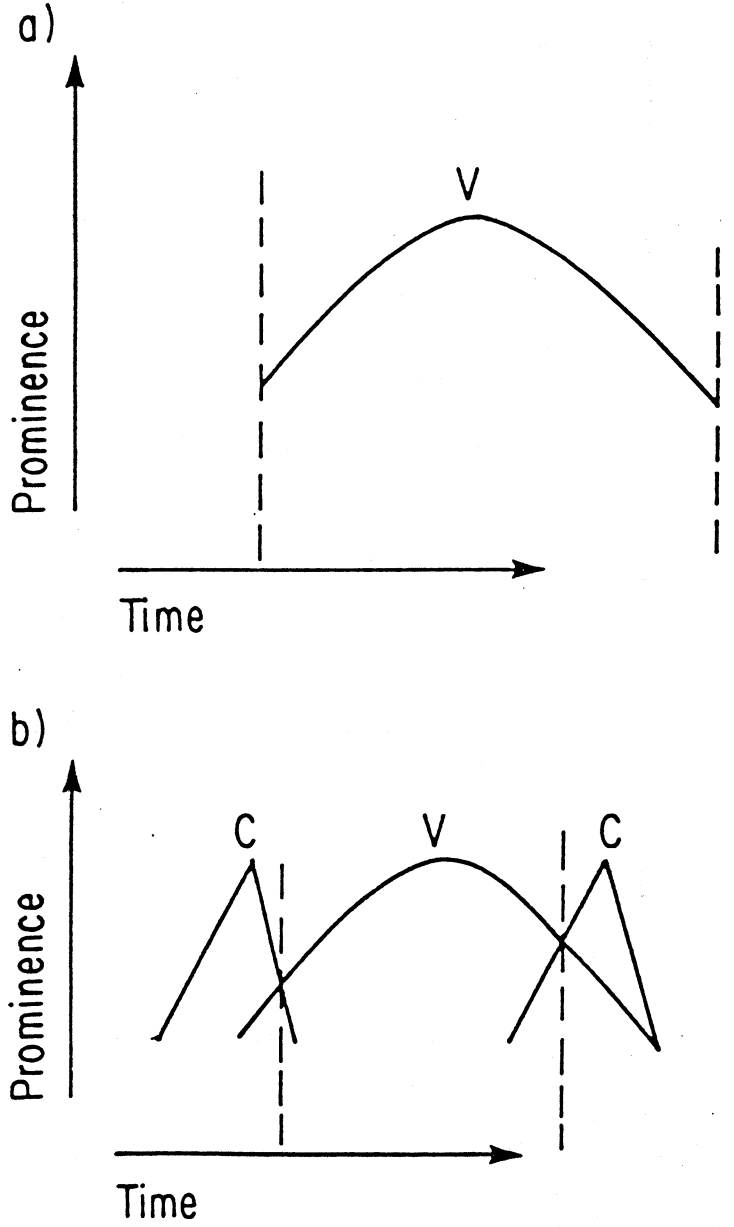
In contrast, in a view of coarticulation as coproduction, segments are temporally extended in more complex ways. Their constituent gestures are not initiated concurrently and they typically overlap with gestures for neighboring segments. In this way, segments literally are coarticulated (Fujimura, 1981). For example, onset of lip rounding for a rounded vowel generally precedes onset of gestures of the jaw and tongue body for the same segment and it overlaps with gestures for preceding segments. In a view of coarticulation as coproduction, despite its temporal precession of jaw and tongue body movement, the lip rounding gesture is tied to the other gestures producing the rounded vowel and not with gestures for the earlier segments with which it overlaps.

Evidence that coarticulation is, at least in part, coproduction of neighboring segments is perhaps strongest in respect to movements of the tongue body and jaw for vowels. In VCVs, gestures for the first vowel overlap with those for the following consonant so that, for example, a bilabial consonant has a lower jaw position during closure if the vowel is /ae/ than if it is /ɛ/ and lower if /ɛ/ than /i/. Likewise, the jaw shows less elevation in the closing gesture for the bilabial consonant if the second vowel is /ae/ than if it is /ɛ/ and less if it is /ɛ/ than /i/ (Sussman, MacNeilage, and Hanson, 1973). Data from three studies already described (Carney and Moll, 1971; Kent and Moll, 1971; Ohman, 1966) suggest that during consonant closure in a VCV, gestures of the tongue body are smooth gestures from V_1 and V_2. Ohman's proposal that vowels are produced as diphthongal gestures with consonants superimposed seems to fit the extant data on VCVs more naturally than a proposal that tongue body positions are features spread from both vowels to the consonant and given differential weightings throughout the consonantal closure.

The specific proposal that vowels are partially coproduced with consonants appears to have independent support from the data on durational shortening described earlier. Vowels in a syllable are measured to shorten as consonants are added to the syllable. However, although they are measured to shorten, in fact they may not shorten at all in their articulatory extents; indeed, they may even lengthen a little. Measuring conventions select as measured onset of a segment a point where the segment begins to predominate over neighbors in an acoustic signal; they do not include the whole extent of a segment's influence on the acoustic signal. If consonants overlap with vowels, then even were the vowels' articulatory extents invariant in isolation and in the context of a consonant, they would be measured to shorten in the second context. Figure 6–11 illustrates this idea.

In the figure, the horizontal axis represents time and the vertical axis an abstract dimension of "prominence." This refers to the extent to which relevant articulators are given over to the production of a particular segment and similarly to the extent that the acoustic signal's dominant character is that of the segment. Dashed lines in the figure represent places in an acoustic signal where a given segment begins or ends its phase of predominance. If a vowel is produced in isolation, as in Figure 6–11*a*, the

Figure 6-11. Schematic view of syllable production in which vowel and consonant production overlap.



dashed lines will be at onset and offset of voicing. For a vowel in a CVC, however, one line will be drawn where the constriction for the first consonant has given way sufficiently for the vowel's formant pattern to begin to dominate in the acoustic signal; another line will be drawn at closure for the final consonant. In a CVC, then, the vowel's produced extent is greater than its measured extent. This is less true in an isolated vowel.

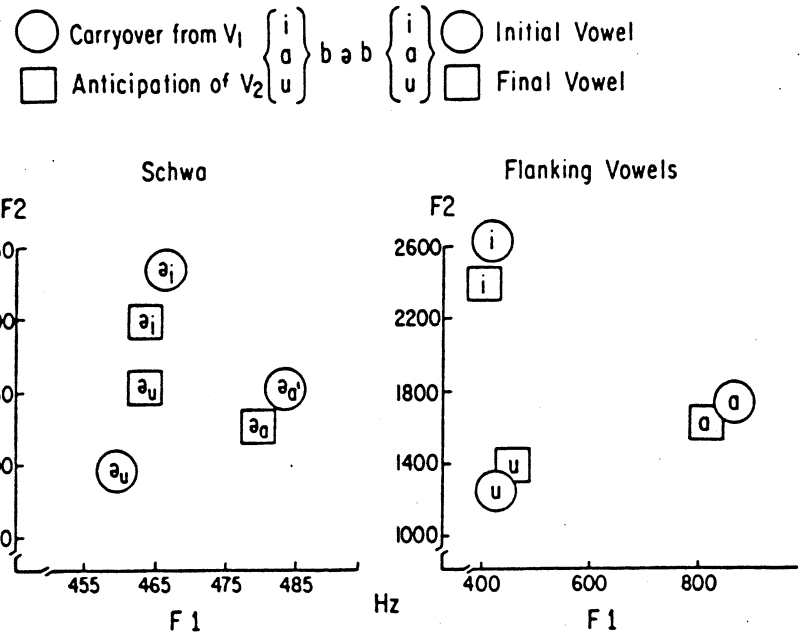
The measured shortening of a vowel when a consonant is added to a syllable is less than the duration of the added consonant. This may be (as Figure 6-11 indicates) that the vowel's articulatory extent does not span the whole syllable, or it may be because the vowel in fact lengthens.

Shortening at the foot or word level has a similar explanation within a view of coarticulation as coproduction. Figure 6-12 illustrates coarticulatory effects of preceding and following stressed /i/, /a/, and /u/ on a medial unstressed schwa (data from Fowler, 1981b). The figure shows substantial carrry-over influences on schwa and lesser anticipatory effects. On a coproduction view, the medial unstressed vowel is produced as a brief deflection of gestures of the tongue body and jaw from their stressed-vowel to stressed-vowel trajectory. That is, the medial unstressed vowel is coproduced with both stressed vowels, but it overlaps with the first one more than with the second.

Figure 6-13 depicts the asymmetry schematically. It appears to reflect the foot structure of English reported by linguists (Abercrombie, 1964; Catford, 1977; Selkirk, 1980a) and encountered earlier in this manuscript as a production unit proposed by Sternberg and associates (1978). In the linguistics literature, the foot emerges from analyses of the metrics of English and other languages as a structure superordinate to the syllable that organizes the placement of strong and weak syllables in words. It consists of a strong (stressed) syllable followed by any weak syllables up to the next strong one. That is, in a foot, weak syllables cohere more with preceding than with following strong syllables. Figure 6-12 shows that this coherence pattern is also reflected in coarticulatory relationships between stressed and unstressed syllables and Figure 6-10 shows that it is reflected in shortening patterns at the foot or word level. Figure 6-13 illustrates why coarticulation and shortening patterns are similar.

The dashed lines in Figure 6-13 represent segmentation lines that would be drawn if the syllables in the figure were to be measured. If the utterance were a disyllable VV (the presence of any consonants in the sequence will be ignored here), measurement lines would be drawn at points a, c, and e in the figure. The measured duration of the first vowel, then, would be c-a and of the second e-c. In a VvV utterance (where V is stressed and v unstressed), however, segmentation lines are drawn at a, b, d, and e and

Figure 6–12. Coarticulatory effects of a stressed vowel on unstressed schwa. The plot on the right gives F1 and F2 values for preceding and following /i/, /a/ or /u/. The plot on the left gives the F1 and F2 values for schwas in the context of preceding or following /i/, /a/, or /u/.



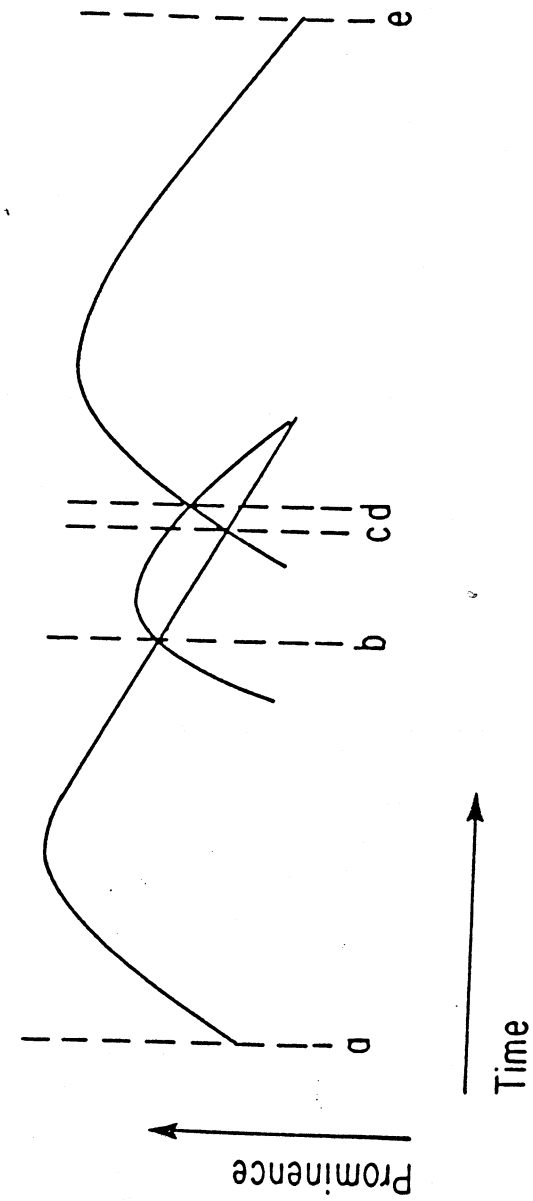
the measured durations for the stressed vowels are b–a and e–d, respectively.

Just as a vowel in a CVC is measured to shorten by its extent of coproduction, so is a stressed vowel in a VvV utterance. The measurement lines in Figure 6–13 also show that the first vowel shortens more than the second because, as the coarticulation data reveal, the overlap between a weak syllable and a strong syllable in the same foot is greater than cross-foot overlap.

Recent research of my own (Fowler, 1981a) directly compares coarticulation and shortening effects in sequences of stressed and unstressed syllables. The research shows that measures of the two effects are correlated, and that a version of Lindblom, Lyberg, and Holmgren's equation given above, which was designed to predict shortening effects, does a good job of predicting coarticulatory variability of F_2 in an unstressed vowel as a function of its context of stressed vowels.

The view of coarticulation as coproduction probably does not cover all instances of contextual influence in speech. For example, it may not

Figure 6–13. Schematic representation of foot production in English showing putative overlap of stressed by unstressed syllables.



explain the shift in place of articulation of /d/ in "width" (Daniloff and Hammarberg, 1973). It does not immediately explain why lip rounding and velum lowering migrate as far as they do. However, the explanation has several strong points in its favor.

1. It provides a more natural account of data on vowel production in VCVs than do other accounts.

2. It provides a unified account of these coarticulation effects and a foot- and syllable-level shortening.

3. It is the only account of the three reviewed here that provides any sort of connection to evidence from language production research that the metrical structure of language organizes speech output.

How Vocal Structures are Regulated to Realize Phonological Segments: Motor Equivalence

Motor equivalence is a flexibility in skilled activity so that a goal may be reached in a variety of ways. This allows an actor to accommodate to the different contexts in which the activity occurs and to compensate for unforeseen perturbations in the course of his performance. Uttered segments exhibit motor equivalence in a variety of ways.

First, uttered segments exhibit motor equivalence in ongoing, unperturbed speech. For example, the relative contributions of the jaw and the lips to bilabial closure varies with the height of coarticulating vowels (Sussman et al., 1973). Second, they show compensation for perturbations both foreseen, as in the bite-block studies (e.g., Lindblom et al., 1979), and unforeseen, as in studies of resistive jaw loading (Folkins and Abbs, 1975; 1976; Kelso et al., 1982). In the last example, compensation by the upper lip to jaw loading closing for a bilabial consonant begins only 15 to 30 ms after onset of a perturbation (Kelso et al., 1982).

Motor equivalence provides a strong hint concerning regulation of vocal structures that realizes phonetic segments. In particular, it reveals that vocal structures are coordinated. The implications of this observation for theories of speech motor control will be explained below. First, however, let us look at Perkell's model to see how it realizes the characteristics of articulated speech outlined above: phonological segments realized as targets or movements toward targets, durational shortening, and coarticulation.

A Summary Model: Perkell (1980)

Perkell (1980) intended his model to provide a heuristic framework for discussion and research rather than to serve as a detailed simulation

of processes supporting speech production. The model has three broad phases, depicted in Figure 6-1*b*: preplanning, determination of motor commands, and realization of commands as peripheral vocal tract activity. The first two phases are elaborated further.

In the model, planning for speech production starts with a sequence of phonetic segments. Each consists of a column of features specified as "orosensory goals." Planning provides timing specifications based on several different properties of the sequence of segments. First, each segment has a stored intrinsic duration. This duration is modified by the context in which the segment is produced. For example, research reveals trading relations among velocity of movement, the distance that the articulator has to move (with more extensive movements generally faster than shorter movements, e.g., Kuehn and Moll, 1976), and the required precision of movement. Mechanical interactions among articulators involved in realizing different goals may also affect timing.

The output of the timing component of the model is a temporally specified sequence of orosensory goals. In the model, the temporal regularities described earlier under Duration would arise here along with others, for example, the increased duration of a segment before a voiced consonant.

Next, orosensory goals are converted to motor goals. The strategy incorporates knowledge of the relationships between sensory goals and the patterns of muscle activations required to realize the goals. The conversion is complicated by the fact that orosensory goals having overlapping temporal specifications may make demands on the same articulator. The strategy implemented here "sums," as it were, the different requirements of different orosensory goals to create one composite motor goal.

According to Perkell, it is this strategy that, with minor adjustments, enables compensation for bite blocks or artificial palates (e.g., Hamlet and Stone, 1976). However, because this phase is still in the planning realm, it cannot generate the on-line compensations; these must occur during execution.

A final aspect of planning introduces coarticulatory influences. Presumably, this strategy adjusts the timing specifications determined earlier to accomplish anticipatory look-ahead of "noncompeting aspects of articulation." A reason for look-ahead is to prevent abrupt, and hence needlessly effortful, gestures. "Urgencies," which specify how soon a motor goal has to be reached, are assigned to motor goals during this planning phase.

In the final planning phase of speech production, motor goals are transformed into motor commands. Perkell rejects the idea that the commands selected based on motor goals are fully elaborated, as they would be if speech were fully regulated centrally. Instead he proposes that

the strategy for converting motor goals to commands involves internal feedback (that is, feedback entirely within the central nervous system) and possibly peripheral feedback as well. Internal feedback is used in something like a predictive simulation (Lindblom et al., 1979) that enables appropriate motor commands to be computed. A hypothetical "higher motor center" (see Figure 6–1*b*) receives moment-to-moment information via internal feedback about the state of the vocal tract. This is necessary to compute motor commands that will realize motor goals. The internal feedback is supplied by a "lower motor center" that represents the state of the vocal tract to the nervous system. Premotor commands are selected based on the discrepancy between the state of the vocal tract and the state required by the next motor goal. The lower motor center then translates premotor commands into commands to individual muscles. Presumably it is at this stage that on-line compensations occur.

Evaluation

The model is useful in offering an account of most of what a speech production theory must handle. In particular, it provides an account of the three phenomena identified earlier as central: durational patternings, coarticulation, and motor equivalence. The model requires revision or elaboration along three lines.

First, the planning component of the model clearly was not offered as a *minimal* specification of what must be "computed" during speech planning. The planning components of the present model overlap little or not at all with the planning specified in language production models, yet they are complex and include several stages. A job for future theoretical efforts will be to reduce the hypothetical computational component of speech production.

Second, and perhaps relatedly, the concept of coordination—in particular, the observation that speech is a highly coordinated activity—has little significance in the model. Yet (as is argued later), it may be the most important fact of activity on which researchers should focus their efforts; indeed, incorporating coordination into a speech production model may enable substantial reduction of the computational component. In particular, it can help to explain in one stroke how it is possible for talkers to regulate as many separate vocal structures and muscles as they can (the "degrees of freedom" problem; Bernstein, 1967; Turvey, 1977), how talkers can restrict themselves only to performance of coordinated vocal activity (Weiss, 1941), and concomitantly, how talkers can engage in performances that are at once physical (activities of the vocal tract) and "cognitive" or "mental" (utterance of linguistic units). This last question is the problem with which the chapter was introduced and which is addressed in the present subsection on speech production. The concept of coordination is further in the section that follows.

Third, the model does not address the distinction between linguistic units and metrical structures that emerges both in research on language production and from linguistic analysis. It was proposed earlier that metrical structures may require explanation in a theory of speech production more than in one of language production. An explanation is not yet at hand, but one direction in which to search is proposed in the section entitled Cyclicity in Behavior later in this chapter.

Coordination

Coordination is a fundamental property of speech and other biological activities. Understanding what coordination is and how it is achieved may be essential to understanding both the regulation of speech production specifically and, more generally, the sense in which vocal structures can be said to realize linguistic units.

Coordination is not well understood, however. According to Weiss (1941), nearly exclusive focusing on the properties of individual neurons and neural transmission has led to

> the neglect of the problem of how transmission has come to be so discriminatory and selective as to lead to coordinated responses rather than to unorganized convulsions. (p. 3)

More recently, and speaking more generally about biological organization, Pattee (1976) remarked:

> However, in spite of our knowledge of the "palpable detail" which is said to be normal chemistry for all known cellular reactions, the origin and nature of the *coordination* of these reactions remains an obscure and evasive question.

That biological activity is coordinated implies two related consequences for activity and its regulation. First there is a selective loss of degrees of freedom in the regulated physical system. If two variables of a system are coordinated, then they are not independent; changes in the value of one variable imply corresponding changes in the value of the other. Therefore, only some of the conceivable pairings of values of the coordinated variables wil occur—namely, only those compatible with the nature of the coordinative relations between them. For example, if the variables in question are the positions of the two front wheels of a car, then the only pairings of values of the variables that will occur are pairings in which the values are the same. In the example, and generally, loss in

the number of possible outcomes in a system owing to coordinative relations among variables is advantageous. The driver happily gives up independent control over the front wheels of a car because he or she never wants to turn the wheels in different directions. Moreover, turning them in the same direction to the same degree is easier if the wheels are coordinated by an axle than if they require separate control.

Coordination, then, is selective loss of independence of the parts of a system, with the result that unwanted outcomes are avoided. At the same time, the task of regulation is made easier because fewer independent choices need to be made and enforced.

A corollary of this may be that coordination enables avoidance of certain errors. In the introduction to this chapter, it was pointed out that walkers do not make global ordering errors. For example, they never take two steps with the right foot without stepping with the left foot in between. Errors *do* occur in nonverbal activities, however (e.g., Norman, 1981). Errors are avoided in locomotion because functional and physical linkages essentially enforce the intended orderings of events. Errors arise, it seems, where sequences of actions are not physically coordinated, or, perhaps equivalently, where the intended sequencing of actions is arbitrary with respect to the system that realizes it.

If misorderings of phonological segments occur during "planning" rather than during execution of speech, as most speech-errors researchers claim, perhaps that is because during planning the ordering of segments *is* arbitrary with respect to the implementing physical (neural) system, whereas it is no longer arbitrary when the vocal tract becomes organized to implement the segments.

Coordination has another consequence that is implied in the quotation from Weiss and that is explicitly studied by Pattee (1973; 1976; 1977; see also Polanyi, 1962). Coordination gives rise to an abstact level of description and functioning in the coordinated system. If there were no coordination in the motor system, then Weiss's unorganized convulsions would be a probable outcome of motor activity. In addition, such an outcome would not be distinguished from any apparently "coordinated" gestures that might occasionally eventuate. (Compare the monkey at the typewriter who occasionally types a letter sequence that people recognize as a word; from the monkey's perspective, there is nothing special about the word.) In another example, if there were no grammatical constraints on word order in English, then no sequence of words in the language would be special or distinguished from others and no sequence would have the superordinate meaning that sentences have as compared to random word strings (Pattee, 1976). By restricting the outcomes of a system to a principled subset of all conceivable ones, coordination creates a new, more abstract level of functioning in the system embodying it. The result is functional activity if the system is a motor system and meaningful sentences if it is also linguistic.

This consequence of coordination—that of creating what Pattee (1973) calls an "alternative description" of a biological system—is exactly what is needed if populations of neurons or if vocal tracts are to realize linguistic units in the way depicted in Figure 6-8*b*. Linguistic units are the alternative, more abstract, description of the vocal tract producing speech. The problems for a theory of speech production, from this perspective, are to identify the coordinations involved in speech and thereby explain in part, how alternative descriptions can be planned and realized in a vocal tract.

Coordinative Structures and Their Properties

"Coordinative structures" are functionally specific units of action defined over groups of muscles and articulator degrees of freedom. Their components are constrained to effect coherent activity. Activities characteristic of coordinative structure regulation have two general characteristics by which they can be identified (Kelso, Tuller, and Harris, 1983). First, their constituent muscle activity or articulator-joint movement exhibits both invariant and variable properties. In particular, over changes in rate of production, or sometimes amplitude of movement, the relative timing of muscle activations remains invariant while the magnitude of muscle activity varies. This means of changing rate is characteristic of locomotion (e.g., Grillner, 1975), handwriting (Viviani and Terzuolo, 1980), and typing (Terzuolo and Viviani, 1979). It may be characteristic of speech as well. Tuller, Kelso, and Harris (1982) reported invariant relative timing of V_1 and C_2 related muscle activity in $C_1V_1C_2V_2C_3$ sequences varying in both rate and stress pattern.

The separation of invariant and variant properties of the coordinative structure suggests a separation of "coordination," responsible for the global form that an activity will take, and "control," responsible for its specific character (Kugler, Kelso, and Turvey, 1980).

A second property of coordinative structure regulated activities is that they interact in various ways with other such activities to form larger functional systems. This is clearly evident in von Holst's classic studies of fish fin activity (1973). It is also characteristic of the limbs in locomotion (Shik and Orlovskii, 1965). Again, it may be characteristic of speech as well. Kelso, Tuller, and Harris (1983) report that production of a repeated syllable interacts with concurrent finger tapping. A talker is asked to produce syllables at a constant amplitude and rate, but to tap his or her finger with alternating long and short finger excursions. The result is an

apparent coupling of the two activities—long finger movements are accompanied by higher amplitude syllables than short movements of the finger. Similarly, asked to produce alternating stressed and unstressed versions of a syllable, but to tap evenly, taps accompanying a stressed syllable have a greater amplitude than those accompanying an unstressed syllable. This outcome, it seems to me, it is not predicted by a theory of speech production in which muscles are controlled by independent motor commands.

How is Coordination Achieved?

The example of a car and its axle was used earlier to illustrate a benefit of selective degrees of freedom loss. The example is misleading, however, because in speech activity, coordinative coupling of vocal structures is transient.

How is activity coordinated when the particulars of the coordinations apparently undergo moment-to-moment revision? One possibility is that peripheral reflexes are selectively potentiated and disabled in the course of talking. There is evidence of this in regulation of activities other than speech (Fukuda, 1961; Gottlieb, Agawarl, and Stark, 1970; Grillner, 1975) and some in the speech domain as well.

McClean, Folkins, and Larson (1979) have studied a possible role of the perioral reflex in speech. The perioral reflex is elicited in the laboratory by stretch of the lip or by electrical stimulation. It has a short latency response (10 to 15 ms) and a longer latency response (35 ms). Both McClean (1978) and Netsell and Abbs (1975) have shown an increase in the amplitude of the perioral reflex recorded from the orbicularis oris muscle in the latent interval before muscle activation for a bilabial consonant. In addition, Netsell and Abbs reported a suppression of the reflex during production of /a/ in /pa/. This outcome is similar to outcomes reported for other skilled activities. As reported in this literature, reflexes are potentiated when their action will promote a voluntary movement. Activating reflex has the consequence, in addition, of suppressing excitability of reflexes with action antagonistic to the intended movement. That the perioral reflex itself is inhibited during /a/ would seem to imply a reciprocal relationship between it and antagonistic reflexes.

However, there is reason to doubt that these data should be read in this way. Recently Abbs and Cole (1982) have expressed strong doubt that brainstem reflexes (having latencies in the range of that of the perioral reflex) play an important role in speech. They point out that if the perioral reflex were recruited for speech, several consequences should be realized:

it should inhibit its antagonists; effects of stimulating it should be local; and the reflex itself should be responsive only to stimuli generating local movement. According to Abbs and Cole, none of these outcomes is obtained. Actions antagonistic to those promoted by the perioral reflex are excited by stimuli eliciting the reflex. Loading and unloading the orbicularis oris muscle both give the same excitatory response. In addition, the lip stretch that elicits the perioral reflex also gives rise to responses in distant facial and neck muscles. Finally, excitatory oris responses can be elicited by distant stimuli.

Most telling, perhaps, according to Abbs and Cole, the perioral reflex is not very sensitive to movement velocities in the speech range. They propose that the reflex is part of a generalized response to aversive or potentially injurious stimuli.

A more promising source of information concerning coordination in speech may be offered by studies of temporary, functional coordinations among articulators as studied by Abbs and his colleagues (Folkins and Abbs, 1975, 1976; see also Kelso et al., 1982). Specifically, a talker can achieve the "orosensory goal" of bilabial closure in a variety of ways involving different contributions of the jaw and the two lips. Indeed, in different syllable contexts, talkers will exhibit different relative contributions of these three structures (Sussman et al., 1973). For example, in the context of a high vowel, the jaw will contribute relatively more to closure of a bilabial consonant than in the context of a low vowel. This "context-conditioned variability" can be viewed in a variety of ways, as discussed earlier, but so far as efficiency of regulation is concerned, an efficient way to ensure bilabial closure but allow contextual influences would be to establish a coordinated relationship among the articulators so that they bear, in effect, a negative relationship to each other. Any decrease in the contribution of the jaw, then, by virtue of the jaw's relationship to the lips would give rise to a corresponding increase in the contribution of the upper or lower lip, or both.

This, in fact, appears to characterize the jaw-lip relationship during bilabial consonant production (Folkins and Abbs, 1975, 1976; Kelso et al., 1982). Unexpected jaw loading during closing for a bilabial stop gives rise to compensatory lip movement so that closure is achieved. The same compensatory relationship is observed both during unperturbed repeated productions of bilabial consonants in a vocalic context (Hughes and Abbs, 1976; but see Sussman, 1980), and, as already noted, during production of bilabial consonants in the context of vowels varying in height.

In contrast to the general excitability of the perioral reflex, excitatory lip activity is not observed to the same degree if jaw closing is impeded during production of a nonlabial consonant (Kelso, Tuller, and Fowler, 1982). That is, for example, if jaw closing for /z/ in /baez/ is impeded,

the upper lip does not show the same magnitude of increase in excitation that it shows when jaw closing for final /b/ in /baeb/ is impeded. This suggests a temporary coupling of jaw and lips established only when bilabial closure is a goal of articulatory gestures.

Abbs and Kennedy (1982) refer to this mode of control as "open-loop, feed-forward" control. It is open loop in the sense that afferent information about jaw postion appears not to feed back to jaw closing muscles as it would if the system were "closed loop." Instead, it feeds "forward" to other articulators—here, the lips—thereby coordinating the activities of the different articulators for the achievement of an articulatory goal.

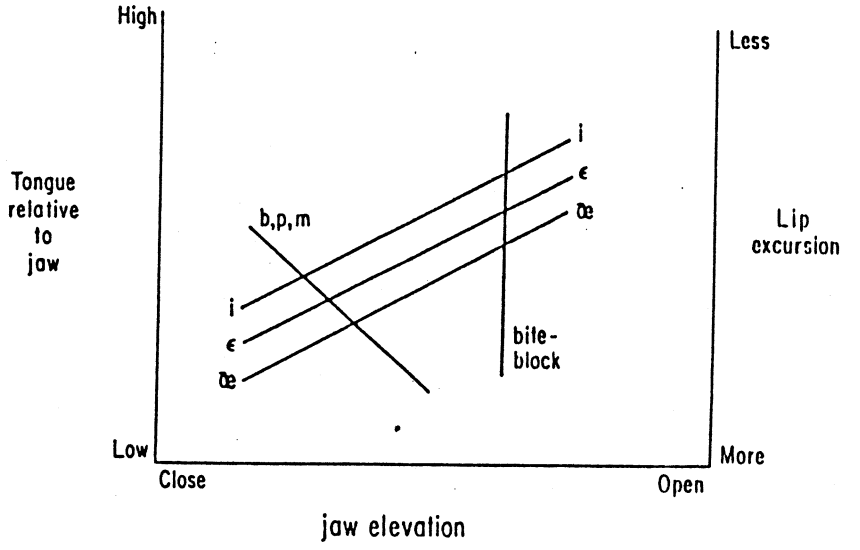
Similar couplings between jaw and tongue for vowels and lingual consonants have not been studied systematically. However, Chuang, Abbs, and Netsell (1978) find negative correlations between jaw and tongue positions in repeated productions of some vowels. As for lingual consonants, Kelso and colleagues (1982) find on-line compensatory activity of the genioglossus to unpredicted jaw perturbations during closure for /z/ in /baez/, but not during closing for final /b/ in /baeb/.

These coordinations of articulators to achieve a common outcome also achieve the selective loss of degrees of freedom characteristic of a coordinated system. Moreover, by interacting, the jaw-lip and jaw-tongue systems may further reduce controlled degrees of freedom. Figure 6-14 is a schematic illustration of this idea. It plots the jaw-lip-tongue relationship suggested by the study of Sussman and colleagues (1973) and by the studies of compensatory behavior just described. The figure indicates that jaw height during bilabial closure varies due to vocalic context. This implies that neighboring (coarticulated) segments help to choose which particular values of jaw height and lip positions will occur in production of a particular bilabial-consonant token. In addition, however, the jaw-height requirements of the consonant will help to select the relative contribution of the jaw and tongue movements to vowel production. It is as if the bilabial consonant coordinate structure and the vowel coordinative structure share control for a period of time over a common set of articulatory variables. Because both sets of functional coordinative relationships must hold during overlapping time slots, each constrains the values that the shared variables will take. To the extent that there is a unique or nearly unique value of the shared variables that allow both coordinative relationships to hold, the number of controlled degrees of freedom in the system is further reduced.

Cyclicity in Behavior

Research on language production and that on systematic properties of linguistic utterances both uncover "metrical structures." These are

Figure 6-14. Schematic illustration of jaw, lip and tongue coordination for production of a bilabial consonant/vowel syllable, and for production of bite-block speech. See the text for elaboration.



remarkable in two respects as already discussed. They are not necessarily coextensive with linguistic units, and they suggest underlying cyclicity in talking.

I have not found any rationale for these structures in language itself; yet the literature on speech production has suffered no insights, either. In this final section of the paper, the question is asked whether a rationale could be developed within the study of speech production.

Research findings on performance of cyclic motor skills other than speech recently have been interpreted in the light of the physical theory of the dynamics of "open" physical systems (e.g., Kugler, Kelso, and Turvey, 1982). For reasons to be discussed shortly, such physical systems—which include biological ones—are inherently cyclical in nature. The following section examines this view of physical systems in relation to the performance of skilled activity. (The discussion is taken from Kelso, Tuller, and Harris [1983] and Kelso, Holt, Rubin, and Kugler (1981), who provide more detail.) Next the relevance that this theory may have for understanding metrical structures in speech is considered.

According to Kelso and associates (1981), until recently, the physical theory of dynamics has had little to offer in the way of explaining biological activity because it has dealt largely with "closed" physical systems. Closed

systems do not exchange energy and matter with their environments. Whereas closed systems very evidently "obey" the second law of thermodynamics—that is, they move toward states of decreasing organization—open systems, including biological systems, "accumulate negentropy" (Schrodinger, 1945). Thereby they are able to develop and maintain their organized forms and functions over long periods of time.

Open systems maintain their forms and functions by offsetting energy losses with periodic energy gains. Energy flow into and out of the system is a chief organizing property of living systems; moreover, tracking the flow of energy from "source" to "sink" identifies one cycle in a necessarily cyclic life style of the system (Morowitz, 1978, cited in Kelso et al., 1981). How is the cyclicity realized? To illustrate cyclicity in living systems, Kelso and associates (1981; see also Kelso et al., 1983) begin by describing a familiar oscillatory system: a linear damped mass-spring system. This system is described by the following equation:

$$m\ddot{x} + c\dot{x} + kx = 0$$

(where m is mass, k is stiffness, and c is a frictional component; x, $\dot{x}$, and $\ddot{x}$, refer to displacement, velocity, and acceleration, respectively). Set in motion by a displacement, its oscillatory motion decays over time as the second law of thermodynamics predicts. To prevent decay, the system has to be supplied with an energy source:

$m\ddot{x} + cx + k\dot{x} = F(\Theta)$ (where k is stiffness, F is force, $F(\Theta)$ is a forcing function, and other variables are as above).

To be usable, the energy source has to be tapped by the system at the proper phase in the cycle. In living systems, and in some nonliving systems, this requirement is met by an "escapement"—a nonlinear element that taps an energy source and injects it into the system at the proper phase in its cycle. (Pendulum clocks have escapements which ensure that potential energy from a hanging weight is delivered to the swinging pendulum only in the middle of its arc where the pendulum is most effectively influenced by outside forces; see Fitch, Tuller, and Turvey, 1982.) An example of the escapement principle in skilled activity is provided by research or Orlovskii (1972). Orlovskii found that stimulation of the red nucleus of the cat excited flexion in the resting limb. Stimulation of Deiter's nucleus excited extension. During locomotion, however, continuous stimulation of the red (Deiter's) nucleus energized flexion (extension) only during the swing (stance) phase of the stepping cycle. In short, the supraspinal "energy source" supplied by stimulation of the red and Deiter's nuclei was tapped by the spinal locomotion system only during appropriate phases of the stepping cycle.

Nonlinear oscillators with the escapement property are called "limit cycle oscillators" and they are believed to characterize systems such as living ones that maintain themselves consistently far from thermodynamic equilibrium (e.g., Yates, 1980; Yates and Iberall, 1973; and others cited in Kelso et al., 1981).

According to Kelso and associates, limit cycle oscillators have at least some of the characteristics exhibited by muscle systems engaged in functional activity. In addition to their cyclicity and the escapement property already discussed, two other characteristics are salient. First, because of the escapement feature, "power" and timing are independent. That is, for example, in a pendulum clock, increasing the amount of energy that is injected into the system in the middle of a pendulum swing does not affect *when* in the cycle the injection occurs. The same feature, invariant relative timing, over changes in amplitude of muscle activation has already been described for speech and other activities. Second, limit cycles exhibit at least some varieties of "equifinality." Their cycles are stable and return to normal very soon after perturbation.

What, if anything, can the concept of limit cycle contribute to understanding metrical structures in speech? An answer will have to await more careful study and investigation. However, the concept is promising because it rationalizes cyclicity in activity and that is partly the aspect of metrical structures that needs understanding.

The "breath group," and ideally, the domain of declination, appear to be natural "cyclers" of the limit cycle variety. From this perspective, inspiration provides a source of potential energy used during utterance production. For different reasons, a syllable also appears subsumable under the idea of a limit cycle mode of functioning (see Kelso and Bateson, 1983, for an elaboration of this idea). Finally, stressing may be a reflection of energy injection into the articulators. An idea compatible with this concept is that of Catford (1977), who has suggested that "isochrony" of the stress foot may not be isochrony in fact, but rather "isodynamism"—an injection of about the same amount of "initiator power" (that is, work done per unit time by a pulmonic pressure pulse) for each foot. This idea does not shed light on why the foot is the domain of the pressure pulses. Conceivably, the principle is simply one of alternation—one pulse approximately every other syllable.

CONCLUDING REMARKS

The intent of this chapter was to review most of what researchers are researching and writing about language and speech production. Several

concluding observations are suggested by the review.

1. Researchers have devoted little attention to the relation of speech to language. This has allowed models of each to be developed that have limited bearing one on the other and that together posit, in my view, implausibly many computational planning and execution stages of production. In addition, the default view of the relation of speech to language suggested by researchers—namely, that planned linguistic units are *translated* into articulatory gestures—raises what seem to be insurmountable difficulties that can be avoided if the view of the relationship of speech to language depicted in Figure 6-8*b* is adopted.

2. Figure 6-8*b* allows for physical systems, by virtue of their superordinate organization, to embody psychological (cognitive, mental) functions. In my view, an important avenue for future research is one of understanding this characteristic of complex living systems, and of seeing how it is exploited to reduce the number of computationally or representationally controlled aspects of talking.

3. Research, theorizing, and modeling in speech and language production have been somewhat narrow in scope. If a language production researcher studies and models the structure of pausing in language production, he or she is unlikely to constrain theorizing and modeling by reference to data on spontaneous errors of speech, for example, or even to data on durational shortening in speech production. Collectively, researchers in the fields of language and speech production know a great deal about the superficial systematic properties of utterances. It becomes important now to attempt a realistic comprehensive theory of their underlying causes.[1]

FOOTNOTES

[1]This research was supported by NSF Grant BNS 8111470 and by NICHD Grant HD 16591-01 to Haskins Laboratories. I thank Elliot Saltzman and George Wolford for their comments on parts of the manuscript.

REFERENCES

Abbs, J., and Cole, K. (1982). Considerations of bulbar and suprabulbar afferent influences upon speech motor coordination and programming. In S. Grillner, B. Lindblom, J. Lubker, and A. Persson (Eds.), *Speech motor control.* Oxford: Pergamon.

Abbs, J., and Kennedy, J. (1982). Neurophysiological processes of speech movement control. In N. Lass (Ed.), *Speech, language and hearing* (Vol. 1). Philadelphia: Saunders.

Abercrombie, D. (1964). Syllable quantity and enclitics in English. In D. Abercrombie, D. Fry, P. MacCarthy, N. Scott, and J. Trim (Eds.), *In honour of Daniel Jones.* London: Longman.

Baars, B. (1980). On eliciting predictable speech errors in the laboratory. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand.* New York: Academic Press.

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

Barry, W., and Kuenzel, H. (1975). Co-articulatory airflow characteristics of intervocalic voiceless plosives. *Journal of Phonetics, 3,* 263-282.

Bell-Berti, F. (1980). Velopharyngeal function: A spatial-temporal model. In N. Lass (Ed.), *Advances in basic research and practice.* New York: Academic Press.

Bell-Berti, F., and Harris, K. (1976). Some aspects of coarticulation. *Haskins Laboratories Status Report on Speech Research, SR 45/46,* 197-204.

Bell-Berti, F., and Harris, K. (1979). Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America,* 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. A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica, 30,* 41-55.

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

Bladon, A., and Al-Bamerni, A. (1982a). One-stage and two-stage temporal patterns of velar coarticulation. Paper presented to the Acoustical Society of America, Orlando, FL.

Bladon, A., and Al-Bamerni, A. (1982b). Nasal coarticulation of pharyngeal and glottal consonants: A deductive account. Paper presented to the Acoustical Society of America, Orlando, FL.

Bolinger, D. (1963). Length, vowel, juncture. *Linguistics, 1,* 1-29.

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

Butcher, A., and Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics, 4,* 59-74.

Carney, P., and Moll, K. (1971). A cinefluorographic investigation of fricative-consonant vowel coarticulation. *Phonetica, 23,* 193-201.

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

Chuang, C. K., Abbs, J., and Netsell, R. (1978). Possible role of tongue-hard palate contact in vowel production. *Journal of the Acoustical Society of America, 63,* Supplement 1.

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

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-Verlag.

Cutler, A. (Ed.) (1982). *Slips of the tongue.* The Hague: Mouton.

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

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., and Reich, P. (1980). Toward a unified model of slips of the tongue. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand.* New York: Academic Press.

Dell, G., and Reich, P. (1981). Stages in speech production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20,* 611-629.

Donegan, P., and Stampe, D. (1979). The study of natural phonology. In D. Dinnsen (Ed.), *Current approaches to phonological theory.* Bloomington, IN: Indiana University Press.

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.

Fitch, H., Tuller, B., and Turvey, M. T. (1982). The Bernstein perspective. III. Tuning of coordinative structures with special reference to perception. In J. A. S. Kelso (Ed.), *Human motor behavior: An introduction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

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,* 113–133.

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,* 127–139.

Fowler, C., Rubin, P., Remez, R., and Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production,* Vol. 1. London: Academic Press.

Fowler, C., and Turvey, M. T. (1980). Immediate compensation for biteblock speech. *Phonetica, 37,* 306–326.

Freud, S. (1958). *The psychopathology of everyday life.* New York: New American Library [1901].

Fromkin, V. (Ed.) (1973). *Speech errors as linguistic evidence.* The Hague: Mouton.

Fromkin, V. (Ed.) (1980). *Errors in linguistic performance: Slips of the tongue, ear, pen and hand.* London: Academic Press.

Fujimura, O. (1980). Elementary gestures and temporal organization: What does articulatory constraint mean? In T. Myers, J. Laver, and J. Anderson (Eds.), *The cognitive representation of speech.* Amsterdam: North-Holland.

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

Fukuda, T. (1961). Studies on human dynamic postures from the viewpoint of postural reflexes. *Acta Otolaryngologica* Supplement 161.

Garding, E. (1979). Sentence intonation in Swedish. *Phonetica, 36,* 207–215.

Garrett, M. (1975). The analysis of sentence production. In G. Bower (Ed.), *The psychology of learning and motivation,* Vol. 9. New York: Academic Press.

Garrett, M. (1976). Syntactic processes in sentence production. In R. J. Wales and E. Walker (Eds.), *New approaches to language mechanisms.* Amsterdam: North-Holland.

Garrett, M. (1980a). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production,* Vol. I. London: Academic Press.

Garrett, M. (1980b). The limits of accommodation: Arguments for independent levels in sentence production. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand.* London: Academic Press.

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

Gottleib, G., Agawarl, G., and Stark, L. (1970). Interactions between voluntary and postural mechanisms of the human motor system. *Journal of Neurophysiology. 33,* 365–381.

Grillner, S. (1975). Locomotion in vertebrates. *Physiological Reviews, 55,* 247–304.

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

Hamlet, S., and Stone, M. (1976). Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses. *Journal of Phonetics, 4,* 199–218.

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.

Henke, W. (1966). *Dynamic articulatory model of speech production using computer simulation.* PhD thesis, Massachusetts Institute of Technology, Cambridge.

Hockett, C. (1960). The origin of speech. *Scientific American, 203,* 89–96.

Huggins, A. W. F. (1975). On isochrony and speech. In G. Fant and M. Tatham (Eds.), *Auditory analysis and perception of speech.* London: Academic Press.

Huggins, A. W. F. (1978). Speech timing and intelligibility. In J. Requin (Ed.), *Attention and performance,* Vol 7. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hughes, O., and Abbs, J. (1976). Labial mandibular coordination in the production of speech: Implications for motor equivalence. *Phonetica, 33,* 199–221.

Kelso, J. A. S., and Bateson, E. (1983). On the cyclical basis of speech production. *Journal of the Acoustical Society of America, 73,* Supplement 1, S67.

Kelso, J. A. S., Holt, K., Rubin, P., and Kugler, P. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear oscillators: Theory and data. *Journal of Motor Behavior. 13,* 226–261.

Kelso, J. A. S., Tuller, B., and Fowler, C. (1982). The functional specificity of articulatory control and coordination. *Journal of the Acoustical Society of America, 72,* S103.

Kelso, J. A. S., Tuller, B., and Harris, K. (1983). A "dynamic pattern" perspective on the control and coordination of movement. In P. MacNeilage (Ed.), *The production of speech.* New York: Springer-Verlag.

Kent, R. (1983). The segmental organization of speech. In P. MacNeilage (Ed.), *The production of speech.* New York: Springer-Verlag.

Kent, R., Carney, P., and Severeid, L. (1974). Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research, 17,* 470–488.

Kent, R., and Minifie, F. (1977). Coarticulation in recent speech production models. *Journal of Phonetics, 5,* 115–117.

Kent, R, and Moll, K. (1972). Tongue body articulation during vowel and diphthong gestures. *Folia Phoniatrica, 24,* 286–300.

Kenstowicz, M., and Kisseberth, C. (1979). *Generative phonology: Description and theory.* New York: Academic Press.

Klatt, D. (1976). Linguistic uses of segment duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America,* 1976, *59,* 1208–1221.

Klatt, D. (1980). Speech perception: A model of acoustic-perceptual analysis and lexical access. In R. Cole (Ed.), *Perception and production of fluent speech.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Klima, E., and Bellugi, U. (1979). *The signs of language.* Cambridge, MA: Harvard University Press.

Kuehn, D., and Moll, K. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics, 4,* 303–320.

Kugler, P., Kelso, J. A. S., and Turvey, M. T. (1980). On the concept of coordination as dissipative structure. I. Theoretical lines of convergence. In G. Stelmach and J. Requin (Eds.), *Tutorials in motor behavior.* Amsterdam: Elsevier/North-Holland.

Kugler, P., Kelso, J. A. S., and Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso and J. Clark (Eds.), *The development of movement control and coordination.* New York: Wiley.

Kupin, J. (1979). *Tongue twisters as a source of information about speech production.* PhD thesis, University of Connecticut, Storrs.

Lamb, S. (1966). *Outline of stratificational grammar.* Washington, DC: Georgetown University Press, 1966.

Lass, N., and Davis, M (1976). An investigation of speaker height and weight identification. *Journal of the Acoustical Society of America, 60,* 700–703.

Lehiste, I. (1980). Phonetic manifestations of syntactic structure in English. *Bulletin of the Research Institute of Logopedics and Phoniatrics, 14,* 1–27.

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

Lindblom, B. (1971). Phonetics and the description of language. *Seventh International Congress of Phonetic Sciences.* The Hague: Mouton.

Lindblom, B., Lubker, J., & 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., Lyberg, B., & Holmgren K. (1981). *Durational patterns of Swedish phonology: Do they reflect short-term memory processes?* Bloomington, IN: Indiana University Linguistics Club.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Papers in Linguistics from the University of Stockholm, 21,* 1–59.

Lockwood, D. (1972). *Introduction to stratificational linguistics.* New York: Harcourt, Brace, Jovanovich.

Lyberg, B. (1979). Final lengthening—partly a consequence of restrictions on the speech of fundamental frequency change? *Journal of Phonetics, 7,* 187–196.

Lyberg, B. (1981). Temporal properties of spoken Swedish. *Monographs in linguistics from the University of Stockholm,* Vol. 6.

MacKay, D. (1982). The problems of flexibility, fluency and speed-accuracy tradeoff. *Psychological Review, 89,* 483–506.

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

MacNeilage, P., & DeClerk, J. (1969). On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America, 45,* 1217–1233.

MacNeilage, P., & Ladefoged, P. (1976). The production of speech and language. In E. C. Carterette & M. Friedman (Eds.), *Handbook of perception: Language and speech.* New York: Academic Press.

Maeda, S. (1976). *A characterization of American English intonation.* PhD thesis, Massachusetts Institute of Technology, Cambridge.

McClean, M. (1978). Variation in the perioral reflex amplitude prior to lip muscle contraction for speech. *Journal of Speech and Hearing Research, 21,* 276–284.

McClean, M., Folkins, J., and Larson, C. (1979). The role of the perioral reflex in lip motor control. *Brain and Language, 7,* 42–61.

McClelland, J., and Rumelhart, D. (1981). An interactive activation model of context effects in letter perception. *Psychological Review, 88,* 375–407.

Merringer, R., and Meyer, K. (1895). *Versprechen und Verlesen: Eine Psycholigisch-Linguistiche Studie.* Stuttgart: Goschensche Verlagsbuchhandlung. (Reissued: Amsterdam: John Benjamin, 1978).

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.

Morowitz, H. J. (1978). *Foundations of bioenergetics.* New York: Academic Press.

Netsell, R., & Abbs, J. (1975). Modulations of perioral sensitivity during speech movements. Paper presented to the Acoustical Society of America, San Francisco.

Norman, D. (1980). Copycat science or Does the mind really work by table look-up? In R. Cole (Ed.), *Perception and production of fluent speech.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Norman, D. (1981). Categorization of action slips. *Psychological Review, 88,* 1–15.

Ohala, J. (1971). Monitoring soft palate movements in speech. *Project on linguistic analysis reports.* (Phonology Laboratory, Department of Linguistics, University of California, Berkeley), 13, JO1–J05.

Ohala, J. (1981). The listener as a source of sound change. In M. F. Miller (Ed.), *Papers from the parasession on language behavior.* Chicago: Chicago Linguistic Association.

Ohala, J., & Ewan, W. (1973). Speed of pitch change. *Journal of the Acoustical Society of America, 53,* 345(A).

Ohman, S. (1966). Coarticulation in VCV utterances: Spectrographic measures. *Journal of the Acoustical Society of America, 39,* 151–168.

Orlovskii, G. (1972). The effect of different descending systems on flexion and extensor activity during locomotion. *Brain Research, 40,* 359–371.

Pattee, H. H. (1973). The physical bases and origin of hierarchical control. In H. H. Pattee (Ed.), *Hierarchy theory: The challenge of complex systems.* New York: Braziller.

Pattee, H. H. (1976). Physical theories of biological coordination. In M. Grene and E. Mendelsohn (Eds.), *Topics in the philosophy of biology.* Dordrecht, Holland: Reidel.

Pattee, H. H. (1977). Dynamic and linguistic modes of complex systems. *International Journal of Complex Systems, 3,* 259–266.

Perkell, J. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study.* Cambridge, MA: MIT Press.

Perkell, J. (1980). Phonetic features and the physiology of speech production. In B. Butterworth (Ed.), *Language production,* Vol. I. London: Academic Press.

Polanyi, M. (1962). *Personal knowledge.* Chicago: University of Chicago Press.

Prince, A. (1983). Relating to the grid. *Linguistic Inquiry,* 19–100.

Reich, P. (1970). *A relational-network model of language behavior.* PhD thesis, University of Michigan, Ann Arbor.

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

Ryle, G. (1949). *The concept of mind.* New York: Barnes & Noble.

Sampson, G. (1980). *Schools of linguistics.* Stanford, CA: Stanford University Press.

Scherer, K. (1981). Speech and emotional states. In J. Darley (Ed.), *The evaluation of speech in psychiatry.* New York: Grune & Stratton.

Scherer, K. (1982). Methods of research on vocal communication: Paradigms and parameters. In K. Scherer and P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research.* Cambridge: Cambridge University Press.

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.

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

Shik, M., and Orlovskii, G. (1965). Coordination of the limbs during running of the dog. *Biophysics, 10,* 1148–1159.

Shrodinger, E. (1945). *What is life?* London: Cambridge University Press.

Simon, H. (1980). How to win at twenty questions with nature. In R. Cole (Ed.), *Perception and production of fluent speech.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Smith, E., & Medin, D. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press, 1981.

Sorenson, J., & Cooper, W. (1980). Syntactic coding of fundamental frequency in speech production. In R. Cole (Ed.), *Perception and production of fluent speech.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Stemberger, J. (1982). *The lexicon in [illegible]*

Sternberg, S., Monsell, S., Knoll, R., and Wright, C. (1978). The latency and duration of rapid movement sequences: Comparison of speech and typewriting. In G. Stelmach (Ed.), *Information processing in motor control and learning.* New York: Academic Press.

Sternberg, S., Wright, C., Knoll, R., and Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. Cole (Ed.), *Perception and production of fluent speech.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Stevens, K., & Blumstein, S. (1981). The search for invariant correlates of phonetic features. In P. Eimas and J. Miller (Eds.), *Perspectives on the study of speech.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Studdert-Kennedy, M. (1980). Language by hand and by eye: A review of Edward S. Klima and Ursula Bellugi's *The signs of language. Cognition, 8,* 93–108.

Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics, 7,* 71–79.

Sussman, H. (1980). Methodological problems in evaluating lip–jaw reciprocity as an index of motor equivalence. *Journal of Speech and Hearing Research, 23,* 699–702.

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

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

Tarttar, V. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics, 27,* 24–27.

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

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.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw and J. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Viviani, P., & Terzuolo, C. (1980). Space-time invariance in learned motor skills. In G. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior.* Amsterdam: Elsevier/North-Holland.

von Holst, E. (1973). *The behavioral physiology of animals and man: The collected papers of Erich von Holst.* London: Methuen (originally published in 1937).

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

Wickelgren, W. (1969). Auditory or articulatory coding in verbal short-term memory. *Psychological Review, 76,* 232–235.

Wickelgren, W. (1976). Phonetic coding and serial order. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception: Language and speech.* New York: Academic Press.

Williams, C., and Stevens, K. (1972). Emotions and speech: Acoustical correlates. *Journal of the Acoustical Society of America, 52,* 1238–1250.

Yates, F. (1980). Physical causality and brain theories. *American Journal of Physiology, 238,* R277–R290.

Yates, F., and Iberall, A. (1973). Temporal and hierarchical origin in biosystems. In J. Urquhart and F. Yates (Eds.), *Temporal aspects of therapeutics.* New York: Plenum.