14

Implications for Speech Production of A General Theory of Action

C. A. Fowler — *Dartmouth College and Haskins Laboratories*
P. Rubin — *Haskins Laboratories*
R. E. Remez — *Indiana University*
M. T. Turvey — *University of Connecticut and Haskins Laboratories*

I. Introduction

Phonetic and phonological segments have substantial linguistic/theoretical support as real and universal units of language systems. Quite naturally, investigators of speech production have taken these units to be constituents of a speaker/hearer's linguistic competence and have focused investigation on discovering their correlates in the articulatory and acoustic records of utterances. These searches have met with little success, however, and the recent literature betrays some disenchantment with this as an investigatory strategy. Several investigators have proposed as an alternative that production theory develop its own units and concepts (e.g. MacNeilage and Ladefoged, 1976; Moll *et al.*, 1976). These categories would be based on production records as interpreted from the perspective of theories of coordinated movement perhaps, but would be unbiased by *a priori* notions borrowed from linguistic theory as to what the units ought to be like.

Our theoretical approach is related to this, but is somewhat less radical. We agree that concepts general to the control of coordinated movement are relevant to understanding the control of the articulators and need incorporation into production theory (cf. Abbs and Eilenberg, 1976; Moll *et al.*, 1976). In addition, we agree that correlates of linguistic segmental units, as production theorists have described them, are absent in articulation. But we are not yet ready to agree, therefore, that investigations of speech production ought to be conducted without reference to linguistic segmental units.

This strategy is likely to exacerbate the difficulty of reconciling the units that a speaker/hearer is assumed to know with those that he uses. Furthermore, speech *perception* theorists are faced with the same kind of disparity (cf. Studdert-Kennedy, 1978; Pisoni, 1978). If they were to take an analogous tack to that proposed by the production theorists cited above, doubtless they would discover a third set of units irreconcilably distinct both from those of linguistic theory and those of production theory.

The strategy that we have adopted involves a reassessment of some of the properties of linguistic units, based in part on concepts general to the control of coordinated movement. Our aim is to discover some way to characterize these units that preserves their essential linguistic properties, but also allows them to be actualized *unaltered* in a vocal tract and in an acoustic signal.

As the production literature describes them, abstract linguistic units have three properties that lack articulatory correlates: they are discrete, static, and context-free. In contrast, units of production, whatever they may be, are essentially dynamic and continuous. Furthermore, to the extent that articulatory approximations of linguistic units are discovered, they appear to be context-adjusted.

We suspect, perhaps naively, that these incompatibilities are only apparent and derive from a misunderstanding of abstract linguistic units on the part of production theorists. Linguistic theory expresses what is systematic in language. In that sense, and only in that sense, it captures what a language-user knows about his language. It does not claim to specify what a language-user knows, *as he knows it*; thus it does not necessarily concern itself at all with dimensions of description of its constituents that can be realized either in a vocal tract or even a mind. (Whether or not this is a sensible strategy is not at issue here.)

For example, in reference to phonetic segments, linguistic theory characterizes each one in terms of its locus in a system of phonetic relationships. Thus, a given segment is assigned values (features) along all and only the dimensions of description that are relevant to specifying its relationship to other segments. (These dimensions are given articulation-or acoustic-based names, but it would not change the *abstract* system to relabel them, "north-south", "X_1-X_2", and so on.)

"Discrete/continuous" is not a phonetic dimension of description, and one can imagine two alternative reasons why it is not. One is that *all* phonetic segments are discrete, and featural descriptions only specify distinctive features. But we favor the other possibility that "discrete" is unspecified because it is an irrelevant dimension of description of an *abstract* linguistic segment.

In speech production, "discreteness" has to do with the spatiotemporal overlap of production units. Due to coarticulation, articulatory approximations to linguistic units are not discrete in this sense; rather they overlap considerably. However, this has nothing to do with the linguistic description of a segment. The latter does not concern itself with sequences of segments, but only with locating a vocalic segment in a vowel space and a consonantal segment in a consonant space. Nor is discreteness of segments relevant to a phonetic or phonological representation of a word or sentence. The essential *linguistic* property here is "sequence" or "serial order" (so that, for example, /æ k t/, /t æ k/ and /k æ t/ are distinct). Discreteness is the way in which *orthographic* or *phonological-symbolic* representations of these sequences preserve their serial order, but it is not the way that articulation preserves it. Nonetheless, serial order is preserved in *some* way in articulation since /æ k t/, /t æ k/ and /k æ t/ are differently produced.

Consider also the property "static". This, too, is a value along an irrelevant dimension as far as abstract linguistic units are concerned. Although the essential properties of linguistic segments can perhaps be realized in a static medium, as when language is written down, a dynamic realization need not be considered necessarily destructive to those properties.

We will consider the dimension context-free versus context-adjusted in some detail later. Here we simply suggest that linguistic segments as known and as uttered must have context-free or invariant properties. In order for a phonological category to *be* a category for a language-user, its essential properties must be context-free. Theories of coordinated movement offer a way of understanding how the superficial variability of articulation may have underlying invariance.

It has been our suggestion that the mismatch between abstract linguistic units and articulatory categories may only be an apparent one. Such a mismatch may arise when an immutable assignment of values (e.g. discrete, static) to linguistic units is assumed by production theorists. Designations such as these, however, are along dimensions (discrete/continuous, static/dynamic) to which these values as abstract linguistic units are, in fact, indifferent.

We do not believe that a theory of speech production can afford to disregard the systematicities among linguistic units as described by linguistic theory. Nor can it ignore, as linguistic theory may allege to, the fact that these systematicities are known in a particular way and are used in particular ways by speaker/hearers. The obligations of a production theory with respect to these units is to give them values on those dimensions of description that may be irrelevant to abstract linguistic theory, but that are central to the realization of the units as they are embodied in the knowledge and vocal tract of a talker.

Our suggestion is that linguistic units, as they are known to a language-user and as they are spoken and perceived by him, are: *qualitatively separate and serially ordered* (but not discrete), *dynamic*, and *context-free*. These dimensions and their values seem to us to do no violence to the essential *linguistic* properties of segments. Yet they permit the claim that there is no essential difference between the segments that a language-user knows and those that he speaks and hears.

Support for these proposals is derived from a particular perspective on coordinated movement (e.g. Bernstein, 1967; Greene, 1972; Turvey, 1977a). Shortly we will review the relevant information from that source, but first we would like to provide some negative support for our view that linguistic segments as known and uttered are compatible through a critique of the conventional view which says that they are not. We call proposals in the spirit of the conventional view "translation theories", and suggest that this label covers all, or virtually all, extant theories of speech production.

II. Translation Theories: an Evaluation

In contrast to the view that we have put forward, it is generally agreed that the distinction between abstract and actual segments cannot be avoided—that neither the linguistic nor the articulatory descriptions of speech segments is inaccurate and that the two are irreconcilably distinct. Each is said to characterize a different *level* of abstraction in the speech system and to be invoked at different *stages* in the production of speech (Daniloff and Hammarberg, 1973; Hammarberg, 1976; MacNeilage and Ladefoged, 1976). In particular, abstract phonological segments are pre-motor cognitive entities that thereby can serve both as outputs of perceptual processing and as inputs to an articulatory mechanism. Dynamic articulation is the final product of the articulatory mechanism which alters the properties of abstract segments to fit the demands of vocal tract production and of the ear (see below).

The translation process between a sequence of discrete, static, context-free segments that a talker has in mind to say and the continuous, dynamic, context-adjusted ones that he in fact utters is not a simple one. According to Liberman and Studdert-Kennedy (1978) (see also Liberman *et al.*, 1967) the translation involves a "drastic restructuring" of planned phonetic segments. They believe the restructuring to be necessary in order to match the cognitive requirements of communication with properties of the vocal tract and ear. In their view, were abstract segments reproduced literally in an utterance, listeners would be unable to comprehend the spoken message of a talker. One cognitive requirement of communication is that a spoken message be produced rapidly. This requirement has to do with the fact that the semantic information jointly provided by phonological segments is spread over long stretches of speech. Thus, in order for a hearer to extract a talker's semantic message, a meaningful chunk must be produced rapidly enough that its initial part is still available in the hearer's memory when the final part is produced. A second cognitive requirement is that the order of the segments is extractable from the acoustic signal so that, for example, such words as "tack", "cat" and "act" may be distinguished.

Vocal tracts, however, cannot produce discrete segments rapidly enough to satisfy the first cognitive requirement; nor can auditory systems extract the temporal order of rapidly produced isolated sounds (e.g. Warren, 1976). For these reasons, according to Liberman and Studdert-Kennedy, abstract linguistic segments are complexly restructured in articulation. The result is coarticulated speech. In a spoken utterance, information about each abstract segment is spread over several acoustic segments and each acoustic segment provides information about more than one abstract segment. In this way, more linguistic segments can be signaled in a given amount of time than could be possible were each segment temporally distinct from its neighbors.

Current theories of speech production differ from this account in detail, but are like this account in that they tend to be translation theories; that is, they acknowledge the incommensurability of abstract and actual segments and thereby are obligated to describe how the one transforms into the other in the processing surrounding speech production. What follows is a brief evaluation of the logic and adequacy of translation theories.

A. The Logical Status of Translation Theories

A decision that the differences between abstract and actual segments are irreducible invites complicated theories both of speech production and of speech perception where simpler ones might otherwise have sufficed. The translation theory's main explananda arise solely from the irreducibility claim. Were it the case that talkers had only one kind of sound segment to work with instead of two, then the aspects of speech production currently held to surround the translation process would no longer require explanation.

In addition, we should point out that the implications of the irreducibility claim extend also to theories of speech perception. Any transformation that takes discrete, static, context-free segments into continuous, dynamic, context-adjusted gestures of the vocal tract must be considered destructive of essential properties of the abstract segments. Therefore, a perceiver has to reconstruct the talker's phonological intent from an impoverished acoustic signal. (For a well-known and graphic description of the destructive effects of coarticulation on the acoustic signal, see Hockett's easter egg analogy (1955).)

In contrast, if one were to adopt a non-translation view of speech production, which supposed that the essential properties of phonological segments are preserved in an utterance, the possibility would remain that the acoustic products of vocal-tract gestures fully specify the speaker's phonological intent. That is, a theory of "direct" speech production leaves open the possibility of a theory of direct speech perception as well.

The particular point we wish to make with regard to theories of speech production and speech perception is an instance of a larger point concerning the construction of general theories of acting and perceiving. To illustrate this larger point we will focus our arguments on visual perception. For it is not unreasonable to claim that the problem of perception as defined and the proposed solutions to that problem both originate in, and receive continued impetus from, the attempts of past and present philosophers and psychologists to understand visual experience and its physical support. Moreover, it will be apparent that the received orientation to coordinated movement is conceptually consistent with the received orientation to visual perception.

It has become a matter of doctrine that the light to an eye is equivocal about properties of the environment (see Turvey and Shaw, 1979). More precisely, it is assumed that the laws which map environmental properties into the light as an energy medium are "destructive" in the sense that the light lawfully reflected from (and presumably patterned by) environmental properties is not specific to those properties. Very often this doctrine has reduced to a comparison of two very different sets of descriptors: On the one hand there are the descriptors of the environment as it relates to the animal (for example, surfaces of support for locomotion, places to hide, courting behavior of a conspecific) and on the other hand the descriptors of the light as identified by the physicist, namely, photons or individual rays varying in intensity and wavelength. When stated in this inequitable-descriptors form, the doctrine of an intractably nonspecific relationship between environment and light (cf. Turvey and Shaw, 1979) lays the ground rules for identifying the problem of perception. If the descriptors of the proximal stimulus (the light at the eye) are as they are defined, then they are clearly incommensurate with the descriptors of perceptual experience, for the latter are in reference to such things as surfaces of support for locomotion, places to hide, etc. It follows, therefore, that the problem of perception is the problem of identifying the sequence of internal cause and effect relationships or the sequence of internal transformations that *translate* the impoverished and fine-grained descriptors of the proximal stimulus into the rich and coarse-grained descriptors of perceptual experience.

If the mapping from environment to light is *destructive*, then the mapping from light to perceptual experience must be *reconstructive*; and to achieve this reconstruction perceptual theorists have been compelled to postulate a large variety of epistemic mediators (cf. Turvey, 1977b), among which may be numbered hypothesis testing, matching to memory and innate principles of organization. The gist of the conventional orientation to visual perception (and to perception in general) can be stated briefly: given two mismatched vocabularies or descriptor sets that putatively define the starting point and end point of perception, that is, the proximal stimulation and the perceptual experience, how is the former translated into the latter? The gist of an alternative orientation, and one that we endorse, can be given an equally brief statement: Can a *single* vocabulary or descriptor set be found in which can be defined both the proximal stimulation and the perceptual experience, so that no translation is needed? At the heart of this alternative orientation is the assumption that the mapping from environment to energy medium (say, the light) is not destructive, as time-honored doctrine would have it; rather, the light to an eye is lawfully structured in ways that are specific to the properties of the environment. On this alternative view, when the light is structured by environmental properties, there are descriptors of the light to be found that correspond unequivocally to those properties, and that some of these descriptors can be intercepted by any given animal. But since there is no longer an assumed mismatch between the descriptors of the environment and those of the proximal stimulation, there is no longer reason for assuming a mismatch between descriptors of the proximal stimulation and those of perceptual experience (see Mace, 1977; Shaw *et al.*, in press; Turvey and Shaw, 1979).

This alternative orientation to perception that we have espoused was first proposed by Gibson (1950, 1966). As the logician Hintikka (1975) has observed in reference to Gibson's theory: "the conceptual moral is that the *perceptions* that can surface in our consciousness *must be dealt with in terms of* *the same concepts as what we perceive.* The appropriate way of speaking of our spontaneous perceptions is to use the same vocabulary and the same syntax as we apply to the objects of perception" (see p. 60). It is this "conceptual moral" which we believe ought to be similarly pursued in the theory of action. The orientation to coordinated movement, as intimated above, has been consistent with the familiar orientation to perception; the vocabulary of intentions, like the vocabulary of perceptions, has been conventionally construed as distinct and largely separate from the vocabulary of activity. If there is a mismatch between the descriptors of intention and the descriptors of muscle-joint activity, then "translation" identifies the problem of action. However, in an alternative orientation to action, which echoes the alternative orientation to perception, the problem of action is to find a single vocabulary or descriptor set in terms of which intentions and muscle-joint activity conflate.

Returning to the instance of action theory that is of present concern, namely, speech production, we might enquire as to the *a priori* plausibility of the translation-type theory. In our view, the concept of abstract forms that cannot be realized nondestructively in any medium is implausible for several reasons. One is the absence of any obvious way in which they could be acquired. It is one thing for an adult perceiver to bring these concepts to bear *a priori* on an acoustic signal (cf. Hammarberg, 1976). But if the signal can only be interpreted by an individual who *has* these concepts, how does a child learn them? Evidently, we have to suppose that they are innate. But this is to invoke "innate knowledge" not *as* an explanation, but instead of one. It is inappropriate to pawn the acquisition problem off on the biologist if it is as unclear how the concepts could be acquired by a species in evolution as it is how they could be acquired by an individual (Turvey and Shaw, 1979).

Second, it is not evident *why* these concepts should have arisen in evolution. If evolution, like ontogeny, is considered a process whereby an organism maintains or enhances its compatibility with selected aspects of the world (cf. Shaw *et al.*, 1974), it must be disadvantageous for an organism to seek to impose fictitious categories on a spoken utterance as a translation theory demands. ("Fictitious" here refers to categories that are not "in the world", but only "in the mind"; see Hammarberg, 1976, for an expression of this view.)

The above identify reasons why translation theories of speech production are logically insecure. Let us now proceed to a brief overview of their adequacy as accounts of speech production.

B. The Adequacy of Translation Theories

A common thread (perhaps *the* common thread) through all psycholinguistic accounts is the appreciation for Lashley's (1951) identification of the issues in, and potential solutions to, the control problems posed by language. His particular rendition was so convincing that many individual theories of production which followed it appear to stand in its review, especially on the issues of priming and integration. In the original, these issues respectively involved: (1) the selection and ordering of abstract entities (themselves possessing no temporal valence) and, (2) the expression of these ordered entities in a hierarchically controlled manner by the response mechanism. These two aspects of production have been variously termed (higher level) activation, facilitation, patterning, preplanning and sequencing for the first, as opposed to (lower level) expression, execution, articulation and motor control for the second.

The argument for separating the act of speaking into largely distinct stages recognizes a simple fact about language; ignoring the ineffable shadings of meaning, a single idea can have infinitely many equivalent realizations from the standpoint of both the actual linguistic entities employed and the alternative progressions of peripheral events which immediately create the speech sounds. Because of this lack of straightforward isomorphism among thought, word and muscle, some intermediary agency must obviously determine the ideally appropriate constellation of linguistic elements and response parameters to convert idea to utterance.

The underlying unanimity in the various elaborations of the problem of serial order should not dissuade us from noting several essential failures in the program. These are not inconsequential to the choice of hypothesis of control, as would be a deletion or refinement of a rule of syntax. Rather, they urge the reconsideration of the idea of two distinct levels of description. The goal of higher level functions (in theories which have this specialty) is assumed to be a sequence of linguistic elements, usually phonemic, in which the derived order is prospective of the intended utterance as well as faithful to the semantic content. While this output is a description, more or less, of the abstract structure of the message, it is not really a plan for action; it would be such only if language were a commerce involving the exchange of though-phrases expressed in IPA notation. Articulation, the vehicle by which the message is actually delivered, requires that a lower level neuromuscular process fulfill the plan of the higher level "mental" process through vocal tract efforts. One way to view this relationship of higher level to lower is that the tailoring of the plan to fit the peculiarities of the motor system, is essentially a detail of physiology; while interesting to consider, the facts of the motor system largely agree with the production concept of a higher level plan transferred to a lower level system for routine completion.

The inadequacy of this approach to the serial ordering problem is brought out by a brief consideration of the aforementioned job of tailoring (or translating). Attempts to discover segments in the speech stream—for example in muscle events, articulatory movements, vocal tract area functions, or in the acoustic signal—that correspond to the segments listed in the higher level representation, have met with little success. Given this predicament, it appears that, as the problem is conventionally conceived, the adaptation of linguistic plans for the requirements of articulation involves a great deal of creativity on the part of the tailoring function.

An example of the creativity of a tailoring function occurs in vowel production. Producing vowels seems to involve approaching a relatively invariant shape of the vocal tract, but from various directions depending on the vowel's articulatory context (MacNeilage, 1970). A second example is provided by the compensation findings of Folkins and Abbs (1975) and of Lindblom and Sundberg (1973) (see also Lindblom *et al.*, 1979). Both studies show that talkers can compensate immediately, or almost immediately, for imposed disruptions of normal articulation (unexpected resistive loading in the experiment of Folkins and Abbs, and a bite block in that of Lindblom and Sundberg). These capacities are not handled in the typical higher level plan, but neither can they be dismissed as details of the physiology.

The foregoing considerations suggest that any putative tailoring functions cannot be simple. More than this, however, Fowler (1977) argues that the hypothetical distinction which necessitates a tailoring function—namely, that between the higher order plan and its executor—is a false distinction. Among other things, it has fostered a view that plans represent timeless (static) segments of speech, even though the view is falsified by quantities of data. (See for example, Lisker (1974, 1976) for reviews and discussions of this point.) In addition, many models of speech production that have adopted the distinction are unable to account either in the plan or in the tailoring function for many of the characteristics of running speech. In particular, they do not adequately handle coarticulatory phenomena (Kent and Minifie, 1977), adjustments for speaking rate and stress, and speech rhythms (Fowler, 1977).

One conclusion inspired by the dynamic aspect of speech is that solutions to the serial ordering problem that have proposed higher level plans written in a static vocabulary may be inadequate by design. In order to deflect, rather than to confront, the important constancy problems in production, as well as the remainder of the serial ordering problem, the tailoring assumption was invoked. We believe that this will not do if what is desired is a description of language *production*. Instead of viewing the peripheral machinery as a necessary evil, as linguistically oriented dualistic models are prone to do, a more appropriate description might consider production to be the acts, not merely the thoughts, of the speaker. A *plan* in this sense would be a physiologically considerate description of the act, and not only an abstract summary of its significance. This kind of plan would be sensitive to what can actually be said, rather than to that which can only be cogitated.

Now, one of the fundamental points we wish to argue from a general perspective on action is that the style of control of any activity must reflect the contingencies of execution while complying with the aim of the activity. In a similar vein, the objective of natural phonology (Donegan and Stampe, 1977) in some cases is to remove the distinction between competence and performance, the mediate and the immediate. In that view, the traditional operating principles of modern phonology, formal simplicity and abstract distinctiveness, represent the needs of the professional linguist more than they characterize the structure of language. In an analysis kindred to ours, natural phonologists claim that the solutions for many of the conundra in linguistics are to be sought in the dynamics of talking, rather than in the bloodless domain of formal grammars. But what possibilities for understanding organization during speech become available through scrutiny of the putatively lower, more basic properties of speech acts? The remainder of the paper may be viewed as a response to this question.

III. Principles of a General Theory of Coordinated Movement

We have argued that a critical assumption tends to characterize current accounts of speech production: it is of an irreducible difference between the properties of linguistic segments as they are *known* to a language-user and those of speech sounds as they are uttered and realized acoustically. The assumption bears major responsibility for the character of theories of speech production and perception. For their part, speech production theories are compelled to characterize a translation process from mental plan to vocal performance. The translation process is supposed to destroy crucial information about segment identity. Perception theories are obliged, thereby, to explain how abstract segments are reconstructed on the basis of acoustic stimulation that is impoverished in respect to specifying abstract segment identity.

We have given reasons for trying to avoid translation type theories; it becomes appropriate, therefore, to consider the possibility of a theory of direct (non-translational) speech production. A reconciliation was proposed above between the characters of abstract and actual segments to the effect that segments, both abstract and actual, are essentially dynamic; their defining properties are context-free, and in relation to other segments in an utterance they are qualitatively separate, one from the other, and they are serially ordered. The considerations that lend plausibility to this characterization have two sources. The primary source is a theoretical perspective on coordinated movement owing to Bernstein (1967), Greene (1972) and Turvey (1977a). This perspective offers a way to understand how the essential properties of abstract classes of acts may be preserved unaltered in various instantiations of them. The second source is the literature on speech production itself. First we will discuss some of the key concepts encountered in studies of coordinated movement, then we will examine such concepts in light of the relevant speech production data.

A. Some Relevant Properties of the Control of Coordinated Movement

It is commonplace for speech production accounts to characterize a talker's control over his articulators in the following way. Planned phonological segments are given a distinctive featural representation. Some of the members of the feature bundles characterizing each segment are spread or shared among neighboring segments in the plan. The effect is to increase the similarity of adjacent feature bundles and, thus, to avoid abrupt changes in articulatory specification at segment edges (Daniloff and Hammarberg, 1973). Feature bundles are executed in sequence as their component features are translated into commands to muscles. The accounts of Daniloff and Hammarberg (1973), MacKay (1969, 1970) and Liberman and Studdert-Kennedy (1978) are proposals of this sort.

A number of considerations render the last (and perhaps least controversial-seeming) stage of this proposed sequence at best implausible. These considerations, detailed below, are the context-conditioned relationship between innervation of muscle and its outcome in movement, the "degrees of freedom problem", and the inherent organization of the musculature into systems. By reformulating this stage of articulatory control, we believe that we can obviate its being considered separate from its predecessors. The result in principle could be a non-translational, or direct, theory of speech production.

1. *Context-conditioned variability*

Bernstein (1967) observed, and after him many others (Hubbard, 1960; Grillner, 1975; Turvey, 1977a; Turvey *et al.*, 1979), that the relation between a central "command" to a muscle and its consequences in movement is equivocal. One source of this equivocality is the context of extant forces, muscular and otherwise, with which a commanded muscular contraction interacts. A simple example makes this clear. Intuitively, if an actor wishes to raise his forearm by rotating it about the elbow joint, he has only to send down a command to contract the flexion agonist. But, in fact, contraction of the brachioradialis or the biceps only leads to flexion at the elbow under restricted conditions—namely, conditions in which none of the extant forces on the arm when combined with the new forces alter the intended character of the movement. If the forearm had been extending at the elbow when the command arrived, the result might only have been a deceleration of extension. If the forearm were fixed, the command's consequence might be a rotation of the body at the elbow joint. In short, a given command will have different consequences for movement depending on the nature of the force field into which it was inserted.

Thus, an alternative is evidently required to the proposal that plans for movement include commands coded in the vocabulary of muscles which completely specify an intended movement. Somehow the control over movement must be realized by *bending* the current field of forces in an intended direction (cf. Bernstein, 1967; Hubbard, 1960; Fowler and Turvey, 1978). In addition, the nature of this bending must be contingent on the character of the field to be bent. An implication for speech production must be that there is no simple relationship between the distinctive features of a phonological segment and contractions of particular muscles.

An example, in speech production, of the context-conditioned variability in the relationship between a muscle and its role concerns the muscles of articulation (see, also, below). According to Zemlin (1968), the palatoglossus muscle may either lower the soft palate or raise the back of the tongue, depending on which is the freer to move. The mylohyoid may raise the hyoid or depress the jaw, and the geniohyoid may raise the hyoid and protrude the tongue or depress the mandible. Which of the alternatively possible roles are made manifest will depend on whatever other forces are operative on the different articulatory structures. Presumably, this kind of context-conditioned variability underlies the observed absence of invariant electromyographic (EMG) correlates of phonetic segments.

There is another source of context-conditioned variability that is physiological, rather than mechanical or anatomical, in nature; it is due to descending neural fibers synapsing on interneurons primarily, rather than directly on motoneurons. This indirect innervational pathway allows other influences than hypothetical muscle commands on the activities of motoneurons.

Muscles of the vocal tract are innervated by the cranial nerves whose cell bodies are in the brain stem. In primates some limited proportion of fibers from the motor, premotor and somesthetic cortical areas synapse directly on these cranial nerves. However, the system of direct fibers is only supplemental to the more substantial system of cortical fibers that innervate interneurons of the brain stem (Kuypers, 1958; Carpenter, 1972). Interneuronal networks typically receive input from many descending and afferent neural fibers. Thus, the state of an interneuronal network depends on inputs from several sources. Indeed, Evarts *et al.* (1971) speak of interneurons as nodal points of integration of information. Apparently they are not passive recipients of supraspinal inputs. Rather, supraspinal influence constitutes one among many influences on the state of interneuronal networks.

A system of the kind just described seems to preclude movement control via cortical *commands* to muscles. Apparently any commands would be changed when they reached the system of interneurons. None of this precludes a role for messages of central origin in movement control. But it does exclude a role of supplying commands to muscles. What is needed, we would argue, is a set of muscular forces that relates to the set of nonmuscular forces in such a way as to yield the needed movement. In this view the role of supraspinal influences looks to be organizational rather than executive.

2. *The degrees of freedom problem*

A description of any arbitrary time-slice of an act of speaking would be very lengthy and complicated if it were to catalogue all of the muscles being regulated in the service of the speech act, and the manner of their interregulation. Not only are the muscles of the vocal tract controlled and intercoordinated, but they in turn are coupled, at the very least, to the muscles of the respiratory system. And the muscles of respiration themselves are subject to a special speech-mode of regulation (e.g. Ladefoged, 1968). Described in this detail, and recognizing the very rapid rate at which articulatory events are perpetrated, talking seems to constitute an extraordinary feat of motor control.

In fact, virtually all naturally occurring activities are feats of this sort. Bernstein characterizes the "degrees of freedom problem" as the theorist's problem of explaining how the many degrees of freedom of the body (however they are counted: in terms of muscles, joints, trajectories of limb segments, etc.) are regulated in the course of an act. Precluded is a proposal that each degree of freedom is individually controlled at every point in time during an act. There are too many degrees of freedom (Greene, 1972, estimates dozens in the head-neck system alone), and there are too many points in time in an act, for a like manner of control to be workable. Clearly, there must be some way in which many degrees of freedom can be automatically regulated through the individual control of very few. The next section will describe how this seems to be done in the case of some activities. It will be instructive first, however, to look at the implicit solution to the problem of degrees of freedom in models of speech production.

Although writers in the speech literature often remark on the extraordinary complexity of articulatory events, extant models of speech production rarely confront more than a small part of the degrees of freedom problem. For example, the prototypical speech model given earlier is both remarkable and intriguing in regard to the aspects of the production act that its articulatory plan leaves out. All that the plan *does* represent, in fact, are the values along just those dimensions of description (of a talker) that distinguish a particular speech segment from any other. Unspecified are dimensions of description of a talker's activities not directly relevant to the speech act; dimensions (if any) whose values are invariant to all speech segments, and most importantly, dimensions that *are* relevant to speaking and whose values are controlled, but at a slower rate than segmental controls—for example, speech breathing and the suprasegmental aspects of an utterance.

We recognize that not all proposals or models conform to this prototype (for example, the hierarchical model of Kent, 1976). But many do, and they are particularly interesting in respect to their implicit solution to the problem of degrees of freedom. The contents of their articulatory plans conform to a general strategy in the theory of language for the systematic description of languages. That strategy is to simplify the description of a language, and to express its orderliness by separating its regular, invariant properties from properties that are idiosyncratic to particular components. Thus, abstracted from a listing of the rules of a particular language are universal rules of grammar. Similarly, removed from the lexical entry of a word in a particular language are any of its properties that are predictable by the general rules of the language (e.g. Chomsky, 1965). Likewise, in the case of the articulatory plan, according to production theory, the strategy is to exclude from the moment-to-moment description of the planned utterance everything except its aspects that are idiosyncratic to each moment.

By itself, of course, a plan of this sort is inadequate because it fails to explain how the slower, systematic motor events surrounding segmental articulation are regulated, or how they are integrated with segmental events. But they do express a strategy that may be as useful to a talker as its grammatical counterpart is to a linguist and presumably to a child learning a language.

One can think of the relationship between a grammatical rule, say, and a particular instantiation of it in a sentence as like that between the equation for a circle ($(x-h)^2+(y-k)^2=r^2$) and its realization as a particular circle ($(x-5)^2+(y-2)^2=45$). The regularity that the general rule expresses provides a *frame* which sets out what is essentially invariant across all of its instantiations. Equally important, it establishes slots or parameters (h, k, r) which constrain the particular ways in which instances are allowed to be idiosyncratic. The parameters h, k and r are not *substantial* properties of a circle, but are *attributes* or adjectival properties of them. *And it makes no sense to speak of the parametric values without expressing what they are parameters of.* We suggest that speech production may be similarly characterized. There are regularities to which all utterances and all parts of all utterances conform. These include the special mode of breathing, the phonation mode of laryngeal adjustment (Wyke, 1967, 1974), suprasegmental timing constraints, the near alternation of consonants and vowels and, in English, the near alternation of stressed and unstressed syllables. These regularities provide an organizational frame (of the musculature, as we will suggest below) for an utterance that both establishes the *possibility* of producing segments, and constrains the ways in which particular segments can be idiosyncratic. Viewed in this way, the idiosyncratic aspects of an utterance—that is, the distinctive features of its successive segments—are parametric values *of* an extant organization; they are not substantives in themselves.

In summary, what seems to us attractive about the articulatory plans of translation theories is their reasonable (if tacit) solution to the degrees of freedom problem. That solution recognizes that some aspects of an act of talking need not be regulated from moment-to-moment because their time-scale of execution is longer than a segment. And, by virtue of that, their moment-to-moment manifestations are determinate. All that is specified from moment-to-moment in a plan are those aspects of talking that are idiosyncratic to each moment. What is unattractive about the proposals, however, is that they mistakenly impute substantive properties to attributes. Evidently the idiosyncratic features of particular segments in a plan are not things in themselves, but rather are attributes of things. Hence, they make no sense described apart from a characterization of their frame (cf. Garner, 1974, for a general discussion of features as attributes of a system). What is required in a specification of a plan for an utterance is a way to preserve this kind of solution to the degrees of freedom problem, while capturing the necessary support provided features by coarser-grained regularities of talking.

3. *The inherent organization of the musculature into systems*

All of the acts that animals perform intentionally are coordinated. This implies that they are products of organized relationships among muscles, since in the absence of those relationships, different muscles would tend to compete and oppose each other's effects (cf. Weiss, 1941). A "commands to muscles" view of motor control proposes, in effect, that an actor's plan fails to recognize or to exploit those necessary organizations (cf. Easton, 1972).

However, there is substantial evidence that muscle systems are marshalled in the service of various aims. This is one important way in which the "degrees of freedom problem" is minimized. The muscle systems, called "coordinative structures" (Easton, 1972; Turvey, 1977a; Turvey *et al.*, 1979), are functional groupings among muscles. A system like a coordinative structure that performs a function incorporates an optimal balance between its freedom to undergo change and limitations on its freedom (cf. Pattee, 1973). A rigid object, a table for instance, does not function because all of its degrees of freedom are tightly constrained. At the other extreme, an aggregate, say of grains of sand, is hardly constrained at all, but it does not function either. Systems that perform a function are somewhere in between. They are specialized by virtue of the limitations they place selectively on their degrees of freedom, but they are not so tightly constrained as to effect rigidity. Rather, their freedom is to perform a coherent activity of some particular type.

Coordinative structures are "built" via physiologically mediated biasings of the musculature called tunings (see Turvey, 1977a for a review of this evidence). The biasings facilitate or inhibit the excitability of certain muscles and thereby both alter the relationships among muscles and determine the *kind* of activity they will promote.

Some properties of the coordinative structure are modulable. Because of that, these muscle systems are said (e.g. Greene, 1972) to promote an equivalence class of acts, much as the general equation for a circle delimits an equivalence class of instances. The modulable properties of the coordinative structure are the properties along which the class members differ and are like the *parameters* of the equation for a circle. This property of a coordinative structure gives it its versatility and adaptiveness, because it enables a single organization of the muscles to govern activities with different superficial properties. (For example, as we will see, a single organization of the muscles may govern locomotion at different rates and with different gaits over rough or smooth terrains.)

An act is believed to be governed by functionally embedding (as opposed to temporally concatenating) coordinative structures. Each nesting level delimits a broader equivalence class of movements than the finer-grained level nested within. That is, the definition of "equivalence" broadens in that the essential properties defining the class weaken and reduce in number. The more coarse-grained nesting levels are established by altering the relationships among smaller coordinative structures, and at the same time they act as constraints on the lower ones.

Once a nesting of coordinative structures has been marshalled, the control task is minimal, because each coordinative structure performs its function autonomously (see below). Thus, at most, the actor addresses these muscle-systems as units (rather than addressing their constituent muscles individually) when he assigns values to parameters of the mappings that describe the muscles' interorganization.

Locomotion provides a supreme example of an act that conforms to this mode of description. The neuromuscular organization of a locomoting quadruped is one in which small systems of muscles are nested within larger systems whose role is to coordinate the activities of two limbs. These systems, in turn, are parts of a superordinate system whose domain of governance is the whole step (Easton, 1972; Grillner, 1975).

Shik and Orlovskii (1965), among others, have monitored the stepping cycle of a single limb of an animal who is suspended over a treadmill so that only the single limb touches the treadmill belt. This procedure allows the investigators to examine the muscular organization of a single limb when it is stepping. The relative onset times of flexion and extension of the hip, knee and ankle joints are patterned invariantly over cycles. The same patterning is observed in a spinal animal (Grillner, 1975). These observations suggest quite strongly that stepping by a limb is the product of systemic relationships among muscles that are established at the spinal cord. The organization is one that generates a stereotypic pattern of muscular activation and movement. It can do so relatively independently of supraspinal influences (see Shik and Orlovskii, 1976) and absolutely independently of commands to muscles. The role of supraspinal influences, in part, seems to be to establish the locomotor "mode" of organization of the spinal cord and musculature (Shik *et al.*, 1966) and not to trigger components of the step.

The participation of the intra-limb coordinative structures in a larger system of muscles is observed by allowing two limbs, or all four limbs, to touch the treadmill. Under conditions of undisturbed stepping, the relative onset times of flexion and extension in the two or four limbs are also patterned in a stereotypic way in time. Thus, not only does each limb evidence a relatively fixed pattern of activity, but also the stepping cycles of pairs of limbs are intercoordinated.

The intercoordination among the limbs is made more apparent when one limb is disturbed while the animal is walking on the treadmill (Shik and Orlovskii, 1965). If a limb is prevented temporarily from initiating its transfer stage (the stage during which the limb is off the surface of support), so that its stepping cycle is out of its usual phase relation to the other limbs, it is gradually brought back into phase over several cycles. This is effected by the muscles in a way that alters the stepping cycles of the *unperturbed* limbs, as well as the perturbed limb itself. These findings confirm that *intra*-limb coordinative structures are organized into larger systems that regulate *inter*-limb stepping.

The nestings of muscles govern not just one sequence of movements, but rather an equivalence class of acts. Locomotion at different rates and with different gaits can be described as the products of a single nested system of muscles with one or two modulable parameters (Arshavskii *et al.*, 1965; Easton, 1972). Likewise, Shik and Orlovskii's (1965) mathematical model of interlimb coordination incorporates the animal's ability to accommodate its stepping cycle to the inhomogeneities of environments. These investigations of locomotion provide a particularly well-understood example of an equivalence class of acts that is governed by relatively autonomous and adaptive nested systems of muscles.

In summary, one way in which the coordinative-structural mode of organization minimizes the degrees-of-freedom problem for an actor is by allowing him to address several muscles as a unit. Thereby he controls the many degrees of freedom that a nesting of coordinative structures regulates by deliberately providing values for just a few.

There is a second way as well. In an act of walking, coordinative structures provide only some of the forces necessary to control locomotion. The remainder are provided by gravity, friction and the reactive forces created by the actor's contacts with the surface of support. Moreover, in walking, some movements require no muscular control at all—most of the transfer phase, for example (Grillner, 1975)—because the actor's momentum brings them about. Muscular control is only required, and is only used by a skilled actor, to change inertial states (Hubbard, 1960).

IV. An Action Theoretic View of Speech Production

A. Taking Advantage of Nonmuscular Forces

As pointed out above, the kinds of considerations that are relevant to a general description of the control of coordinated movement are paralleled in the more local domain of speech production. A brief examination of some recent literature in this area should serve to illustrate these concerns. An excellent review of the role played by the context of nonmuscular forces in the overall speech production process is presented by Abbs and Eilenberg (1976). Let us identify some of their key observations as they relate to the kind of story of motor organization that we have been discussing.

Abbs and Eilenberg present a hypothetical example of the sort of phenomenon that we have labelled as "context-conditioned variability". Consider the problem of opening the jaw, a movement that is due in large part to the effective muscle force generated by the anterior belly of the digastric muscle (ABD). A schematized model of this system (see Fig. 1) assumes simple mandibular rotation about the axis at the condyle. With fixed skull and hyoid bone position, a specification of the effective muscle force as a function of jaw position is possible. Mandibular rotation can be accounted for in terms of the component of the total ABD muscle force that is tangent to the arc of rotation (Ft), with magnitude of the Ft vector being proportional to the angle between the ABD force vector and the tangent vector. Put simply, this model illustrates a case where a descending constant motor influence would yield a variable output, that is, variable jaw opening movement, depending upon the position of the jaw. Abbs and Eilenberg point out that observations of EMG activity undertaken while studying a movement such as this must be considered in

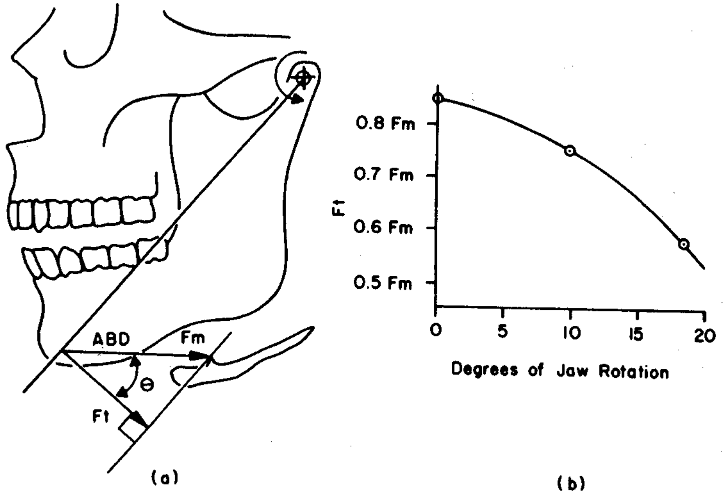


FIG. 1. [From Abbs and Eilenberg, 1976, Figure 5.6, p. 150.]

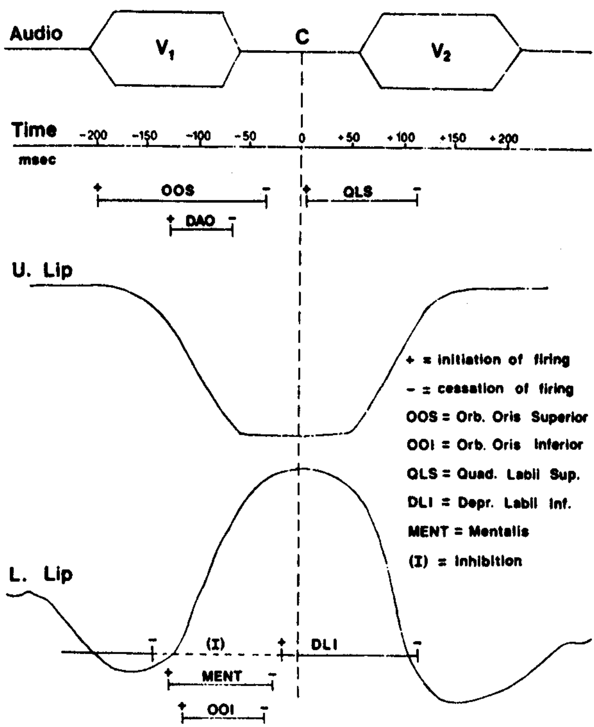


FIG. 2. [From Sussman, MacNeilage and Hanson, 1973, Figure 2, p. 403.]

conjunction with the effective jaw lowering force of the particular muscle in question—a force, then, that is varying.

Let us consider a further example, one that expresses the exploitative use of nonmuscular forces, in this case the inherent elasticity of the muscles. EMG activity in the orbicularis oris muscle (see Fig. 2) is seen to cease just prior to release in the production of bilabial closure with a vowel-consonant-vowel frame (Sussman *et al.*, 1973: Abbs, 1973a; Abbs and Netsell, 1973). It may be claimed that the release of a bilabial stop in a case such as this takes advantage of the inherent change in the elasticity of the lips, which covaries with changes in the level of orbicularis oris contraction. Bilabial stop release seems to result from an apparent cessation of this muscular activity—a cessation that results in a decrease in the stiffness of the lips that is necessary for closure, with the concomitant result of lip opening.

Abbs and Eilenberg extend this kind of description even further by attempting to separate the influences of passive components and active muscle properties in just such an articulatory system. An indication of the organization among such subsystems can be found in a series of experiments (Abbs and Netsell, 1973; Abbs, 1973a) which examines the contributions of muscle forces and passive elasticity in lower lip movements. Subjects were required to alternately move and relax their lower lip. Recordings of EMG signals during these activities from the orbicularis oris and the depressor labii inferior muscles indicate that the passive elastic component of lower lip movement can be separated from those components that represent active muscle forces. An examination of Fig. 3 reveals that, with passive elasticity considered as a "springmass" system, lower lip displacement is to be

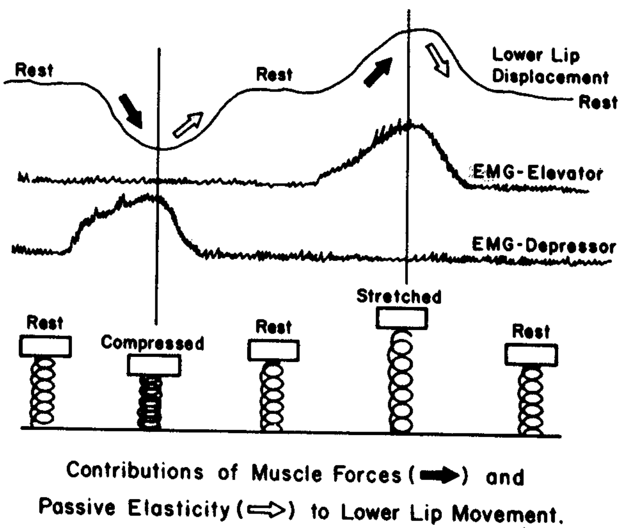


FIG. 3. [From Abbs and Eilenberg, 1976, Figure 5.7, p. 152.]

accounted for in terms of the conjoint influences of muscular and nonmuscular forces.

These illustrations speak to the same concerns as those which must be confronted in *any* consideration of the organization of movement. Descriptions solely in terms of command signals to muscles must give way to a more structured understanding of the complementary relation between forces supplied muscularly and the forces supplied reactively and otherwise (cf. Fowler and Turvey, 1978).

The principle of control that is being discussed can, as a first approximation, be conceptualized as *exploitative* rather than compensatory. The different nonmuscular forces that are manifest in speech—namely, those arising from the mass and elasticity properties of articulatory structures, those arising from air pressures and flows, that due to gravity and, in general, those that are biomechanical in nature, both passive and active—should not, in our view, be considered negatively as sources of variability to be overcome by deftly timed and weighted commands to muscles. On the contrary, they should be construed positively as usable aspects of a highly flexible organization. A quote from Ashby (1956) pinpoints this theme: "when a constraint exists advantage can usually be taken of it" (p. 130). And the issue in narrower perspective is given due expression by Abbs and Eilenberg (1976):

> Just as the back swing in ball throwing is an attempt to add momentum, it may be that some of the bizarre patterns of movement observed during speech production result from system utilization of inertial or elastic properties. Further, it is possible that some of the movements and muscle activity patterns that have been implicated as evidence of *higher level control* in abstract models (such as *coarticulation*) may represent system activity utilization of mechanical properties. In any case, it is clear that one cannot observe movements and/or EMG patterns and make inferences to underlying neuromuscular mechanisms without an appreciation of these passive principles. (p. 152)

B. The Equation-of-Constraint Perspective on The Concept of Coordinative Structure

There is an especially useful perspective that we might take on the concept of coordinative structure in order to advance our understanding of the concept and to highlight what it entails, particularly for speech. We consider this perspective in this section and we do so prefatory to a literature review which provides tentative evidence for coordinative structures in speech production.

Here, we review Saltzman's (1977) discussion of degrees of freedom and equations of constraint. The degrees of freedom for a given system are the least number of independent coordinates required to specify the position of the system elements without violating any geometrical constraints (Groesberg, 1968; Timoshenko and Young, 1948). Consider two elements in a two-dimensional space as depicted in Fig. 4(a).

A total of four independent coordinates are needed to specify the respective positions of A and B, namely, x_1y_1 (for A) and x_2y_2 (for B). The system

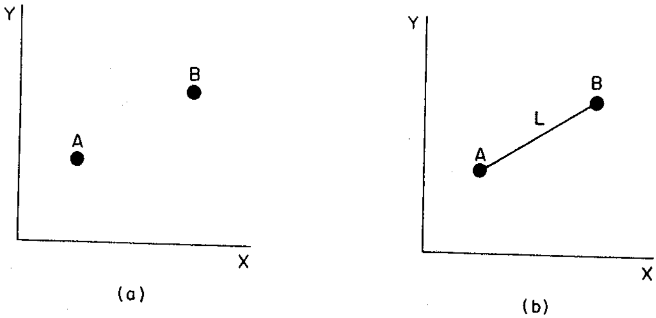


FIG. 4 (a). A two-element system with no equations of constraint.
(b) A two-element system with one equation of constraint.

depicted in Fig. 4 has four degrees of freedom, since a change in one coordinate will not limit the possible values of any of the others. Consider now the two elements in Fig. 4(b). These elements are connected by a line of fixed length L and their respective coordinates are not independent, since they must satisfy the following *equation of constraint*:

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 = L^2$$

In this case, given the specification of three of the coordinates, the fourth is not free to vary. The system depicted in Fig. 4(b) has three degrees of freedom. As a general rule, the number of degrees of freedom of a system is given by $(Dn - c)$, where D is the dimensionality of the space, n is the number of elements in the system and c is the number of equations of constraint. For Fig. 4(a) the space is two-dimensional, the number of elements is two, and the number of equations of constraint is zero; hence there are four degrees of freedom in this system. For the system depicted by Fig. 4(b) the dimensionality and number of elements are both two, but there is one equation of constraint, hence this system, by the general rule, has three degrees of freedom.

With these preliminary remarks at hand, let us proceed to develop what might be called an *equation-of-constraint* perspective on the concept of coordinative structure. We do so through an example used elsewhere (Turvey *et al.*, in press), but one which will receive, in these pages, a slightly different and, we believe, more useful interpretation.

Imagine a rather simple airplane with five hinged parts: the ailerons on the rear edge of the wings which can be moved up or down regulating roll; two elevators on the horizontal portion of the tail section which can likewise be moved up or down regulating pitch; and a rudder on the vertical tail fin that can be moved right or left regulating yaw. Let each of the five hinged parts be free to adopt one of nine positions; the ailerons and elevators can go into four different positions above and four different positions below the zero position (zero being defined as flush with the wing for the ailerons and flush with the horizontal tail portion for the elevators); and the rudder can go into four positions to the right and four positions to the left of its zero position (defined as flush with the vertical tail fin). If the five hinged parts are independent of each other then, in the pilot's perspective, the airplane has five degrees of freedom, since for each of the five hinged parts ($n=5$) only one coordinate is needed to define its position ($D=1$) and there are no equations of constraint ($c=0$).

Let us reduce the degrees of freedom by introducing the following three equations of constraint, given the convention that movements up from and to the right of zero are positive positions and movements down from and to the left of zero are negative positions. One equation of constraint is given by yoking the left aileron (LA) to the rudder (R) so that as the left aileron moves up, the rudder moves to the left by the same number of positions; that is, $LA=-R$. A second equation of constraint is given by the yoking of the right aileron (RA) to the rudder so that when the right aileron moves up, the rudder moves to the right by the same number of positions; that is, $RA=R$. And the third equation of constraint is given by yoking the right elevator (RE) and the left elevator (LE) so that they move in unison up and down by the same number of positions; that is, $RE=LE$.

The five hinged parts have now been reduced to two subsystems, namely, the ailerons-rudder subsystem (since the ailerons are linked via the rudder according to $RA=-LA$), and the elevators subsystem, each subsystem having one degree of freedom. In short, a system of five degrees of freedom with a possible 54 049 states (9^5) has been reduced to a system of two degrees of freedom with a more manageable set of 81 possible states (9^2). But a further reduction is possible, albeit through a constraint of a coarser grain. Preserving the airplane theme, the two subsystems themselves can be linked through the joystick according to an equation of constraint such as ($P=kQ$), where P is the ailerons-rudder subsystem responsible for banking and turning. Q is the elevators subsystem responsible for rising and falling and k is a constant. We now have a system of one degree of freedom; the dimensionality is one (as before), the elements are two in number and there is one equation of constraint.

The airplane example nicely captures what we take to be a prominent feature of coordinated movement: *When a group of muscles, or at coarser grain, a group of muscle collectives, functions as a unit (that is, as a coordinative structure), indices of the individual muscles, or individual collectives, appear to covary in terms of a relatively fixed relationship that is indifferent to overall magnitude changes in the indices.* (Elsewhere (Turvey *et al.*, in press), following Boylls (1975), the invariant relationship has been referred to as a structural prescription and the magnitude of the indices as a metrical prescription.) More generally, the airplane expresses the view, noted above, that coordinated movement is engendered by functionally embedded (as opposed to temporally successive) equations of constraint defined over elements of different grains. On this view, and in keeping with the gist of the preceding sections, coordination is to be understood not in terms of the "issuing of commands", but rather in terms of the "arising or specification of constraints" (Remez and Rubin, 1975; Fitch and Turvey, 1978).

What kind of indices upon muscles or muscle collectives might "covary" owing to an equation of constraint? The index for a muscle might be its level of innervation or the time, relative to the group's behavior, at which a muscle is innervated. Where the focus is the organization over a number of links in a biokinematic chain—such as an individual limb in running, or the cervix, thorax and pelvic girdle in breathing (Gurfinkel *et al.*, 1971—the index might be coarser, such as joint angle (e.g. Arutyunyan *et al.*, 1968, 1969) or the torque generated at a joint. And where the focus is the relation among the limbs, say during locomotion, the index might be even more coarse-grained, such as cycle time. Thus, in the alternate step gaits, limbs of the same girdle are 0·5 of a cycle out of phase with each other and obstinately preserve this relation in the face of manipulations that alter disproportionately the cycle times of the individual limbs (Kulagin and Shik, 1970).

A set of elements together with a set of equations of constraint identifies a control system in this sense: the degrees of freedom of the set of elements so constrained is less than the degrees of freedom of the set in free-variation. Given the *equation of constraint* perspective on coordinative structures that we have just developed, we now ask what kind of system is it that equations of constraint (written, as it were, on the neuromuscular apparatus) give rise to? Our tentative answer is that the kind of system so created is analogous to a self-regulating and biasable vibratory system (Fitch and Turvey, 1978; Turvey, 1977a). The motivation for this answer derives from the work of Asatryan and Fel'dman (1965) and Fel'dman (1966a, 1966b) which receives a brief description in the paragraphs that follow.

Under some conditions a neuro-muscular system can be established by an actor (to govern forearm movement relative to the elbow joint) that is similar in its properties to a spring system. A linear spring system can be described by the equation: $F = -s(l - l_0)$, where l_0 is the length of the spring when there are no forces operating on it; l is the current length of the spring; s is a stiffness parameter; and F is the force developed by the spring. The behavior of an actual spring is governed completely by the external forces ($-F$) exerted on it. In contrast to a real spring, the elbow system, which is shown to behave like a spring, incorporates some internal controls. It is not governed solely by the value of $-F$. Significantly, it can alter l_0, the zero-state of the system. (l_0 is the joint angle at the elbow that the system will assume in the absence of any forces on it. It is an abstract variable whose value is established by the collaborative activities of several muscles.)

Asatryan and Fel'dman (1965) demonstrated the spring-like properties of the joint-muscle system in the following way. A subject attains and holds some joint angle. His forearm is resting on a mobile horizontal platform whose axis of rotation allows elbow flexion and extension. Suspended from the platform are a set of weights (at the subject's wrist) that promote extension. To hold steady the specified joint angle, the subject has to counteract exactly the moments of the forces of the weights exerted on his arm. In terms of the equations above, he must establish some l_0 such that the difference between l_0 and l (the joint angle specified by the experimenter) multiplied by s just counteracts the $-F$ due to the weights. If the spring analogy is appropriate, then:

(1) If the investigator releases some of the weights, thereby changing $-F$, and if the subject allows his angle to change passively, it should move to a new l such that the difference between the new l and the established l_0 counteracts the new $-F$. A graph plotting F by l should be linear if the elbow behaves like a linear spring. But, in any event, if it is analogous to a spring system, the curve should be simple and lawful.

(2) A family of *parallel* curves should characterize a family of experiments across which the subject is asked to hold different steady-state joint angles. (To counteract the same weights, a different l_0 must be set by the subject when the experiment prescribes that he holds different ls. l_0 is the x-intercept in an F by l plot, and s is the slope. Hence, $F \times l$ plots of experiments evoking different l_0s should yield parallel curves with one curve intersecting the x-axis to the right of the other.)

These predictions were borne out in the experiments. The curves for all experiments were nonlinear, but lawful. Across experiments (that is, across different prescribed ls) the curves were parallel indicating that the subject altered l_0 to perform the task. The experimenters concluded from these results that for the purpose of the experiment the subjects established a joint-muscle system with spring-like properties. When asked to hold a particular joint angle at the elbow, the subjects did so by resetting the "zero-state" of the joint-muscle system (i.e. l_0).

There are several features of the coordinative structure as a vibratory system metaphor that we find attractive and which we believe to be of special relevance to issues in speech production. First, a vibratory system is intrinsically cyclic or rhythmic *but it need not behave cyclically or rhythmically*. Suppose we modeled a mass-spring system by the following equation:

$$mX''(t)+kX'(t)+sX(t)=0$$

where $X(t)$ is the displacement of the system from the equilibrium at time t, $X'(t)$ and $X''(t)$ are the first and second derivatives respectively, and where m, k, s stand, in that order, for the mass, the damping (or frictional) parameter and the stiffness parameter. Where these parameters are related as $0 < k^2 < 4$ ms) the system oscillates with the amplitude of oscillations decreasing with time. But if the parameters are related as ($k^2 \geq 4$ ms) the system does not oscillate; to the contrary it rapidly achieves the equilibrium position in less than a cycle. The point is that it is often tempting to conceive of rhythmic and nonrhythmic behavior as distinct cases requiring possibly separate mechanisms. In a mass-spring system we have a concrete example of a single mechanism that can do both; whether its behavior is rhythmic or not depends simply on the current parameterization of the system. (As an aside, in pursuing the vibratory system metaphor we take note that the intuitive object that we are trying to understand might be more judiciously labeled coordinative *cycle* rather than coordinative *structure*.)

A second significant feature of a vibratory system follows from the first, namely, that the steady-state or equilibrium position is indifferent to the initial conditions and is determined only by the system parameters. Within limits, a mass-spring system can be stretched or compressed to any degree yet it will, on free-vibration, equilibrate at the same length. And we may note that the kinematic details of its equilibrating behavior will differ from one set of initial conditions to the next. We are reminded that kinematics refers to the descriptions of motions in contrast to dynamics which refers to the explanation of motions. Included in the vocabulary of kinematics are such terms as distance, direction, velocity, etc. while the vocabulary of dynamics includes terms such as force, viscosity, stiffness, momentum, etc. Which brings us to the third notable feature of a vibratory system: the kinematic details of a vibratory system's behavior are determined by its dynamic properties—the system parameters—in conjunction with the initial conditions (e.g. degree of initial stretch or compression); and it would be incorrect to claim that the kinematic details are determined by an *internal representation* of these details, for nowhere within a vibratory system is there a representation, symbolic or otherwise, of the kinematic details of the behavior consequent to disequilibration (Fitch and Turvey, 1978).

There is one further feature of a mass-spring system that we would do well to note: a mass-spring system *is not a servo-mechanism*. The goal of a servo-mechanism's activity is set by an external command that provides the reference signal from which, subsequent to feedback, error signals are derived and on the basis of which corrections are determined. While we could easily speak about a mass-spring *as if* it were a servo-mechanism complete with reference signal, feedback and error signals, we would be guilty of unduly stretching these servo-mechanism concepts; indeed, we would be guilty of introducing fictitious quantities. The "goal" of a mass-spring system in *intrinsic* to the system; it is a necessary *consequence* of the configuration of dynamic properties, rather than something imposed from outside as a causal *antecedent* of the system's behavior. In a mass-spring system there is neither feedback to be monitored nor errors to be computed and corrected and, patently, no special devices for performing these operations.

C. On Explaining "Immediate Readjustments"

We turn now to a curious phenomenon of speech production in order to compare the kind of explanation afforded by the vibratory system metaphor with that afforded by a more conventional view. Lindblom and Sundberg (1971) reported that speakers can produce natural vowels with a bite block between their teeth, without the need for extensive relearning. More recent and complete evidence is to the effect that speakers fitted with bite blocks and instructed to produce isolated vowels do so within the range of variability for normal vowel production. And, most significantly, they do so without the need for acoustic feedback; acoustic measures reveal satisfactory vowel production in the first pitch pulse of speech (Lindblom *et al.*, 1979). A similar outcome is obtained when lip protrusion is impeded during the production of rounded and unrounded vowels. Compensation for impedance of lip protrusion is achieved for each vowel by differentially lowering the larynx to a degree that preserves the vocal tract length characteristic of each vowel. Even on the first trial, under these conditions, vowels are acoustically normal (Riordan, 1977).

How might the phenomenon of "immediate readjustment" be accounted for? Substantial reasons have already been given elsewhere (e.g. MacNeilage, 1970) for dismissing open-loop explanations of phenomena of the kind under analysis, and we will not pursue these reasons here. The issue of interest to us is how well a *closed*-loop explanation fares in this regard. All such explanations lie, of course, within the purview of the servo-mechanism metaphor. Thus, in closed-loop explanations a sensory referent is proposed, tied either to the environmental goal of the movement such as, say, a spatial target or an acoustic pattern, or to the commands to produce the movement (compare Adams, 1971; MacNeilage, 1970; Pew, 1974; Powers, 1973; Schmidt, 1975). The comparison of sensory feedback, a perceptual signal (p), with the referent signal (r) defines an error signal (E) which provides a basis for correcting the lower-level mechanism(s) responsible for controlling the referent (see Fig. 5).

In the phenomenon under analysis, feedback related to the acoustic

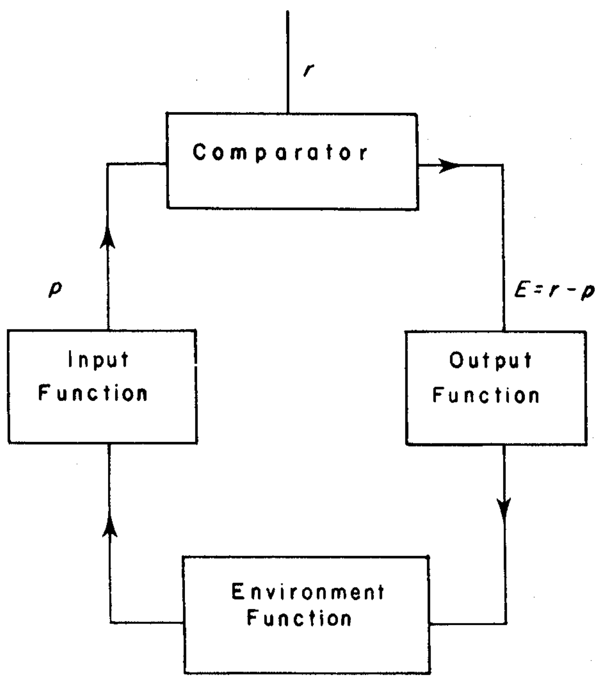
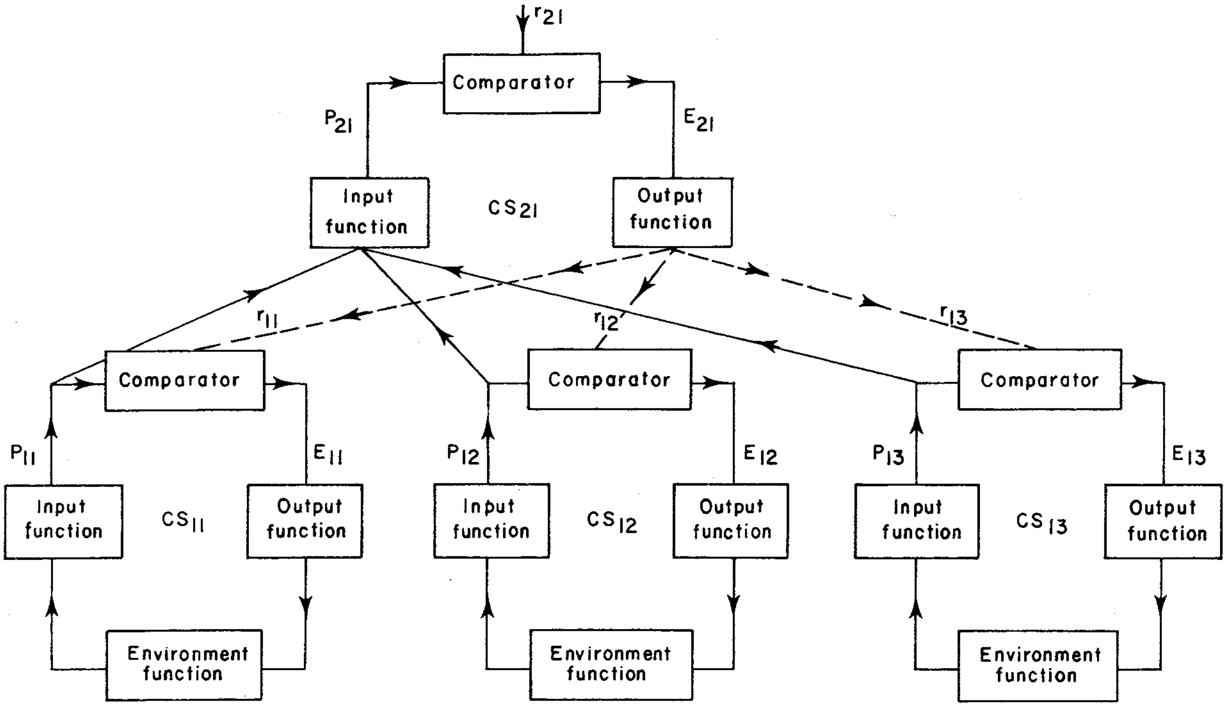


FIG. 5. An individual control system.

consequence of the movement may be absent—we are reminded that in the Lindblom *et al.* (in press) demonstration vocalization is normal in the first pitch pulse. This fact would seem to rule out a closed-loop explanation in which the referent is tied to the acoustics. But it does not rule out a closed-loop explanation in which the referent is felt vocal tract shape. Nor does it rule out the following special version of a closed-loop explanation. The manifest plasticity and context-appropriateness of movement has invited some students of coordinated movement to invoke a more cognitive style of control than that embodied by a bare-bones servo-mechanism. The claim is that more central mechanisms use the techniques of model-referenced (or schema-referenced) control (e.g. Arbib, 1972; Ito, 1970) in order to *predict* the commands needed for goal-directed movement given variation in initial conditions (see also Schmidt, 1975). Let the model be of the actual servo-mechanism (Eccles, 1969) in which the motor commands and sensory consequences are simulated. We can then see how the requisite commands for the actual servo-mechanism problem are arrived at through a process of error-correcting, precisely, that set of simulated commands is selected which produces simulated sensory consequences matching the reference signal. All of this, of course, is assumed to precede the actual commands to the musculature; and so it must if the model-referenced control notion is to account for the "immediate readjustment" phenomenon. But let us examine closely the error-correcting process, whether it be in an actual or simulated closed-loop, and whatever the referent signal, for we have serious doubts about its tractability.

Suppose that three (relatively) independent muscles, when acting together, bring about a change in some notable vocal tract parameter. And suppose, not unreasonably, that of all possible combinations of the states of these muscles only a subset will yield a given value of the aforementioned parameter. The problem is this: When the actual value (the perceptual signal, p) and the desired value (the reference signal, r) do not match, how are the states of the individual muscles adjusted (that is, how is a more apt combination of their states found) so as to reduce this mismatch? How is the means of adjusting informed about the adjustments to be made? The magnitude of error indicates how near the collective action of the three muscles is to the desired values, but it fails to indicate how their individual states are to be adjusted to reduce the perceptual signal—reference signal mismatch.

The problem can be illustrated briefly (see Fowler and Turvey, 1978, for a more complete discussion). Consider Fig. 6 which depicts a two-tiered control system composed of three lower level or first-order control systems (CS_{11}, CS_{12}, CS_{13}) and one higher level or second-order control system (CS_{21}). Each first-order system supplies CS_{21} with a perceptual signal. In keeping with the general servo-mechanism strategy of unidimensional feedback, the perceptual signal of the second order system, p_{21}, is a linear transformation of the three, first-order perceptual signals, p_{11}, p_{12}, p_{13}. Thus, $p_{21} = a_1 p_{11} + a_2 p_{12} + a_3 p_{13}$ (cf. Powers, 1973). That signal is subtracted from the reference signal, r_{21}, of the second-order system. The result, $E = r_{21} - a_1 p_{11} -$



CS_{ij} --control system j of the i-th level of the hierarchy

p --perceptual signal

r --reference signal

E-- error signal ; $E = r - p$

$p_{21} = a_1 p_{11} + a_2 p_{12} + a_3 p_{13}$

$E_{21} = r_{21} - a_1 p_{11} - a_2 p_{12} - a_3 p_{13}$

FIG. 6. A hierarchy of control systems: Three first-order systems nested under one second-order system.

$a_2p_{12}-a_3p_{13}$, is the error signal of the second-order system. This error signal is the basis for selecting the reference signals, r_{11}, r_{12} and r_{13}, of the first-order systems.

Unfortunately, for any given error signal there are very many possible combinations of values for r_{21}, p_{11}, p_{12} and p_{13} that might yield that error, even if some boundaries are set on the possible ranges of values that each might take on. The error signal might be entirely due to an error of one of the first-order systems; or it could be one of many combinations of errors on the part of all three first-order systems. In summary, unidimensional feedback must rarely be informative in a hierarchical closed-loop system because, typically, there is a one-to-many mapping between an error signal and the conditions that may have provoked it.

The servo-mechanism conception is attractive when the entity commanded to produce a reference-matching perceptual signal has but one degree of freedom (such as the commonly described servo-mechanism—the thermostatically controlled home-heating system). The attractiveness of this mechanism, in its unqualified form, decreases, however, as we increase the number of degrees of freedom in the commanded entity. In the absence of any well-defined, and possibly special-purpose (Gel'fand and Tsetlin, 1962, 1971), search procedure, the error-correcting process on n degrees of freedom would be essentially random and temporally indefinite.

Let us now consider an interpretation of the "immediate readjustment" phenomenon in the equation of constraint perspective. Given a set of variables related by an equation of constraint, if one of the variables is fixed, the others are not free to vary but must assume those values that preserve the equation. In the airplane example above, if the equation of constraint linking the ailerons-rudder subsystem and the elevators subsystem identifies an invariant ratio, for example, $P=kQ$, then if P's value is frozen Q *must* assume whatever value is necessary to keep invariant the ratio k.

The ability of speakers to generate acceptable vowel qualities in the face of a fixed position of the mandible or of the lips, can be accounted for in very much the same way. Like the ailerons, elevators and rudder of the airplane, the components of the vocal tract, including the mandible, tongue, lips and larynx, can be described as linked by equations of constraint. More simply, the components are represented as variables in an equation that defines the relationships among them. The constrained relationships among components of the vocal tract "create" a vibratory system—that is, a system with an intrinsic goal which it attains from any starting point by virtue of its dynamic configuration. When a bite block is introduced in the system, it fixes the values of the variable for jaw position. As we have indicated, under conditions in which the value of a variable is fixed, the remaining variables assume values that preserve the equation of constraint. So long as the requisite values are attainable by the components of the vocal tract, the effect of a bite block should be negligible.

The assumption here is that an invariant vibratory-like organization of the muscles underlies vowel production. In very large part the aim of the final part of this paper is to substantiate this claim. At this point, we proceed to consider evidence that coordinative structures are involved in speech.

D. Fine-grained and Coarse-grained Coordinative Structures in Speech

An emphasis on volitional aspects of motor behavior has resulted in a viewpoint that would be conducive to coordinative structures as we speak of them, but only in terms of handling nonvolitional, or automatic, functions. We prefer, instead, to emphasize that this form of analysis of motor behavior extends across all domains, both volitional and nonvolitional. Some brief examples should serve to show how such concerns manifest themselves in aspects of speech production not generally included in various theories, for example respiration, and those more commonly discussed, as in the cooperation between articulators and structures in the supralaryngeal system. (See Fowler, 1977, for a more detailed examination of the role of coordinative structures in speech production.)

An examination of the respiratory system reveals that basic reflexes operate both in vegetative breathing and in speech, with the apparent purpose of regulation of the initiation and termination of inspiratory activity (see, for reviews, Kaplan, 1971; Lambertsen, 1974). Briefly, the medullary respiratory center receives several kinds of vagal and hypoglosseal influence, supplementary to its own intrinsic activity; among these are afferents originating in (1) alveolar stretch receptors (which signal pulmonary distention); (2) aortic chemoreceptors (which monitor blood CO_2 level); and (3) aortic and carotid baroreceptors (which monitor blood pressure). These receptors signal the various changes due to the respiratory cycle, and can be used in a control system for inspiratory and expiratory oscillation independent of higher cortical influence.

At a more macroscopic level vegetative respiration and speech respiration are seen to be quite different. During vegetative breathing, the activity of the inspiratory muscles and the inspiratory portion of the respiratory cycle are in phase (Lenneberg, 1967). Furthermore, except during forced expiration, that phase is typically accomplished passively—that is, by relaxing the muscles of inspiration and by allowing the elastic recoil forces of the lungs to work unaided and unopposed. The expiratory phase occupies about 60% of the cycle. During speech, however, the sequence of events is somewhat different (Draper *et al.*, 1959; Lenneberg, 1967). In speech, the activity of the inspiratory muscles is out of phase with the act of inspiration. Their activity extends into the early part of expiration where they act to check the descent of the ribcage that characterizes passive expiration. Immediately following the decline and offset of activity in the inspiratory muscles, the internal intercostals, that are muscles of active expiration, come into play. If phonation is prolonged, two other muscles that may contribute to expiration, the rectus abdominis and the latissimus dorsi, are also marshalled (Draper *et al.*, 1959).

As a result of this coordinated activity, the proportion of the respiratory cycle occupied by the expiratory phase is about 0.87 (Lenneberg, 1967) and the subglottal pressure is maintained at a nearly constant level, despite the continual decrease in the volume of the air in the lungs (Draper *et al.*, 1959; Lieberman, 1967). Lenneberg (1967) refers to these coordinated activities of the muscles of inspiration and expiration as "synergisms", a term that others have used in a sense that relates to, but is not identical with, the sense in which we use coordinative structures. This particular macroscopic coordinative structure can be conceptualized as a device whose task is to control subglottal pressure, much as the microscopic reflexes regulate stretch and CO_2 levels in the blood. The macroscopic coordinative structure is superimposed on the smaller ones when an individual chooses to speak. Notably, in the experiment of Draper *et al.* (1959), similar sequences of muscular events occurred over a range of controlled subglottal pressures. That is, the same coordinative structure may govern an utterance over all amplitudes of production, the different subglottal pressures or amplitudes of production being *metrical* variations of the governed act, rather than an invariant property.

An examination of the literature provides further evidence for macroscopic coordinative structures in speech at the level of supralaryngeal control. The interpretation of this body of data by speech theorists has been that clusters of articulatory commands to muscles tend to be released simultaneously (Daniloff and Hammarberg, 1973; Kent *et al.*, 1974; Kent and Moll, 1975). But we have seen that, in view of the physiological, mechanical and anatomical indeterminacies of the relationship between a central signal and its peripheral effect, control centers cannot send *commands* to individual muscles. In our view, a more likely interpretation of these data may be that coordinated gestures in speech are evidence of coordinative structures. The coordinative structures are marshalled via biasing operations, but they generate the coordinated gestures autonomously.

Synchrony between articulators is evident in a study by Kent and Netsell (1971), which shows that the gestures of the tongue body and the lips are coordinated during the production of the word *we* in *We saw you.* A figure relating the displacement of the tongue body to that of the lips over different stress patterns of the phrase (*we* saw you; we *saw* you; we saw *you*) shows that the relationship between the two variables is invariant over differences in stress. Kent and Netsell obtain a similar result for the dipthong /oI/ in the noun-verb opposition of *CONvoy* and *conVOY*, and tentatively conclude that, "for sounds like /wi/ and /oI/, which are characterized by coordinated movements of two articulators, the stress contrast must alter both gestures or neither gesture" (see p. 40).

Additional evidence for the cooperation between articulators can be seen in experiments by Kent *et al.* (1974) and Ushijima and Hirose (1974) on velar control and movement. For example, Kent *et al.* (1974), using as a basis for analysis cinefluorographic films of tongue and velum movements, conclude that their data indicates that in many cases "articulatory movements seem to be programmed as coordinative structures so that movements of the tongue, lip, velum and jaw often occur in highly synchronous patterns" (see p. 487). Such patterning can be seen in the articulation of a word like *contract*, where the initiation of velum lowering coincides with the release of oral closure for /k/, and velum elevation starts as the tongue-tip movement begins for the consonant /n/. As the authors point out, the goal of velum opening in such a case has been achieved and reversed *prior* to the time of attainment of the concomitant goal of oral constriction.

Of further interest are cases in which articulatory gestures are not synchronized at a particular moment, but rather occur in a constrained pattern over time. Notice that evidence of this can only be obtained by comparing similar utterances or by comparing the utterance produced at different rates. That is, a nonsynchronized pattern can be detected only if it remains invariant in different contexts. Recent data of Bell-Berti and Harris (1979) appear to show that the onset of lip-rounding precedes the measured acoustic onset of the rounded vowel /u/ by a relatively fixed interval, regardless of the preceding consonantal context. Some of the data of Kent and Moll (1975) may be given a similar reading.

There is some limited evidence that the relative timing of gestures of the tongue body, velum and lips tend to be ivariant over a change in rate of speaking (Kent *et al.*, 1974). For each articulator and for each of two speakers, figures are provided which superimpose the movement tracings at two different rates during the production of "soon the snow began to melt". The rates of speaking were in the ratio 2:1. To facilitate a comparison of the movements' relative timing at the different rates, the investigators compressed the time-scale of the slower movement relative to that of the fast movement. For both speakers, and for all three articulators, the tracings at the two rates of production were nearly identical. Thus, the relative timings of the movements of a particular articulator and the timing relationships among articulators remained nearly invariant over a two-fold increase in rate. These findings are commensurate with the coordinative structure conception, but there is reason for caution. For instance, Gay and Ushijima (1974) show that the muscle activity for consonants is of greater amplitude, and that of vowels of lesser amplitude, during rapid as opposed to slow speech. We should remark, however, that only EMG data were provided and making inferences from muscle contraction to movement is difficult at best and illegitimate at worst (Hubbard, 1960). In addition to the data of Gay and Ushijima, we may note the evidence provided elsewhere (for example, see Klatt, 1976) for articulatory undershoot for vowels, but not for consonants, at a fast rate of speech. The figures of Kent *et al.* (1974), however, do not show any tendency for the movements of the tongue body to be any less extreme in fast speech than they are during slow speech.

E. Techniques For Examining Speech Motor Organization

A final mention should be made of some techniques that might prove useful as tools for isolating and examining the proposed coordinative-structure underpinnings of speech. These techniques all involve procedures that are, in one form or another, disruptive to normal activity. Such approaches have been used with success in more general examinations of action systems (Asatryan and Fel'dman, 1965; Fel'dman, 1966a, 1966b; Shik and Orlovskii, 1965), and their application in the study of speech production could prove valuable. An example of this methodology can be found in a study by Sears and Newsom Davis (1968) in which sudden alterations in mechanical load during breathing were applied. The occurrence of such disruptions during voluntary activation of the respiratory muscles resulted in concomitant changes in EMG activity of the intercostal muscles. Disruptive loading and unloading were accomplished by introducing transient changes in the pressure in an obstructive airway, which yielded small, but rapid, changes in lung volume, and by admitting timed pressure pulses of air into the breathing circuit. Loading, which effectively increased lung volume, resulted in immediate increases in EMG activity, while unloading produced a brief plateau, or period of inactivity, in this signal. The brief latencies of the EMG responses indicated their reflex nature. Further, the results of this study indicated that this reflex structure was sensitive to both changes and rates of changes in lung volume. The authors, then, concluded that their experiments provided, "positive evidence in conscious man of a reflex system that automatically tends to stabilize the voluntary demands for respiratory movement" (p. 190).

In a similar vein, Folkins and Abbs (1975) examined the effect of disruptive jaw loading on lip movements. In their experiment subjects were asked to repeatedly produce the phrase, "a /haepaep/ again". On about 25% of the test utterances, jaw loading was applied randomly by introducing a clamping force to a prosthodentic splint, which effectively provided a transient resistance to upward jaw movement. Simultaneous recordings were made of the EMG activity of the medial ptyerygoid, temporalis, masseter and orbicularis oris superior muscles, and of inferior-superior movements of the upper lip, lower lip and jaw, through the use of a strain gauge. The loading force was applied during the bilabial closure movement that corresponded to the initial /p/ in the test utterance, preventing the jaw from reaching its usual degree of elevation during this activity. Even though this resistance was applied, subjects attained lip closure in all cases. An examination of the data from the physiological monitoring indicated that the compensatory gestures that permitted lip closure were exaggerated upper and lower lip displacements, and slightly exaggerated velocities of lip closing gestures. In a follow-up study (Folkins and Abbs, 1976), the authors provided additional evidence that the compensatory gestures were not merely of a passive nature, as would be the case, for example, if the momentum of the lips had moved them further than normal. The results of this replication study indicated increases in the EMG activity from the depressor anguli oris and the buccinator muscles. Folkins and Abb's interpretation was that the activity of these muscles, which contribute to upper lip depression, provides evidence for *active* compensatory adjustments in the labial muscles. The results of both of these studies are explained most simply in terms of a low-level jaw-lip control system that is responsible for lip closure. Folkins and Abbs suggest that its mode of operation is such that it *adapts* its strategy to current conditions, rather than programming a particular gesture in advance.

Techniques that interfere with normal speech production activities are not restricted to the application of such impedances. Examples of other procedures include the adaptation by subjects to bite blocks—we took note of these above—and dental prostheses. Fairly extensive reorganization is needed when subjects are fitted with dental prostheses that alter the configurations of their vocal tracts. Amerman and Daniloff (1971), relying on listener judgments, found normalization of vowel production within 5 min of the insertion of a prosthesis in a subject's mouth. In a similar procedure Hamlet and Stone (1976) fitted subjects with three different types of prostheses. The effects on the production of vowels were striking. Compensation for vowel changes was variable among the subject pool and was not always successfully accomplished, even after a week of adaptation. In addition, a period of readjustment was required by subjects subsequent to the removal of these protheses. In the case of the bite block studies, we might be viewing the ability of the articulators to attain, simply, a functional reorganization within an equivalence class of permissible, potential organizations. It is possible that when experimental prostheses are used, the subjects have been moved beyond this "ballpark" of permissible organizations and fundamentally restructure their articulatory approach to the new task at hand—an endeavor that they, apparently, find very difficult to undertake.

A final class of disruptive techniques involves the use of differential nerve blocks and selective anesthetization to examine the role played by oral sensory and other forms of feedback and the contribution of lower-level systems to the productive act (cf. Ringel and Steer, 1963; Scott and Ringel, 1971; Borden *et al.*, 1973a, b; Abbs, 1973b; Putnam and Ringel, 1976; Abbs *et al.*, 1976). At present, an integrated understanding of the implications of these studies seems to be hindered, in part, by methodological and interpretive disagreements. However, the application of such procedures in coordination with other forms of physiological monitoring may prove to be valuable as a method for probing aspects of organization in the action system.

One technique for examining *tuning* in action systems involves a form of probing of spinal nervous activity (see Turvey, 1977a, for a review). Examples of this approach can be found in the work of Gottlieb and his associates (Gottlieb *et al.*, 1970; Gottlieb and Agarwarl, 1972, 1973), in which reflex activity was elicited prior to a voluntary movement. In the work of Gottlieb *et al.*, for example, subjects were required to continuously track target levels of force by step effort of either plantar flexion or dorsiflexion. Hoffman reflexes (H-reflex) were elicited by electrical stimulation of the popliteal fossa at various points during the voluntary force-generating activity. Simultaneous measurements were made of the monosynaptic H-reflex and the EMG activity in the gastrocnemius-soleus muscle group (agonist in plantar flexion) and the anterior tibial muscle (agonist in dorsiflexion). If the H-reflex was elicited in the 60 ms period that preceded any EMG indication of voluntary motor unit activation, a marked increase in its amplitude was observed when the voluntary activity was plantar flexion. During dorsiflexion, in which the gastrocnemius-soleus group functions as the antagonist, the amplitude of the H-reflex reflected inhibition. A biasing is thus apparent in which the control structure takes advantage of modulation in the agonist-antagonist opposition.

This technique provides an elegant way of tapping into different neuroanatomic levels, with particular emphasis towards examining the state of the segmental apparatus and its relationship to movement; as in the case considered above, experimental results indicate a pre-movement segmental tuning. Recently, a similar technique has been applied in a speech production context. Netsell and Abbs (1975) have developed an approach similar to that of Gottlieb and his associates that demonstrates a way in which an aspect of the control structure of the facial musculature can be continuously modulated, or tuned. This procedure involves the elicitation of the perioral reflex during static muscle contraction and the production of the syllable /pa/, with concomitant recordings of EMG activity from the orbicularis oris superior muscle. The results show a facilitation of the afferent-efferent pathway during muscle contraction in both the static case and during bilabial closure. The amplitude of the perioral reflex during orbicularis oris shutdown was smaller than that recorded during the resting state, indicating a complementary inhibition. This systemic behavior, a "continuous modulation" in the terms of Netsell and Abbs, represents a further demonstration of the physiological biasing we have referred to as tuning.

In summary, in this portion of the paper we have briefly reviewed some of the relevant literature in the field of speech production with the intention of directing our attention toward the need to view such data from the broader perspective afforded by concepts that have had application in the general study of coordinated movement. This sort of attempt has, of late, begun to be seen as an attractive and necessary alternative to traditional approaches employed in examining speech physiology (Fowler, 1977; Harris, 1976; Moll *et al.*, 1976).

V. Toward a Theory of The Direct Perception and Production of The Sounds of Speech

A. Linguistic Segments as Invariants

We have reviewed the experimental literature suggesting that control over articulation is fundamentally similar to control over other kinds of acts. At this point, it is worthwhile to reconsider the insights that these properties of motor control may provide in reference to two issues raised above; namely, those having to do with the essential properties of linguistic segments as they are perceived and uttered.

1. *Can segments that are dynamic and coarticulated nonetheless have invariant (context-free) properties?*

A coordinative structure is a functional relationship among muscles that can be described by a mapping. In some respects, the mapping is not unlike the equation for a circle given earlier. An example is ($F = -s(l - l_0)$), which describes the spring-like behavior of the muscles governing forearm movement at the elbow under some conditions (Asatryan and Fel'dman, 1965; Fel'dman, 1966a, b). The mapping expresses what is context-free about acts. Importantly, the context-free properties are not realized as movement invariance, but as something closer to functional equivalence. The mapping also expresses the dimensions along which different instances are free to take on different values. In the spring system of the elbow, those dimensions are s and l_0. Thus, the values of these parameters are the distinctive features of the different movements allowed by the mapping, or they may be the dimensions along which the movement may adjust to the context.

Despite the lack of movement invariance (the elbow system governs both extension and flexion over various extents), the acts governed by the elbow system have invariant properites. In essence, these invariant properties are actualizations of the constraints on possible movements as described by the mapping. We suspect that the essential properties of linguistic segments will turn out to be relational properties of this sort. Shortly, we will provide a speculative example to illustrate this. Here our concern is only to show that the coordinative structure type of organization, as noted above, engenders various kinematic outcomes with properties that may be superficially dissimilar (context-adjusted), but have underlying dynamic invariance.

Given the abstractness or depth of the invariance on the action side, it may seem unlikely that any corresponding acoustic invariance can be found. Indeed, it is generally agreed that the perception of speech sounds is not based on acoustic invariants (but see Stevens, 1975). However, the kinds of invariants that have been sought are of a superficial nature (e.g. discrete and context-free representations of a linguistic segment with invariant properties that are visible on a spectrogram). These *evident* acoustic invariants may well be absent, but this in no way implies the absence of invariants of a less obvious sort. Perhaps, instead, we should expect to find acoustic invariants analogous to (or, better, complementary to) the abstract and relational invariants putatively underlying speech production.

We borrow from investigations of visual perception the insight that the invariants supporting perception are abstract. Investigators, in the visual domain, as in speech, have failed to discover simple correspondences between dimensions of the proximal stimulus for visual perception and those of the percept. For example, the perception of a rigid object's size and shape is invariant across any perspective that an observer chooses to adopt on the object, despite enormous variability in the received descriptors of the retinal image. Thus a shape, invariantly perceived, is not represented by an invariant shape on the retina. Nonetheless, under a different analysis that does not equate the description of the light at an eye with the traditional description of the retinal image (cf. Gibson, 1961), the light to an eye can be shown to contain invariant information about shape and other properties of an environment, and about the perceiver's relation to it (Gibson, 1973; Lee, 1974; Johansson, 1973; Shaw and Pittenger, 1977). Invariants prove to be difficult to find and to describe because they do not *resemble* the properties to which they correspond, but they are *specific* to them. The invariants that have been isolated are abstract and relational rather like those that we propose underlie action.

Indeed, in some cases, the invariants underlying activity and its perception seem very close even in the visual domain. Johansson (e.g. 1973) shows that perceivers readily recognize locomoting human forms based on films displaying only point lights of the figures' joints. A vector analysis of the movements of these points of light expresses their regularities. Presumably the regularities correspond to constraints on point-light movement. (That is, no given light can move any arbitrary distance from the others, along any arbitrary path, because in fact it is attached to a body.) Over time, the constraints evidently are detected and specify a locomoting human form to a perceiver. Yet these same constraints follow not only from fixed properties of a walker's form, but also from the particular organization of his body that engenders walking. In short, Johansson's vector analysis most probably expresses the invariants in the motion of the point lights that specify a walking human figure to a perceiver; but simultaneously they express the invariant organization of the actor's body that enables locomotion.

We believe that an analogous story will unfold in the speech domain. The regularities underlying articulation will be as abstract and relational as those underlying walking. These regularities nonetheless impose a patterning on an acoustic signal which specifies them to a hearer.

2. *Can the separateness and serial ordering of segments be preserved in an utterance despite coarticulation?*

As noted earlier, we replace the characterization "discrete" with "separate and serially ordered". Outside the speech domain, we are not always puzzled that two temporally overlapping events are perceived as separate. Consider, for example, a recording of a singer and his musical accompaniment. Perceivers do not find if difficult to hear the singer's voice as separate from the instrumental accompaniment, although the impressions of each source on the compression wave at the ear must be thoroughly intermixed. Evidently, what specifies their separateness is that they constitute different kinds of sound-generating devices and they generate different kinds of acoustic products.

Vowels and consonants are different kinds of articulatory and acoustic events. Vowels are relatively slow, global changes in vocal tract shape effected primarily by moving the tongue body about in the oral cavity. On the other hand, consonants are rapid local obstructions of the vocal tract. Each kind of gesture produces a different kind of acoustic product: vowel gestures produce a relatively slowly changing pattern of formant frequencies, and consonant gestures yield acoustic evidence for obstruction.

In short, coarticulated segments in an utterance are nonetheless separate (but not discrete along the time axis) by virtue of being different kinds of gestures, each producing a different kind of acoustic product. Conversely, the complex of articulations that together constitute a given segment (e.g. lip-rounding, and movements of the tongue body and jaw to produce /u/) are perceived to cohere because they are similar *kinds* of articulations contributing to a coherent kind of acoustic event.

Johansson's work in the visual domain has an analogue to this as well. The spot of light at a walker's wrist may pass very close to that at his hip at some point in each walking cycle, and may even obscure it periodically. However, a perceiver is not persuaded, thereby, to treat the two spots of light as cohering. The *movement* of the spot of light at the wrist closely follows those of the spots at the elbow and shoulder. The three spots move as a collective and specify a coherent entity, the swinging arm. The wrist-spot movement follows that of the spot at the hip much less closely. Hence, the two do not seem attached or part of a single body-segment.

Not only the separateness, but also the serial ordering of segments is preserved in a coarticulated utterance. Two sequential segments in an utterance are never entirely concurrent. For example, movement towards a consonant may begin during a preceding vowel, and it may end during a following vowel, but it does not happen that a consonant's initiation and termination both co-occur with the initiation and termination of a given vowel segment.

Based on the foregoing considerations, we believe it plausible to suggest that linguistic segments as perceived are compatible with uttered segments. A dynamic segment, as well as a static one, can have context-free properties: furthermore, despite coarticulation, a dynamic segment can be separate from, and serially ordered with respect to its neighbors, both articulatorily and *acoustically*.

B. A Speculative Example of a Muscle System Underlying the Production of Vowels

We conclude by describing a rough and speculative model of some aspects of vowel production. It is based on many of the principles of coordinated movement outlined above. However, it does not purport to represent a theory of vowel production; indeed it is a *metaphor* for vowel production rather than a model. Its worth lies in demonstrating that a system embodying a coordinative-structural style of control exhibits some essential properties of vowels, many of which elude expression in theories equating vowels with feature bundles or with spatial targets. We first specify the kinds of properties of vowels that we take to be essential; we then proceed to describe the model.

An adequate theory of speech production accounts for the following kinds of vowel properties:

(1) Vowels constitute a natural articulatory class as distinct from consonants and from all other classes of acts. As a natural class, its members share the defining properties of the class and a theory of speech production optimally specifies these properties and accounts for them.

(2) Similarly, a given vowel must have essential (context-free) properties that are preserved across its instantiations.

(3) The essential properties of either vowels as a class or of a particular vowel are not invariant movements. This is evident, for example, in the production of the vowel /ɛ/. If /ɛ/ is produced from a high tongue position, it involves a lowering of the tongue body. If it is produced from a low tongue position it may involve raising the tongue.

(4) Although their context-free properties cannot be invariant movements, vowels are essentially dynamic (cf. Shankweiler *et al.*, 1977). Thus, while it may be fair to say that movement in vowel production is always toward some particular shape of the vocal tract, it is undesirable to treat the target as if it were the vowel's canonical form. For one thing, targets are often not even attained in running speech. More importantly, perhaps, a spatial target is static, while speaking is only dynamic. If a static shape were a vowel's essential property, movement toward and away from the target would be only a byproduct of the fact that targets are being produced in a flesh and blood system. (That is, by virtue of a vocal tract being a physical system, in order to get from one shape to another, it is necessary to traverse all intervening shapes.) However, if movement is merely a byproduct of planned static targets, it is curious that only these byproducts of production manifest themselves consistently in speech and that the static targets do not.

As we suggested earlier, it is implausible to suppose that linguistic segments as they are known to a language-user are resistant to nondestructive realization in a vocal tract. Therefore, it is unlikely that targets are a vowel's canonical form, and is more likely that a vowel's essential properties are dynamic in nature.

We propose that the class of all vowels is an equivalence class of *gestures* which are equivalent in a broad sense. The members of the class of vowels are all gestures that aim toward some global shape and length of the vocal tract, although the particular shapes and lengths differ for different vowels. This property is not a trivial one, since it is not common to the consonants. Consonantal gestures do not aim toward asymptotic targets. Nor (insofar as it is accurate at all to describe consonants in terms of targets) are their targets *global* shapes of the vocal tract.

Along similar lines, a given vowel is an equivalence class of gestures that are equivalent because they all aim toward some particular limiting shape and length of the vocal tract. As we will argue, this view of vowels as gestures differs from a "targets" view more or less as a curve, say a hyperbola, differs from its asymptotes. We partition vowel production into three components, control of vocal tract length, of tongue shape and of global shape of the vocal tract. Our proposal deals only with that the last of these, but we briefly consider the first two.

Perkell (1969) suggests that the lengthening and shortening of the vocal tract are realized by two neuro-muscular systems that coordinate the lips, mandible, hyoid bone and larynx. One coordinative structure governs vocal tract lengthening, including lip rounding, and the other governs shortening for non-high vowels. With respect to tongue shape, Perkell (1969) suggests, as a first approximation, that it is invariant across the vowels and is characteristic of vowels as a class. Some more recent evidence of Lindblom and Sundberg (1971) suggests that Perkell's claim is a simplification; their data reveals *three* shapes of the tongue for 11 Swedish vowels. Tongue shape, therefore, as a responsibility of the intrinsic musculature of the tongue, may or may not be set invariantly across the class of vowels.

Of immediate concern here is control over global vocal tract shape as determined primarily by the position of the tongue in the vocal tract. Tongue position is governed by the extrinsic tongue muscles and, hence, our metaphor concerns the organization imposed on those muscles when vowels are produced.

To describe the production of vowels, we are looking for a system for positioning of the tongue which captures those properties of the class of vowels that are equivalent across its membership. An equation like that of a vibratory system, say $F = -s(l-l_0)$, with the parameter l_0 unspecified, would satify that criterion because it represents a system which establishes l_0 as a variable (tunable) dimension for the class of vowels. In addition, we are looking for a description of a particular vowel that enables us to describe as equivalent the different gestures that instantiate it in different contexts. (The reader is reminded that for /ɛ/, the tongue is lower following /i/, but raised following /a/.) Possibly l_0 will work to do that, if we describe it at an abstract level as the zero-state of the extrinsic tongue system. (At a less abstract level, each extrinsic muscle may have its own l_0.) l is the actual state of the tongue system (the actual position of the tongue).

Consider what happens to a spring system when F and s are unchanged, but l_0 is reduced in magnitude. To counteract the same $-F$, the system decreases l by the same amount as l_0 was decreased. Thus l alters *in the direction of the new* l_0. (For example, let $l_0 = 10$ in arbitrary units, $F = 50$, $s = 25$. Transforming the equation above, $l = l_0 - (F/s), = 10 - 2 = 8$. If now l_0 is reset to 5, $l = 5 - 2 = 3$.) Again, suppose that l_0 corresponds to the zero-state of the extrinsic tongue system—to the position that the tongue would adopt if $-F = 0$, and l is the actual position of the tongue. If a talker is able to alter l_0 volitionally, as the subjects of Asatryan and Fel'dman (1965) could, then he can change the zero-state of the extrinsic musculature. Suppose that corresponding to a particular vowel is a particular value of l_0, initiated by changing the value of l_0. In consequence of this, and *regardless of its current* value, l, the actual position of the tongue in the mouth, alters its value in the direction of l_0. Hence, in the case of the vowel /ε/ its l_0 is less than that for the vowel /i/ and greater than that for /a/. When $l_{øε}$ is substituted for $l_{øi}$, the tongue is lowered; when it is substituted for $l_{øa}$, the tongue is raised. Although the gesture is different for /ε/ in the different contexts, the parameter $l_{øε}$ is the same.

This proposal is highly speculative, of course, and it is likely to be falsified in its *particulars* (e.g. in the proposal that the extrinsic muscles of the tongue are organized to work as a *linear* spring). What is critical to the proposal, however, is the general suggestion that the invariants for vowels are *organizations* of the musculature. The organizations do not yield invariant movements, nor do they yield static configurations of the vocal tract (nor, for that matter, do they yield simple acoustic correlates). Thus a vowel, /ε/ for example, is essentially the functioning of the coordinative structure encompassing the extrinsic tongue muscles (among other coordinative structures) when a tunable parameter of the system is given its /ε/-characteristic value. Thereby, /ε/ is canonically dynamic. Furthermore, it is context-free since, regardless of context, it is always produced when its invariant parameter value is assigned to the mapping governing vowel production.

It is perhaps worth emphasizing how this view is distinct from one equating a vowel with a spatial target or a feature bundle. The parameter l_0 itself may turn out to be indistinct from a spatial target or a feature bundle. However, l_0 is just a parameter *of* a system. As we argued earlier, a parameter is an *attribute* of something; it is not something in itself. Hence, it makes no sense to speak of the parameter apart from the mapping that it parameterizes.

Furthermore, an advantage of including the mapping in a vowel's specification—that is, of including the description of the muscle organization underlying vowel production—is that the specification captures the dynamic character of vowels. Movement can be considered inherent to vowel production. In contrast, to exclude the muscle organization is to make the implausible claim that, ideally, vowels are static. That is why we suggest that it is as incorrect to equate $l_{øε}$ with /ε/ as it is to equate a curve with its asymptote. $l_{øε}$ is just the limiting shape of the vocal tract towards which /ε/ invariably aims.

To summarize, we suggest that vowels, as they are known and produced are "organizational invariants" of the vocal tract. As such they are dynamic and context-free. Furthermore, they are *separate* from consonants (and hence their serial ordering with respect to consonants can be detected by a perceiver) because the organizational invariants for vowels perpetrate a different *kind* of articulatory and acoustic event than those for consonants.

Finally, we return to our earlier discussion of the degrees of freedom problem and its implicit solution in theories of speech production. As we reported, these theories suppose talkers to include in their articulatory plans only specifications for the vocal tract that require idiosyncratic moment-to-moment control. In particular, suprasegmental regularities are excluded from the plan. This proposal makes the degrees of freedom problem manageable, but does not explain how and why suprasegmental properties are governed, nor how they are coordinated with segments.

We have suggested that vowels as an articulatory class are characterized by an invariant, tunable organization of the musculature. Ohman (1966, 1967) and after him Perkell (1969), argue that, for the most part, different muscles are involved in vowel and consonant production. Thus, vowels and consonants can be, and are, coarticulated. In addition, Ohman suggests that vowels are continuously produced, at least in a vowel-consonant-vowel utterance, and his suggestion is given support by the recent work of Barry and Kuenzel (1975) and Butcher and Weiher (1976). All of this is compatible with the proposal that a talker establishes the invariant organization of his musculature for vowels at the beginning of his utterance and, insofar as possible, maintains it throughout. Thus, his articulatory plan need not repeatedly reestablish the vowel characteristic vocal-tract organization. From moment to moment it has only to provide the appropriate parameter values that select the particular vowels he wishes to produce. This invariant organization of the musculature, with a duration that is suprasegmental, may be responsible for the prosody of speech. Consonant with this proposal, prosody is typically considered to be "carried" by the vowels.

Thus, at least with respect to vowels, theories of speech production may be correct in treating something like the spatial target as a constituent of a plan for the *moment-to-moment* control of speech. However, these targets are assigned *to* a muscle organization that a talker also regulates, though at a slower pace. In short, the degrees of freedom problem may be solved by a talker regulating as much as possible at a slower rate, and as little as possible from moment to moment.

The work reported here was supported by the following grants to Haskins Laboratories: NIH grants HD01994 and NS13617, NSF grant BNS76-82023.

References

Abbs, J. H. (1973a). Some mechanical properties of lower lip movement during speech production. *Phonetica* **28**, 65–75.

Abbs, J. H. (1973b). The influence of the gamma motor system on jaw movements during speech: a theoretical framework and some preliminary observations. *Journal of Speech and Hearing Research* **16**, 175–200.

Abbs, J. H. and Eilenberg, G. R. (1976). Peripheral mechanisms of speech motor control. *In* "Contemporary Issues in Experimental Phonetics" (N. J. Lass, ed.), pp. 139–168. Academic Press, New York.

Abbs, J. H. and Netsell, R. (1973). A dynamic analysis of two-dimensional muscle force contributions to lower lip movement. *Journal of the Acoustical Society of America* **53**, 295.

Abbs, J. H., Folkins, J. W. and Sivarajan, M. (1976). Motor impairment following blockade of the infraorbital nerve: Implications for the use of anesthetization techniques in speech research. *Journal of Speech and Hearing Research* **19**, 19–35.

Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behavior* **3**, 111–150.

Amerman, J. D. and Daniloff, R. G. (1971). Articulation patterns resulting from modification of oral cavity size. *ASHA* **13**, 559.

Arbib, M. A. (1972). "The Metaphorical Brain. An Introduction to Cybernetics as Artificial Intelligence and Brain Theory". J. Wiley & Sons, New York.

Arshavskzy, Yu. I., Kots, Ya. M., Orlovsky, G. N., Rodionov, I. M. and Shik, M. L. (1965). Investigation of the biomechanics of running by the dog. *Biophysics* **10**, 737–746.

Arutyunyan, G. A., Gurfinkel, V. S. and Mirskii, M. L. (1968). Investigation of aiming at a target. *Biophysics* **13**, 642–645.

Arutyunyan, G. A., Gurfinkel, V. S. and Mirskii, M. L. (1969). Organization of movements on execution by man of an exact postural task. *Biophysics* **14**, 1162–1167.

Asatryan, D. and Fel'dman, A. (1965). Functional tuning of the nervous system with control of movement or maintenance of a steady posture—I. Mechanographic analyses of the work of the joint on execution of a postural task. *Biophysics* **10**, 925–935.

Ashby, W. R. (1956). "An Introduction to Cybernetics". J. Wiley & Sons, New York.

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

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

Bernstein, N. (1967). "The Coordination and Regulation of Movements". Pergamon Press, London.

Borden, G. J., Harris, K. S. and Oliver, W. (1973a). Oral feedback I. Variability of the effect of nerve-block anesthesia upon speech. *Journal of Phonetics* **1**, 289–295.

Borden, G. J., Harris, K. S. and Catena, L. (1973b). Oral feedback II. An electromyographic study of speech under nerve block anesthesia. *Journal of Phonetics* **1**, 297–308.

Boylls, C. C. (1975). A theory of cerebellar function with applications to locomotion. II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat. *COINS Technical Report* 76–1, Department of Computer and Information Science, University of Massachusetts.

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

Carpenter, M. B. (1972). "Core Text of Neuroanatomy". Williams and Wilkins, Baltimore.

Chomsky, N. (1965). "Aspects of the Theory of Syntax". MIT Press, Cambridge, Mass.

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

Donegan, P. and Stampe, D. (1977). Natural phonology. Paper presented at the Conference on the Differentiation of Current Phonological Theories. Indiana University, Bloomington, September 30.

Draper, M. H., Ladefoged, P. and Whitteridge, D. (1959). Respiratory muscles in speech. *Journal of Speech and Hearing Research* **2**, 16–27.

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

Eccles, J. C. (1969). The dynamic loop hypothesis of movement control. *In*

"Information Processing in the Nervous System" (K. N. Leibovic, ed.), pp. 245–269. Springer-Verlag, New York.

Evarts, E. V., Bizzi, E., Burke, R. E., DeLong, M. and Thach, W. T. (1971). Central control of movement. *Neurosciences Research Program Bulletin* **9**, No. 1.

Fel'dman, A. G. (1966a). Functional tuning of the nervous system with control of movement or maintenance of a steady posture—II. Controllabe parameters of the muscles. *Biophysics* **11**, 565–578.

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

Fitch, H. and Turvey, M. T. (1978). On the control of activity: Some remarks from an ecological point of view. *In* "Psychology of Motor Behavior and Sports" (R. W. Christina, ed.), Human Kinetics, Urbana, Illinois.

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

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

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

Fowler, C. A. and Turvey, M. T. (1978). Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. To appear in "Information Processing in Motor Control and Learning" (G. Stelmach, ed.), Academic Press, New York.

Garner, W. R. (1974). "The processing of Information and Structure". Lawrence Erlbaum Associates, Potomac, Maryland.

Gay, T. and Ushijima, T. (1974). Effect of speaking rate on stop consonant-vowel articulation. *In* "Proceedings of the Speech Communication Seminar, Stockholm, 1974". Almqvist and Wiksell, Uppsala, pp. 205–209.

Gel'fand, I. M. and Tsetlin, M. L. (1962). Some methods of control for complex systems. *Russian Mathematical Surveys* **17**, 95–116.

Gel'fand, I. M. and Tsetlin, M. L. (1971). Mathematical modeling of mechanisms of the central nervous system. *In* "Models of the Structural-Functional Organization of Certain Biological Systems" (I. M. Gel'fand, V. S. Gurfinkel, S. V. Fomin and M. T. Tsetlin, eds), pp. 1–22. MIT Press, Cambridge, Mass.

Gibson, J. J. (1950). "The Perception of the Visual World". Houghton-Mifflin, Boston.

Gibson, J. J. (1961). Ecological optics. *Vision Research* **1**, 253–262.

Gibson, J. J. (1966). "The Senses Considered as a Perceptual System". Houghton-Mifflin, Boston.

Gottlieb, G. L. and Agarwarl, G. C. (1972). The role of the myotatic reflex in the vocabulary control of movements. *Brain Research* **40**, 139–143.

Gottlieb, G. L. and Agarwarl, G. C. (1973). Coordination of posture and movement. *In* "New Developments in Electromyography and Clinical Neurophysiology" (J. E. Desmedt, ed.), pp. 418–427. S. Karger, Basel.

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

Greene, P. H. (1972). Problems of organization of motor systems. *In* "Progress in Theoretical Biology, Vol. 2" (R. Rosen and F. Snell, eds), pp. 304–322. Academic Press, New York.

Grillner, S. (1975). Locomotion in vertebrates: central mechanisms and reflex interaction. *Physiological Reviews* **55**, 247–304.
Groesberg, S. W. (1968). "Advanced Mechanics". J. Wiley & Sons, New York.
Gurfinkel, V. S., Kots, Ya. M., Pal'tsev, E. I. and Fel'dman, A. G. (1971). The compensation of respiratory disturbances of the erect posture of man as an example of the organization of interarticular interaction. *In* "Models of the Structural-Functional Organization of Certain Biological Systems" (I. M. Gel'fand, V. S. Gurfinkel, S. V. Fomin and M. L. Tsetlin, eds), pp. 382–395. MIT Press, Cambridge, Mass.
Hamlet, S. L. 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.
Harris, K. S. (1976). The study of articulatory organization: Some negative progress. *In* "Dynamic Aspects of Speech Production" (M. Sawashima and F. S. Cooper, eds), pp. 71–82. University of Tokyo, Tokyo.
Hintikka, J. (1975). Information, causality and the logic of perception. *In* "The Intention of Intentionality and Other New Models for Modalities" (J. Hintikka, ed.), pp. 59–75. D. Reidel, Dordrecht, Holland, Boston.
Hockett, C. (1955). "A Manual of Phonology". International Journal of Linguistics. Waverly Press, Baltimore.
Hubbard, A. W. (1960). Homokinetics: Muscular function in human movement. *In* "Science and Medicine of Exercise and Sport" (W. R. Johnson, ed.), pp. 7–39. Harper, New York.
Ito, M. (1970). Neurophysiological aspects of the cerebellar motor control system. *International Journal of Neurology* **7**, 162–176.
Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics* **14**, 201–211.
Kaplan, H. M. (1971). "Anatomy and Physiology of Speech". Second edition. McGraw-Hill, New York.
Kent, R. D. (1976). Models of speech production. *In* "Contemporary Issues in Experimental Phonetics" (N. Lass, ed.), pp. 79–104. Academic Press, New York.
Kent, R. D. and Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics* **5**, 115–133.
Kent, R. D. and Moll, K. L. (1975). Articulatory timing in selected consonant sequences. *Brain and Language* **2**, 304–323.
Kent, R. D. and Netsell, R. (1971). Effects of stress contrasts on certain articulatory parameters. *Phonetica* **24**, 23–44.
Kent R. D., Carney, P. J. and Severeid, L. R. (1974). Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research* **17**, 470–488.
Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* **59**, 1208–1221.
Kulagin, A. S. and Shik, M. L. (1970). Interaction of symmetrical limbs during controlled locomotion. *Biophysics* **15**, 171–178.
Kuypers, H. G. J. M. (1958). Cortico-bulbar connexions to the pons and lower brain stem in man. *Brain* **81**, 364–388.
Ladefoged, P. (1968). Linguistic aspects of respiratory phenomena. *Annals of the New York Academy of Sciences* **155**, 141–151.
Lambertsen, C. J. (1974). Neurogenic factors in control of respiration. *In* "Medical

Physiology, Vol. 2" (V. B. Mountcastle, ed.), pp. 1423–1446. C. V. Mosby, St. Louis.

Lashley, K. S. (1951). The problem of serial order in behavior. *In* "Cerebral Mechanisms in Behavior: the Hixon Symposium" (L. A. Jeffress, ed.), pp. 112–136. J. Wiley & Sons, New York.

Lee, D. N. (1974). Visual information during locomotion. *In* "Perception: Essays in Honor of James J. Gibson" (R. B. MacLeod and H. L. Pick, Jr., eds), pp. 250–267. Cornell University Press, Ithaca, New York.

Lenneberg, E. H. (1967). "Biological Foundations of Language". J. Wiley & Sons, New York.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review* **74**, 431–461.

Liberman, A. M. and Studdert-Kennedy, M. (1978). Phonetic perception. *In* "Handbook of Sensory Physiology, Vol. VIII, 'Perception'" (R. Held, H. Leibowitz and H.-L. Teuber, eds). Springer-Verlag, Heidelberg.

Lieberman, P. (1967). "Intonation, Perception and Language", MIT Press, Cambridge, Mass.

Lindblom, B. and Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw and larynx movement. *Journal of the Acoustical Society of America* **50**, 1166–1179.

Lindblom, B. and Sundberg, J. (1973). Neurophysiological representation of speech sounds. Paper presented at the XV World Congress of Logopedics and Phoniatrics, Buenos Aires, Argentina.

Lindblom, B., Lubker, J. and Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation *Journal of Phonetics*. **7**, 147–162.

Lisker, L. (1974). On time and timing in speech. *In* "Current Trends in Linguistics (Vol. 12)" (T. A. Sebeok, ed.), pp. 2387–2418. Mouton and Co., The Hague.

Lisker, L. (1976). Phonetic aspects of time and timing. Paper presented at the 100th meeting of the American Speech and Hearing Association, Washington, D.C., November, 1975. Also in *Haskins Laboratories Status Report on Speech Research* SR–47, 113–120.

Mace, W. (1977). James J. Gibson's strategy for perceiving: Ask not what's inside your head, but what your head's inside of. *In* "Perceiving, Acting, and Knowing: Toward an Ecological Psychology" (R. Shaw and J. Bransford, eds), pp. 43–65. Erlbaum, Hillsdale, New Jersey.

MacKay, D. G. (1969). Forward and backward masking in motor systems. *Kybernetik* **2**, 57–64.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* **8**, 323–350.

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

MacNeilage, P. and Ladefoged, P. (1976). The production of speech and language. *In* "Handbook of Perception, Volume VII" (E. C. Carterette and M. P. Friedman, eds), pp. 75–120. Academic Press, New York.

Moll, K. L., Zimmerman, G. N. and Smith, A. (1976). The study of speech production as a human neuromotor system. *In* "Dynamic Aspects of Speech Production" (M. Sawashima and F. S. Cooper, eds), pp. 107–127. University of Tokyo, Tokyo.

Netsell, R. and Abbs, J. H. (1975). Modulation of perioral reflex sensitivity during speech production. Paper presented to the *Acoustical Society of America*, San Francisco, California.

Ohman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* **39**, 151–168.
Ohman, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America* **41**, 310–320.
Pattee, H. (1973). The physical basis and origin of hierarchical control. *In* "Hierarchy Theory: The Challenge of Complex Systems" (H. Pattee, ed.), pp. 71–108. Braziller, New York.
Perkell, J. S. (1969). "Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study". MIT Press, Cambridge, Mass.
Pew, R. W. (1974). Human perceptual-motor performance. *In* "Human Information Processing. Tutorials in Performance and Cognition" (B. H. Kantowitz, ed.), Erlbaum, New York.
Pisoni, D. (1978). Speech perception. To appear in "Handbook of Learning and Cognitive Processes" (W. K. Estes, ed.), Erlbaum, Hillsdale, New Jersey.
Powers, W. T. (1973). "Behavior: The Control of Perception". Aldine, Chicago.
Putnam, A. H. B. and Ringel, R. L. (1976). A cineradiographic study of articulation in two talkers with temporarily induced oral sensory deprivation. *Journal of Speech and Hearing Research* **19**, 247–266.
Remez, R. E. and Rubin, P. E. (1975). Is speech like *anything* else people do, or is it like *everything* else people do? Unpublished manuscript, University of Connecticut.
Ringel, R. L. and Steer, M. D. (1963). Some effects of tactile and auditory alterations on speech output. *Journal of Speech and Hearing Research* **6**, 369–378.
Riordan, C. J. (1977). Control of vocal-tract length in speech. *Journal of the Acoustical Society of America* **62**, 998–1002.
Saltzman, E. (1977). On levels of sensorimotor representation. Unpublished manuscript, University of Minnesota.
Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review* **82**, 225–260.
Scott, C. M. and Ringel, R. L. (1971). Articulation without oral sensory control. *Journal of Speech and Hearing Research* **14**, 804–818.
Sears, T. A. and Newsom Davis, J. (1968). The control of respiratory muscles during voluntary breathing. *Annals of the New York Academy of Science* **55**, 183–190.
Shankweiler, D., Strange, W. and Verbrugge, R. (1977). Speech and the problem of perceptual constancy. *In* "Perceiving, Acting, and Knowing: Toward an Ecological Psychology" (R. Shaw and J. Bransford, eds), pp. 315–345. Erlbaum, Hillsdale, New Jersey.
Shaw, R. and Pittenger, J. (1977). Perceiving the face of change in changing faces: Implications for a theory of object perception. *In* "Perceiving, Acting, and Knowing: Toward an Ecological Psychology" (R. Shaw and J. Bransford, eds), pp. 103–132. Erlbaum, Hillsdale, New Jersey.
Shaw, R., McIntyre, M. and Mace, W. (1974). The role of symmetry in event perception. *In* "Perception: Essays in Honor of James J. Gibson" (R. B. MacLeod and H. L. Pick, Jr., eds), pp. 276–310. Cornell University Press, Ithaca, New York.
Shaw, R., Turvey, M. T. and Mace, W. (In press). Ecological psychology: The consequence of a commitment to realism. "Cognition and the Symbolic Processes, Vol. 2" (W. Weimer and D. S. Palermo, eds), Erlbaum, Hillsdale, New Jersey.
Shik, M. L. and Orlovsky, G. N. (1965). Coordination of the limbs during running of the dog. *Biophysics* **10**, 1148–1159.
Shik, M. L. and Orlovsky, G. N. (1976). Neurophysiology of locomotor automatism. *Physiological Reviews* **56**, 465–501.

Shik, M. L., Orlovsky, G. N. and Severin, F. V. (1966). Organization of locomotor synergism. *Biophysics* **11**, 1011–1019.

Stevens, K. N. (1975). The potential role of property detectors in the perception of consonants. *In* "Auditory Analysis and the Perception of Speech" (G. Fant and M. A. A. Tatham, eds), pp. 303–330. Academic Press, New York.

Studdert-Kennedy, M. (1978). Universals in phonetic structure and their role in linguistic communication. "The Recognition of Complex Acoustic Signals" (T. H. Bullock, ed.), Dahlem Konferenzen, Berlin.

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

Timoshenko, S. and Young, D. H. (1948). "Advanced Dynamics". McGraw-Hill, New York.

Turvey, M. T. (1977a). Preliminaries to a theory of action with reference to vision. *In* "Perceiving, Acting, and Knowing: Toward an Ecological Psychology" (R. Shaw and J. Bransford, eds), pp. 211–265. Erlbaum, Hillsdale, New Jersey.

Turvey, M. T. (1977b). Contrasting orientations to a theory of visual information processing. *Psychological Review* **84**, 67–88.

Turvey, M. T. and Shaw, R. (1979). The primacy of perceiving: An ecological reformulation of perception as a point of departure for understanding memory. "Perspectives on Memory Research: Essays in Honor of Uppsala University's 500th Anniversary" (L-G Nilsson, ed.). Erlbaum, New Jersey.

Turvey, M. T., Shaw, R. and Mace W. (1979). Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions. *In* "Attention and Performance, VII" (J. Requin, ed.), Erlbaum, Hillsdale, New Jersey.

Ushijima, T. and Hirose, H. (1974). Electromyographic study of the velum during speech. *Journal of Phonetics* **2**, 315–326.

Warren, R. M. (1976). Auditory illusions and perceptual processes. *In* "Contemporary Issues in Experimental Phonetics" (N. Lass, ed.), pp. 389–417. Academic Press, New York.

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

Wyke, B. D. (1967). Recent advances in the neurology of phonation: phonatory reflex mechanisms in the larynx. *British Journal of Communication Disorders* **2**, 2–14.

Wyke, B. D. (1974). Laryngeal myotatic reflexes and phonation. *Folia Phoniatrica* **26**, 249–264.

Zemlin, W. R. (1968). "Speech and Hearing Science: Anatomy and Physiology". Prentice-Hall, Englewood Cliffs, New Jersey.