542

*Journal of Phonetics* (1986) **14**, 171–196

# Intentional contents, communicative context, and task dynamics: a reply to the commentators

**J. A. S. Kelso,[1,2] E. L. Saltzman[2,3] and B. Tuller[1,2]**

[1]*Center for Complex Systems and Department of Psychology, Florida Atlantic University, Boca Raton, Florida 33431, U.S.A.*
[2]*Haskins Laboratories, New Haven, Connecticut 06511–6695, U.S.A.*
[3]*Department of Psychology, University of Connecticut, Storrs, Connecticut 06268, U.S.A.*

---

## 1. Introduction

In a wonderful little book called *Where Is Science Going*? translated into English in 1933, Max Planck (1933) tells a story about when the first railways spread throughout the countryside, and how the peasants—unaware of any other force than the muscular one generated by beasts and people—used to bet with each other as to how many horses were concealed inside the locomotive. This was no less of an error than the ancient Greeks attributing thunder to the personal anger of Zeus. One cannot help wondering, in these days in which psychological and linguistic explanation rests strongly (if not solely) on the existence of mental states and their assumed contents, whether the same mistake in kind is being made as that by the peasants of bygone days. Even fields such as motor control and neurobiology, which one might think do not require mentalistic language (henceforth "mentalese"), are riddled with computer jargon, as *Grillner* notes. Scratch below the surface and one finds posited an intelligent agent generating an intelligent action. To still others, mental (non-physical) events can actually produce physical consequences (Eccles, 1985). Must the payback of these loans on intelligence (Dennett, 1978) await the scientific equivalent of a Wall Street crash? To be a non-mentalist, it seems, is to be accused of ignoring the very stuff of living systems—goal-directedness and intentionality. Without mentalese (according to some) the chances of explaining the latter are "utterly remote" (*Lindblom & MacNeilage*).[1]

---

[1] In our view the *Lindblom & MacNeilage* position on this issue has a Jekyll and Hyde flavor. In a very recent paper (Lindblom, MacNeilage & Studdert-Kennedy, 1984), the same authors argue strongly that the axiomatic postulation of segments and features (i.e. symbolic contents), by conventional linguistic theory, should be rejected because "It deliberately avoids addressing the issue of explaining where segments and features come from" (p. 187). According to Lindblom *et al.* (1984) this " . . . must be regarded as THE [capitals theirs] most fundamental question of theories aiming at providing truly explanatory accounts of sound patterns" (p. 113). While they criticize us as *non-mentalists* in their present commentary, earlier they shun what they call the "mentalistic" approach, i.e. "any explanatory attempt that involves elements of conscious effort on the part of the originators of the phonemic principle" (p. 184). We prefer their stance of a year ago (the Dr. Jekyll version) than the one they seem to adopt in their KST commentary (the Mr. Hyde version). For example, Lindblom *et al.* (1984) advocate a single research strategy: "*DERIVE LANGUAGE FROM NON-LANGUAGE*" (p. 187, their capitals). Based on a rather superficial treatment of self-organizing systems which "so far linguists do not seem to have discovered—or have not attempted to explore" (p. 185), they suggest the following major guideline (p. 187): Seek " . . . acceptable explanations among accounts that derive segments and features deductively from independently motivated principles". These are the very principles of pattern formation in complex systems, i.e. of self-organization and non-equilibrium cooperative phenomena, that we have sought to

*cont.*

0095–4470/86/010171 + 26 $03.00/0

Of course, the origins of intentional contents are as much a problem, if not *more* of a problem for a mentalist approach, as we shall emphasize below: but first we want to correct an apparent misconception. *Our approach does not ignore intentionality*; to the contrary, we view intentionality as a primary objective property of a complex system that is open to lawful analysis. In fact, a central task for the ecological approach is to come to grips with the laws at the terrestrial scale on which the intentionality of perception–action systems is founded (cf. Gibson, 1979; Turvey, in press; Turvey, Shaw, Reed & Mace, 1981). To do so is to avoid the reduction of intentionality to mental states, or even to *equate* intentions with mental states. A main theme of our paper (henceforth reduced, after *Lindblom & MacNeilage*, to KST) as well as earlier work, is that the same fundamental activity can be instantiated or realized by many different kinds of structures (cf. Section 2.1). Nature is diverse with her forms, but sparse with her principles of functioning. If this ancient conjecture is correct, then the nervous system can best be considered a supportive *medium* for the playing out of dynamical laws, rather than as the producer of intentions *qua* mental states.

As Iberall & Soodak (in press) note, the great merit of physical science is its ability to characterize phenomena—from the laboratory to the farthest reaches of nature—in the same language. This has required impressive powers of abstraction on the part of the scientist. But somehow there are cracks in the edifice when it comes to the issues germane to the linguist, phonetician, or even the biologist, all of whom deal with complex processes. Somehow physics, which our commentators, almost to a person, appear to equate (in the context of speech production) with fields such as *biomechanics* or *acoustics* (or generally, "low-level processes") is inadequate to the task of so-called "higher-order" organizational processes. But KST do not adhere to this seemingly local and concrete interpretation of "physics". In fact, the dynamical approach strips away material properties (whose operation is treated usually by Newtonian mechanics) in a serious effort to understand the organizational processes themselves—especially self-organization. These are formalizable in purely abstract but potentially universal terms, as discussed in Section 4 of our paper and elsewhere. Several of our commentators seem to have taken a similar view quite recently. Thus, in a recent paper, Kent (1985) says: "I propose that speech aquisition can be understood as a self-organizing process" (p. 29), and Lindblom, MacNeilage & Studdert-Kennedy (1984) say: "What we have in mind are the accounts provided by the theory of self-organizing systems" in which "THE INTERACTION OF SUBSYSTEMS—BE THEY SUBSYSTEMS OF MATTER OR INFORMATION OR BEHAVIOR—GIVES RISE TO STRUCTURATION (Haken, 1981)" (p. 185, their capitals). The conceptual and mathematical framework for self-organization is difficult. There are different approaches, some of which work and others do not (for a highly informed review, see Landauer, 1981). Pertinent fields of study range from thermodynamics through non-linear dynamics and statistical mechanics. The potential applications are equally far-reaching. Though the established linguist or phonetician will

---

elaborate in the context of movement coordination and speech production since the late 1970s (cf. KST, Section 4). "Only", according to Lindblom *et al.* (1984) "when such attempts (to derive the pattern of segments and features as a self-organizing system) have been *exhaustive* [italics added] are we entitled to conclude that we are probably dealing with properties that are unique to language" (p. 187). But now, a year or so later, such attempts are considered "naive" and "utterly remote". In their deliberate choice of mentalistic language (e.g. *Lindblom & MacNeilage*, Section 4B) we can only assume they have decided to wait no longer—that they believe attempts to derive intentional contents from independently motivated principles have been exhausted. To the contrary, we say such attempts have only begun; indeed, many do not appreciate the problem or simply sweep it under the rug.

hardly be expected to be familiar with the details (but hopefully the newcomer might), it was a surprise to us nevertheless that those commentators who have explicitly "proposed" self-organizing accounts of sound pattern formation and language acquisition (i.e. *Kent* and *Lindblom & MacNeilage*) chose to remain mute on Section 4 of KST, which describes some of the central concepts. Indeed, of the commentaries only *Ostry*'s seems to appreciate the strategic "physics" advocated by the dynamical approach—more of this below.

In the deliberations that follow, where common themes exist among the commentaries we shall try to highlight them. The extensive critique by *Lindblom & MacNeilage* provides a basis through which such themes may be struck. Moreover, *Lindblom & MacNeilage* raise several detailed concerns about the task-dynamic model which we address in a separate section. An appendix that provides the kernel mathematical details of the task-dynamic approach is also provided. Finally, to the extent that individual commentaries fall outside a thematic approach, we shall address them separately.

## 2. On exposition

### 2.1. *Prefatory remarks*

Some of the commentators, but especially *Lindblom & MacNeilage* have problems with the exposition of our model as well as a number of other, specific topics such as the role of feedback in compensation. If our paper existed in a vacuum or as an isolated entity we could sympathize, but it does not—nor was it intended to do so. Our equations and terms have been explicitly defined in other places. In particular, Saltzman & Kelso was published in preliminary, but detailed, form in 1983 and is now in press in *Psychological Review*. Additional background papers are Saltzman (in press), Kelso & Tuller (1984*a*, 1985*a*); see original target article References Section. We have addressed the role of feedback in compensation in numerous places both in the motor control literature (e.g. Kelso, 1977; Kelso & Holt, 1980; Kelso, Holt & Flatt, 1980; Kelso & Stelmach, 1976) and the speech literature (Kelso & Tuller, 1983; Kelso, Tuller, V.-Bateson & Fowler, 1984; Zimmermann, Kelso & Lander, 1980), including this journal (Tye, Zimmermann & Kelso, 1983). That our work can be interpreted by *Lindblom & MacNeilage* as "ruling the somatosensory component out of court" is a most mysterious statement which we invite the international phonetic community to evaluate for themselves. In addition, we have recently addressed in considerable detail how the concept of *information* can be construed within the dynamical framework (Kelso & Kay, in press).

Our approach in the target article was to provide the gist of what the mathematical formalisms do in our model. In one respect our approach was similar to that adopted in a paper by Lindblom *et al.* (1984) which dealt, in one section, with "simulations of emerging phonetic structure" (see footnote 1). Lindblom *et al.* (1984) preface their modeling work with the following statement (p. 188):

> The following section summarizes the results of two sets of computational experiments described in greater detail elsewhere (Lindblom *et al.*, forthcoming). There are certain assumptions that are common to all the simulations and represent independently given *a priori* information. For their mathematical definition see Lindblom *et al.* (forthcoming).

This is an entirely reasonable caveat, given as *Lindblom & MacNeilage* (this commentary) "realize that the mathematics needed . . . may not facilitate rapid understanding . . ." However, at the Editor's request and in order to correct certain misconceptions, we will here repeat and amplify our stance on issues that we have addressed before on numerous occasions.

### 2.2. *Terminology*

*Lindblom & MacNeilage, Fujimura* and, in a slightly different context, *Grillner* raise questions about some of the terminology used in our paper. Why, *Lindblom & MacNeilage* ask, should we phoneticians be expected to be familiar with terms such as task dynamics, point attractor, limit cycle, etc.? The answer is: "You should not, unless you think, as we do, that such concepts might be of some use in understanding the organization among articulators during the act of speech".[2] In their concluding remarks, *Lindblom & MacNeilage* criticize us for failing to take " . . . the opportunity provided by this *Journal* to make intellectual contact with the international phonetic community", in part, no doubt because of our terminology (see Section 3.11 of *Lindblom & MacNeilage*), *Fujimura* feels forced to "rephrase" some of our "claims" (which he appears to understand perfectly) because our terminology is not based on "generally accepted physical or engineering usage". From *Fujimura*'s remarks, it is clear that he thinks the set of terms used by physics and engineeering is complete—new ones only confuse. We disagree, especially when supposedly "new" terms are needed to convey essential concepts (though actually, they are not new at all; see below). But our point is that physics and engineering are *evolving* sciences: they have not stood still. There are many examples in the history of science (some quite recent) in which the words "not generally accepted" (hence not "rigorous and concise") have provided an easy substitute for ignorance.[3]

*Lindblom & MacNeilage* have themselves dabbled in popularist accounts of self-organization (Lindblom *et al.*, 1984), even a cursory probing of which would reveal that the term "attractor" is a central concept. This is because self-organizing systems are fundamentally dissipative. The significance of this fact for theories of speech production is that dissipation reduces the high dimensionality of the system to one of much lower dimension. Essentially, the *N*-dimensional phase space describing the motion of all relevant system degrees of freedom is compressed (not conserved as in classical Hamiltonian systems) onto a surface of lower dimensionality (e.g. Lichtenberg & Lieberman, 1983). For example, the multi-articulator trajectories, constituting a flow in phase space, that converge—from various initial conditions and despite deviations along their path—to a tongue–palate configuration appropriate for the vowel /i/, can be usefully described in the low-dimensional language of point attractor dynamics (see, e.g. Kelso & Tuller, 1983). Speech "goals" and "dissipation" thus go hand in hand: the final condition, the articulatory system's basin of attraction, is observed as a consequence of dissipative dynamics.

---

[2] The standard call at speech production conferences is always for "more data" and "more precise measurement", e.g. of tongue movement, which is all to the good. But strangely, the call is seldom for "new concepts". The facts, though we must get them, will not arrange themselves.

[3] A prime example was the growth of *Deutsche Physik* which accompanied the rise of National Socialism in Germany. As Rosen (1985) notes, it was basically the threat of a new theoretical framework, especially general relativity, that caused a rebellion among some experimental physicists. Ugly consequences followed. Those who "studied equations instead of nature" were deemed traitors to the "true" science, which could only be carried out by an experimenter who possessed "a kind of mystic rapport with nature through direct interaction with it" (Rosen, 1985, p. 86).

Such concepts arise in the treatment of many body problems in general, and are far from "new" or absent from "general usage". As Wightman (1985) informs us, they are among the oldest parts of physics, though the modern period began with Poincaré in the late 19th century. And, as Wightman (1985) continues, Poincaré's legacy has become "dreadfully fashionable" (p. 22) in the past 15 years. For the community unfamiliar with some of the constructs we used (which, as we mentioned in Section 2.1, are defined explicitly in the context of skilled actions and speech production in earlier papers) we recommend the following:

Abraham and Shaw (1982), Vols. 1–3 (for a wonderful visual introduction to the field of dynamical systems, with few words and fewer equations)
Eckmann (1981)
Feigenbaum (1980)
Haken (see Section 4 of KST)
Helleman (1980)
May (1976)
Shaw (1981)

From a quick purview of these papers it will be immediately apparent that the terminology we use is, in fact, in quite general use (Helleman's article, for example, has 302 references and makes contact with many applications); that it possesses a long and admirable history; and that it represents one of the most exciting developments in modern science (see, e.g. Haken, 1985). If *we* have failed to communicate with the international phonetic community (which we hope—contrary to the opinion of *Lindblom & MacNeilage*—is not the case), then the above references may allow the international phonetic community to communicate with the field of dynamical systems in general, and our work in particular. Phonetics must decide, as (contrary to *Fujimura*'s view) physics and engineering clearly have, whether the terminology we use, here and elsewhere, is helpful or not. It will depend, obviously, on how the problems are defined. A benefit might be that some of the problems of using jargon across disciplines, referred to by *Grillner*, may be avoided rather than exacerbated by using constructs that are linked to dynamical principles. Indeed, they can be usefully applied, we claim, to explain the very property *Grillner* and we wish to understand, namely, pattern generation.

## 3. On the task-dynamic model

In the Appendix we present an expanded treatment of the three mathematical steps (task space, tract variables, and model articulators) of the task-dynamic simulations that were described more briefly in KST. Along with information presented in Saltzman & Kelso (1983 and in press), these added details should allow the reader to reproduce the computational results that KST discuss. Here, we discuss a fourth computational step that was omitted from KST, and its implications for several issues raised by our commentators. We also provide additional analyses of the workings of the model during perturbed (and unperturbed) speech gestures.

There is an important final step in the task-dynamic approach which entails using one of two methods (*control law* or *network coupling*; see Saltzman & Kelso, in press, for details) to make a set of "real articulators" behave identically or near-identically to the set of model articulators previously described in KST. This final step was omitted from KST for two reasons. First, both methods exploit algebraic relations between the equations of motion for the model articulators and the real articulators in order to specify patterns of control parameters over time for the real articulators. Saltzman & Kelso (in press) focused on skilled movements performed by "real" upper limbs (biological, robotic, or prosthetic) whose passive linkage biomechanics and control torques can be captured by a relatively simple geometry and corresponding equation of motion. This is not feasible for the speech articulators whose peripheral biomechanics are much more complex, e.g. the passive tissue properties and muscular forces of the tongue and lips. However, we assume that were they known, the biomechanical equations of motions for the speech articulators would allow the specification of peripheral controls in much the same manner as was theorized for the limbs. Secondly, and relatedly, the essence of the task-dynamic approach lies in its account of the coordinated movement patterns that emerge in a task-specific and posturally conditioned form in the model articulators. Coordinated patterns for skilled actions do not arise from the peripheral biomechanics, although the action system can certainly exploit the constraints supplied by these intrinsic peripheral dynamics. Consequently, we felt that a focus on coordination in the model articulators, and a finessing of the issue of "real" articulatory control was justified in KST for purposes of brevity and simplicity.

As it turned out, however, this omission appears to have muddied the conceptual waters of the task-dynamics section in KST (Section 2.2) more than it simplified or clarified them. In particular, the questions raised by *Lindblom & MacNeilage* regarding the reparameterization of articulatory "targets" and the role of peripheral feedback necessitate a brief overview of the two methods of "real" articulatory control described in greater detail in Saltzman & Kelso (in press). The *control law* method is a viable one for control and coordination of artificial (robotic, prosthetic) systems, and specifies "real" articulatory controls via feedback laws, i.e. as non-linear functions of the on-going peripheral state of the real articulators. This scheme is not biologically plausible for two reasons. First, coordination would be impossible in "deafferented" situations where peripheral state information is unavailable; and secondly, there is no parameter corresponding to articulatory *rest position* or *rest configuration.* Evidence for control of this parameter and its trajectory during the course of limb movements has been provided by Bizzi and his colleagues (e.g. Bizzi & Abend, 1982; Bizzi, Accornero, Chapple & Hogan, 1982). The *network coupling* method is more biologically plausible since it can generate coordinated movements in simulations of deafferented preparations and includes a *rest configuration* parameter that evolves autonomously during a movement according to task-dynamic constraints.

In the network coupling method, schematized in Figure 1, the equations of motion for the model articulators ($\phi$-variables) and real articulators ($\theta$-variables) are coupled to one another in two ways. First, the on-going configuration ($\phi$) of the model articulators is used as the on-going rest configuration ($\theta_0$) for the real articulators. Thus, motions of the model articulators are considered to drive motions of the real articulators. Secondly, peripheral feedback information from the real articulators to the model articulators (hypothesized as $\alpha(\phi - \theta)$ and $\beta(\dot{\phi} - \dot{\theta})$ in Figure 1) serves to modulate motions of the model articulators. Significantly, the real articulators are capable of moving in a task-specific and coordinated (albeit probably degraded) way even when the system is deafferented, i.e. when peripheral feedback information to the model articulators is eliminated. This is due to the fact that coordination of the model articulators can still occur due to the "internal feedback" intrinsic to the model articulators' equation of motion (e.g. equations (4) and (5) in the Appendix). This use of the term internal

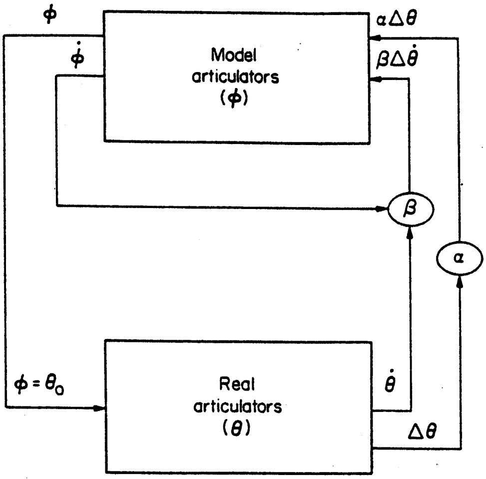

**Figure 1.** Overview of information flow in the network coupling method of task dynamics.

feedback is consistent with Evarts' (1971) definition of information "arising from structures within the nervous system" as opposed to proprioceptive state information arising in the musculoskeletal periphery.

The details given in the Appendix concerning the equation sets involved in task-dynamic simulations may be used to answer some of the *Lindblom & MacNeilage*'s questions concerning the processes involved in modeling speech gestures that are subjected to transient or sustained mechanical perturbations. There are four main points of clarification, as follows.

(1) *Role of feedback.* Mechanical perturbations affecting the articulatory periphery induce corresponding (or at least related) changes in the model articulators via the peripheral feedback terms from the real articulators to the model articulators (see Figure 1). Briefly tugging the real jaw would induce a similar tug on the model jaw; a sustained perturbation (e.g. a bite-block or "clamping" force) that fixes the real jaw at a given position would fix the model jaw at the corresponding position.

(2) *Targets and reparameterization.* Invariant gestural targets are defined at the tract-variable level, e.g. $z_0$ in equation (2) of the Appendix. At the level of the model articulators, coordinated motions evolve so as to attain these gestural targets. If a perturbation is induced in the model jaw, the processes governing the motions of the entire set of model articulators are *exactly the same* as they would be during unperturbed gestures. The tract variable goals are attained, but with different motion patterns (and hence final configurations) for the model articulators. As discussed above, these different motion patterns for the model articulators correspond by definition to different trajectories for the rest configuration vector that serves as a control input to the real articulators (see Figure 1). These different rest configuration trajectories emerge from the natural unfolding of movements in the model articulators. The entire system is only parameterized once at the tract-variable level for a given gesture. The different trajectories for the model articulators (and hence for the real articulators' rest configuration vector) are the automatic consequences of the functioning of the model, and reflect neither reparameterization nor the replanning of trajectories. In fact, there are no explicit trajectory plans for movement

(or "velocity prescriptions" as *Kent* suggests in his commentary) in the task-dynamic simulations. Rather, trajectories for the real and model articulators arise ultimately as implicit consequences of the invariant dynamical systems (control regimes) specified at the tract-variable level.

(3) *Gestural timing*. Task dynamics provides a reasonable account of interarticulator coordination within single speech gestures. It cannot (yet) account for patterns of intergestural coordination, e.g. the relative timing among gestures involved in producing a sequence of vowels and consonants. However, the task-dynamic model can be used to clarify one aspect of relative timing phenomena that appears to be a source of contention in several of the commentaries (*Lindblom & MacNeilage* and *Lubker*). These authors state correctly that when a bite-block prevents the jaw from participating in reiterant speech production, relative timing among the gestures cannot occur with reference to the *jaw's* phase angle in these cases. We agree. *Lindblom & MacNeilage* conclude that the tendency to maintain normal acoustic timing during reiterant bite-block speech implies "that a more abstract representation of speech timing than phase angle was available". This is reasonable and was acknowledged in KST (footnote 4) and elsewhere (Kelso & Tuller, 1985 and in press). However, task dynamics provides an abstract alternative, namely the phase angle of the "vocalic" tract variables associated with the tongue dorsum–palate constriction. The phase portraits of these tract variables (tongue dorsum constriction location and degree) would be defined during reiterant productions despite the bite-blocked jaw due to the coordinated motion of the unblocked tongue relative to the fixed jaw. Presumably these tract-variable trajectories could be computed from X-ray, ultrasound, or magnetometer data in order to provide a more veridical measure of a vocalic gesture during speech production. Similarly, for bilabial reiterant speech, the implication is that consonantal gesture onsets would be best measured by observing the motion trajectory of the tract variable of lip aperture.

(4) *Details of perturbed bilabial closure simulations* (Section 2.2, Example 2 in KST). *Lindblom & MacNeilage* comment that the behavior displayed by the task-dynamic model in normal and perturbed cases is "exciting news" but regret that "the problem of how compensatory manoeuvres come into play nevertheless remains obscure to the reader" (p. 6) and is "far from clear" (p. 6). Here we provide clarifying details concerning the model's behavior during perturbed and unperturbed discrete bilabial closures. (Readers may wish to skip this subsection on their first reading.)

Simulated jaw perturbations do not affect motion along the tract variable of lip protrusion, since the jaw is not part of the articulatory "synergy" for this tract variable. This is strictly a consequence of the way that the simplified model articulatory degrees of freedom are defined in our software articulatory synthesizer. Consequently, our comparison of perturbed and unperturbed gestures will focus mainly on the tract variable of lip aperture and its associated articulatory synergy (jaw, upper lip, lower lip), and particularly on the behaviors of the perturbed articulator (jaw) and the articulator that is most "remote" to the perturbation (upper lip).

In both perturbed and unperturbed gestures, the model articulators begin with the same initial configuration (shown in Figure 2A in KST) and with initial articulatory velocities of zero. Critically damped, second-order dynamics (damping ratio = 1.0; natural frequency = 2 Hz) are specified for the tract variables of lip protrusion and aperture. This results in parameter matrices for the tract-variable motion equation (see equation (2) in the Appendix) whose off-diagonal elements equal zero and whose diagonal elements are as follows: $m_B = 1.0$, $b_B = 25.1$, $k_B = 157.9$. A negative target value was specified for lip aperture's rest position parameter, $z_{02}$, and corresponds to lip

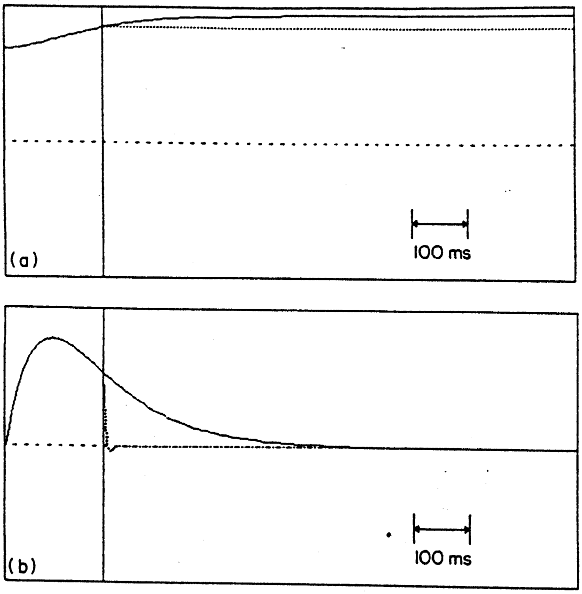

**Figure 2.** Simulated free (——) *vs.* blocked (·····) jaw kinematics. (a) Jaw angle (arbitrary units) *vs.* time (0–1000 ms). Upward motion indicates jaw-raising. (b) Jaw angular velocity (arbitrary units) *vs.* time (0–1000 ms). Dashed horizontal line denotes zero velocity. Solid vertical line at 175 ms denotes perturbation onset in all subfigures.

closure occurring with a given amount of compression ($z_{02}$ = 0.0 corresponds to lip contact or zero aperture, and $z_{02} > 0.0$ corresponds to open lips or positive aperture). In the model articulator motion equation (equation (5) in the Appendix), the articulators are weighted equally by defining the articulatory weighting matrix, $A$, to be equal to the identity matrix. The only difference between the unperturbed and perturbed similations is that during the perturbed similation the jaw is "clamped" in place when it reaches a pre-specified angle. This is accomplished in the simulations by adding a critically damped strong spring (natural frequency = 45 Hz) to the model jaw's rotation axis in the model articulators' motion equation (equation (5) in the Appendix). The rest angle of the perturbing spring is preset to equal this clamping angle. The jaw's behavior in the free and blocked gestures is shown in Figure 2(a) (jaw angle, $\phi_2$, in equations (3), (4), and (5) in the Appendix) and Figure 2(b) (jaw angular velocity, $\phi_2$, in equations (3), (4), and (5) in the Appendix). When the jaw clamp is added, jaw-raising stops within approximately 25 ms of the perturbation onset, and the jaw is "frozen" at the prespecified clamping angle.

However, when one examines the trajectories for lip aperture ($z_2$ in equation (2) in the Appendix) under both free and blocked conditions (Figure 3(a)), two facts are evident: (1) the same degree of lip closure (compression) is attained in the steady-state for both conditions; and (2) the time of attaining lip contact (zero aperture) is hardly affected by the perturbation, with the perturbed contact achieved approximately 10–15 ms after the control. (Note that in other simulations with earlier perturbation onsets, the perturbed contacts could occur earlier than the controls, but again no more than 10–15 ms apart.) In Figures 3(a) and (b) (lip aperture velocity, $\dot{z}_2$, in equation (2) in the Appendix —note that positive velocity corresponds to increasing aperture, i.e. lip opening, while

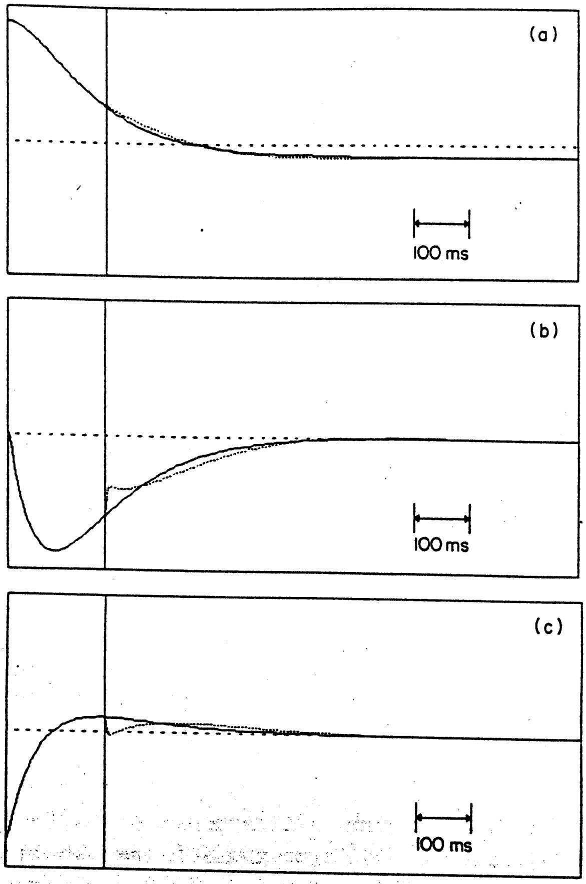

**Figure 3.** Simulated free (—) *vs.* blocked (·····) lip aperture kinematics. (a) Position (arbitrary units) *vs.* time (0–1000 ms). Dashed horizontal line denotes zero aperture or lip contact. Downward motion denotes decreasing lip aperture. (b) Velocity (arbitrary units) *vs.* time (0–1000 ms). Dashed horizontal line denotes zero velocity. (c) Total force (arbitrary units) *vs.* time (0–1000 ms). Dashed horizontal line denotes zero force. Solid vertical line at 175 ms denotes perturbation onset in all subfigures.

negative velocity corresponds to decreasing lip aperture, i.e. lip closing), we see that the jaw perturbation slows the rate of lip closing temporarily relative to the control trial, but that this is soon "compensated" for by a period of increased closing velocity relative to the control.

Figure 3(c) shows the trajectories for the total "force" at the tract-variable level for lip aperture during perturbed and free trials. Here, total lip aperture force is equal to the lip aperture component of the right-hand side of equation (2) in the Appendix, i.e. lip aperture total force ($f_2$) is equal to the sum of the corresponding damping force ($f_{d_2}$) and stiffness force ($f_{s_2}$) components. Since the mass matrix in this equation is the identity matrix, the total lip aperture force is equal to the total lip opening acceleration ($\ddot{z}_2$). Perturbation onset occurs during the acceleration phase of lip opening (i.e. the deceleration phase of lip closing), and the "immediate compensatory" response for lip aperture is a decrease in lip opening acceleration (i.e. an increase in lip closing acceleration) relative to the control.

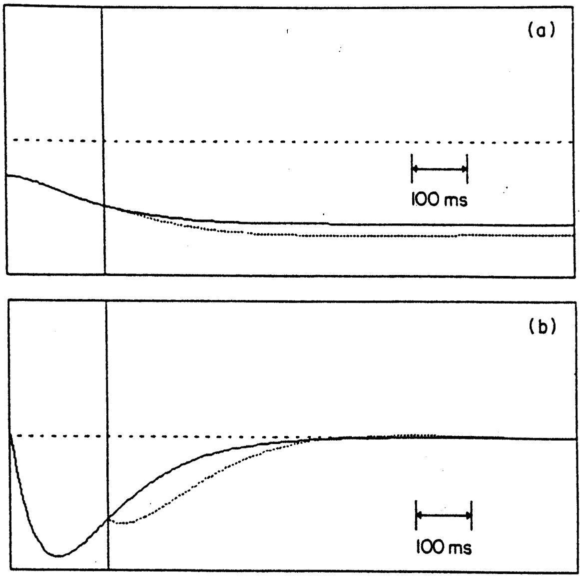

**Figure 4.** Simulated free (—) *vs.* blocked (·····) vertical kinematics of the upper lip. (a) Position (arbitrary units) *vs.* time (0–1000 ms). Downward motion denotes lip lowering. (b) Velocity (arbitrary units) *vs.* time (0–1000 ms). Dashed horizontal line denotes zero velocity. Solid vertical line at 175 ms denotes perturbation onset in all subfigures.

Lip aperture position trajectories are so similar for free and jaw-blocked trials because, in the model, both the upper and lower lips increase their motion "intensities" after the jaw is blocked. (Note that since the lower and upper lips behave as mirror images of each other in these simulations, the lower lip will not be discussed explicitly.) The upper lip ($\phi_3$ in equations (3), (4), and (5) in the Appendix) moves downward to a greater extent in the blocked than the free trials (Figure 4(a)). This is accomplished by an increase in post-perturbation downward velocity ($\dot{\phi}_3$ in equations (3), (4), and (5) in the Appendix) in the blocked trials, while keeping total movement time approximately the same during free and blocked trials (Figure 4(b)).

Figure 5(a) shows the trajectories for the upper lip's total vertical acceleration ($\ddot{\phi}_3$ in equations (3), (4), and (5) in the Appendix) in the free and blocked trials. Referring to equation (5) in the Appendix, it can be seen that $\ddot{\phi}_3$ is the sum of three components: a damping component ($\ddot{\phi}_{d_3}$; see Figure 5(b)), a stiffness component ($\ddot{\phi}_{s_3}$; see Figure 5(c)) and a velocity product component ($\ddot{\phi}_{vp_3}$). Comparing equations (2) and (5), we see that $\ddot{\phi} = \ddot{\phi}_d + \ddot{\phi}_s + \ddot{\phi}_{vp} = J^*\mathrm{F} + \ddot{\phi}_{vp}$, where $J^*$ is the weighted pseudo inverse and $\mathrm{F} = \mathrm{F}_d + \mathrm{F}_s$. Thus, $\ddot{\phi}_3 = \ddot{\phi}_{d_3} + \ddot{\phi}_{s_3} + \ddot{\phi}_{vp_3} = J^*_{31}f_1 + J^*_{32}f_2 + \ddot{\phi}_{vp_3}$. $J^*_{31}$ is always zero since, by definition, lip protrusion ($z_2$) movements are unrelated to upper lip vertical ($\phi_3$) movements, and $J^*_{32}$ varies only slightly about a value of 0.31 in our simulations. Furthermore, in these simulations $\ddot{\phi}_{vp_3}$ is negligible in magnitude relative to the other two components, and can be safely ignored in further discussion. (Note that $\ddot{\phi}_{vp_3}$ is effectively and uniformly zero throughout the free and blocked trials, due to the fact that $\ddot{\phi}_{vp} = J^*\dot{J}\dot{\phi}$ in equation (5) and the elements of $\dot{J}$ are very small, i.e. the Jacobian matrix changes very slowly during these simulated bilabial gestures.) Consequently, $\ddot{\phi}_3 \cong J^*_{32}f_2$ as evidenced by the marked similarity in shape of the trajectories for $f_2$ ($=\ddot{z}_2$; see Figure 3(c)) and $\ddot{\phi}_3$ (see Figure 5(a)).

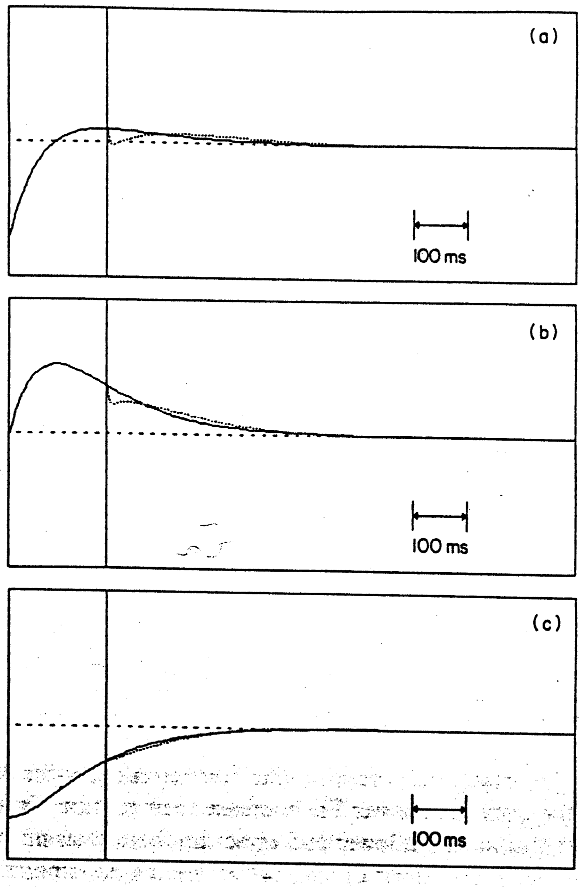

**Figure 5.** Simulated free (—) *vs.* blocked (·····) vertical acceleration components of the upper lip. Total (a), damping (b), and stiffness (c) acceleration components (arbitrary, but identical units in all these parts) *vs.* time (0–1000 ms). Dashed horizontal lines denote zero acceleration. Solid vertical line at 175 ms denotes perturbation onset in all subfigures.

Taking each acceleration component separately, we can interpret the immediate compensatory vertical response of the upper lip as being due to both a stiffness and a damping contribution. The immediate increase in the stiffness component of lip closing acceleration during the perturbed trials reflects the fact that, all other articulatory kinematic factors being equal, a configuration with a more open jaw is associated with a greater "restoring" acceleration toward the lip closure target. The immediate decrease in the damping component of lip opening acceleration (i.e. decreased lip closing deceleration) during the perturbed trials reflects the fact that, all other articulatory kinematics being equal, a reduced jaw closing velocity results in a decreased amount of cross-articulatory damping of the upper lip's downward motion by the jaw's upward motion. When jaw velocity goes to zero, the upper lip is totally released from this kind of damping.

To summarize, coordinated patterns of articulator movements arise in the task-dynamic model due to the manner in which task-specific and state-dependent "forces" at the tract-variable level (e.g. $f_2$ in Figure 3(c)) are partitioned into corresponding patterns of acceleration (e.g. $\ddot{\phi}_3$ in Figure 5(a)) at the model articulatory level. Importantly, the model functions in exactly the same way during perturbed and unperturbed motions.

## 4. On relative timing/phase stability

Both *Löfqvist* and *Lubker* describe experimental extensions of the data reported in Section 3.3 of KST. The *Löfqvist* data are particularly interesting in that they include vowels of different phonological durations. Thus, vowel duration in the *Löfqvist* corpus varies with segmental composition as well as with suprasegmental assignment. The data are remarkably similar to those we report. (1) Overall high correlations are obtained between the duration of the vowel-to-vowel cycle in VCV sequences, and the duration of the consonant-related latency. (2) The observed correlations are higher than what would be expected on the basis of part-whole correlations. (3) In general, although changes in duration of the measured intervals are strongly correlated, the ratio between intervals does not remain constant. We encourage *Löfqvist* to subject the movement trajectories to a phase analysis, analogous to the one we present, as a test of our hypothesis.

The data reported by *Lubker* are more difficult to assess. By his own admission, *Lubker* is uncertain whether any normalization of the data was performed, and if so, which algorithm was used. In fact, the figures suggest no normalization at all. (The reason for normalizing is described in footnote 5 of the target article.) Furthermore, in KST we make a statistical argument for stability, i.e. non-significant effects of stress, rate, or their interaction on the set of tokens of an utterance type produced by each speaker. We do not assert that individual tokens will show no variation at all, although the variation across the set (the standard error) is small. Although such a thorough analysis of *Lubker*'s data should have been possible, given his description of the experiment, all we are shown, in his Figures 2 and 3, are single tokens from one subject's productions of two reiterant and four real-speech utterances. Where are the five normal speakers' data? Why were no statistics computed? Could it be that *Lubker*'s subjects produced only a single token of each experimental word? Regardless of whether we agree with *Lubker* that the main task of science is to disprove hypotheses (we do not; see Munhall & Kelso, 1985, for an elaboration of our position), a far more rigorous analysis of *Lubker*'s data is essential before they can be interpreted as support or disconfirmation of our views. Two additional points follow.

(1) *Löfqvist, Lubker*, and *Lindblom & MacNeilage* all question our use of jaw movements as indicative of a vowel gesture. Let us underscore that we do not equate jaw movements with the vowel but do believe that when jaw movements are allowed in sequences composed of labial consonants and low vowels, jaw lowering will be *related to* vowel production (see Section 3.3 for consideration of bite-blocked speech). Nevertheless, we recognize that production of any sound probably involves a constellation of articulators whose relative contribution is a function of segmental and suprasegmental context, situational speaking style, etc. In fact, our research has not been restricted to jaw movements as indicative of vocalic events: In Harris, Tuller & Kelso (1986), we report results of an X-ray microbeam study that demarcated both vowel-related and consonant-related articulatory events in tongue movements, a compound movement trajectory of the sort described by *Kent*.

(2) Because we, as investigators, can describe a system as stable in its durational characteristics, this does not necessarily mean that the system controls duration *per se*. Thus, the observation that bite-block speech preserves the temporal characteristics of unimpeded speech does not necessarily imply that speakers adjust their articulatory patterns *in order to* maintain durations of individual elements. Yet we do not propose, as *Lubker* seems to think we do, that the temporal stability "simply 'fall(s) out' from the planning for displacement and velocity", nor do we require "velocity prescriptions" as *Kent* suggests (see also Section 3.2). Surely the past decade of motor systems work indicates strongly that single-joint movements, and complex multiple-joint movements viewed from the end effector, act qualitatively like mass–spring systems. In such dynamical systems, duration and velocity of movements to an equilibrium position are largely functions of system stiffness. Neither duration nor velocity (nor for that matter, displacement) are "planned" system parameters.

The mistaken assumption that what the experimenter chooses to measure must be what the speaker controls is also apparent in *Lubker*'s assertion that speakers directly control timing because "In Swedish, for example, duration information can be critical in differentiating between otherwise identical words". Tellingly, phonologically long and short vowel pairs in, for example, Estonian also differ slightly in formant frequency (cf. Lehiste, 1970). To our knowledge, investigators have yet to characterize acoustic formant transitions, or better yet, articulator trajectories, to and from vowels that differ "only" in duration.

## 5. On falling short of objectives

Our target article represents a very small part of a program that is geared to elucidating the (lawfully based) constraints that link animals to their environments. A flavor of the broader scope of the program is provided in Warren & Shaw (1985). A paper such as ours has obvious limitations, most notably its confinement to certain perceived issues in speech production. Invariance (stability) in the face of phonetic variation (change), controlling a system of speech articulators, and articulatory pattern formation (the assembly of a functional unit of action) were central themes. We were pleased that some of the commentators advanced implications for speech perception (*Lindblom & MacNeilage* and *Kent*), ontogeny of speech organization (*Kent*), phonological structure (*Fujimura* and *Löfqvist*), and speech motor control (*Grillner* and *Ostry*).

On the other hand, the comments indicate some profound misunderstandings that we find ourselves unable to ignore. References are made, for example, to our supposed adherence to a "solely bottom-up", approach, our failure to consider the actor's intentions, our lack of "ecological validity" (whatever that means), our advocation of a scenario that ignores the listener, the environment and the communicative function of speech in general. *All of these are just plain wrong*. Lack of emphasis in the present paper should not be mistaken to represent crucial failures of the overall program or lead to false "impressions" on the part of the commentators or the community. The main perpetrators are *Lindblom & MacNeilage* to whom we now address ourselves.

### 5.1. *On intentionality ignored*

*Lindblom & MacNeilage* argue that our program cannot be successful as long as we limit our inquiry to an exclusively bottom-up approach and ignore "mental" goals (see footnote 1). They say:

> Without such goals . . . [to] constrain the selection of neural connections, metabolic processes, muscles and joints etc. that participate in the formation of functional units the CS (coordinative structure) could not possibly be derived.

And further

> The CS is not derivable from a purely bottom-up approach, since it arises as an interface between intention and its output.

In their concluding remarks they reiterate the latter point sharply. Our "chances" of responding to this crucial criticism are " . . . non-existent not only because of the magnitude of the task but also because *CSs must be seen as arising from interaction between intentions and the motor system*" [emphasis added].

It seems that we are supposed to take these remarks as highly critical of the dynamical approach. But how can *Lindblom & MacNeilage* tag us as a bottom-up approach (a term we have never used, see *Fowler* reply), ignoring goals and intentions? Other commentators appear to imply the same deficit. *Ostry* wonders whether our emphasis on physical principles creates difficulties for "relating behaviors to the structure of nervous systems or to actor's intentions". We say yes, but no more and probably much less than other frameworks not grounded in dynamics (see below). *Kent* quotes a study reporting slower lip and jaw velocities in children than adults that he finds interesting because the authors conclude that "higher-order organizational factors and lower-level, physical structures and capabilities" account for the differences (although the link between the observation in this case and the conclusion is mysterious). Presumably *Kent* believes that we embrace the latter to the exclusion of the former. In his discussion of compound trajectories (which seems akin to what KST (Section 2.2) refer to as gestural coproduction), *Kent* stresses that phonetic goals need to be properly considered. No argument—unless he thinks that we have ignored phonetic goals.

As for *Lindblom & MacNeilage*, whose interpretation of us as "bottom-uppers" is unambiguous, even though incorrect, we suspect their "impressions" are vestiges of the past. The term "coordinative structure" was first coined by Easton (1972) and was reflex-based. Fowler's (1977) thesis, Turvey's (1977) earlier work (which was available some years earlier) and a consequent article in this journal (Fowler, 1980) all tended to equate the CS with a reflexive kind of organization. The CS was associated more with the *actualization* of an action, rather than a unitary assemblage of components that carried its own intrinsic meaning or significance. Readers can evaluate for themselves from the following quotes whether *Lindblom & MacNeilage* have fairly represented the CS as an intentionless, bottom-up, and hence sterile construct.

For example, at a conference held by MacNeilage in 1981 (see MacNeilage, 1983) we said:

> Coordinative structures are functional units in the sense that the individual degrees of freedom constituting them are *constrained by particular behavioral goals or effectivities* (Turvey & Shaw, 1979; emphasis ours). Sharing the same degrees of freedom without reference to the effectivity engaged in by an actor would not constitute a functional unit (Kelso, Tuller & Harris, 1981/1983, pp. 141–142).

In 1982, Kelso, Tuller & Fowler (1982) concluded a study in which highly distinctive forms of compensation were observed in remote articulators to the same perturbation as follows:

> These highly distinctive patterns suggest that the jaw, lips and tongue may be controlled and coordinated as a single, functional unit (a coordinative structure) that *is unique and specific to the intended act.* [emphasis added]

And, in 1984, Kelso, Tuller, V.-Bateson & Fowler (1984) proposed an experimental hypothesis about the CS (p. 813):

> In the case of speech, the components of the articulatory apparatus would cooperate in such a way as to preserve the *linguistic intent of the speaker.*

Our position (ibid., p. 813), which proved correct, was that:

> Different patterns of articulator cooperation (coordinative structures) should occur tailored to the *particular phonetic requirements.*

Do these quotations reflect *Lindblom & MacNeilage*'s claim that our program "limits its inquiry to an exclusively bottom-up approach" and hence must be doomed to failure? Or that we " . . . seem to believe that the only constraints on speech production come from the speech motor system itself?" (*Lindblom & MacNeilage*). *The answer is no, on both counts.*

### 5.2. *On situational context ignored*

After Sommerhoff (1969), "goals" can be conceptualized as objective properties of a system, not subjective, consciously purposed things. We can assume them—at the very least—to be compatible with natural law. We might go further and say that intentions marshall lawful relations among a system's component structures (see below). But the action or response of a system is not itself a sufficient condition for a goal event. The latter is conditional on having a specific relation to the environment. Hence, in the case of speech, we continually refer to the importance of *linguistic intents* and *phonetic goals* as special boundary conditions on a softly assembled group of neuromuscular components[4]. By definition, the CS has a communicative role in speech. In our experiments, the task of the subject is to produce a sound or sequence of sounds that can be clearly understood by a listener, i.e. produce a specific environmental consequence. Does this suggest a non-ecological scenario in which the communicative context is ignored? Clearly not.

---

[4] A theme reiterated recently in a *Neurosciences Abstract* by Abbs, Shaiman, Gracco & Cole (1985). They refer to data that " . . . support the supposition that neural networks underlying a given class of voluntary motor acts are configured and reconfigured, flexibly, by subtle changes in intended goals". This is another way of stating the important fact that the kind of coordination one *observes* among muscles and articulators is task-dependent. But no one has identified or observed "neural networks" in these kinds of experiments, as *Grillner* notes.

[5] Grillner (1985) reports a series of findings that " . . . indicate that the CNS and the spinal cord contain central pattern generators (CPGs) that can produce a complex motor output *similar* to that of locomotion" (p. 144, emphasis added). But what does the word "similar" mean here? Does it or does it not produce the locomotor pattern? "Similarity" is a very tricky concept (see Munhall & Kelso, 1985) and not one we would have expected in the context of CPGs.

In 1981 in a paper on apraxia (Kelso & Tuller, 1981) we made our position very clear on the importance of situational context:

> We wish to propose an alternative theoretical perspective . . . in which a movement (or any event) is not defined independent of the context in which it occurs. In this view, in response to questions like "Who wants to go with me?", "How many oranges do you have?" and "How high can you reach?", the gesture of an outstretched hand with all five fingers extended has very different meanings. In short, the hand gesture is *not* functionally equivalent in different contexts. [p. 232]

> Just as the situational context provides boundary conditions or constraints on the possible meaning of the movement, so also the possible meanings of the movement provide boundary conditions on the movement's dynamic forms. [ibid., p. 233]

> Accordingly, the interaction between the individual and the context or environment must be an adaptive one whose fit is functionally defined by the particular behavioral goal. [ibid., p. 233]

Though we apologise for the *ex cathedra* style we believe it necessary to lay to rest the erroneous conclusion by *Lindblom & MacNeilage* that KST offers a view that is " . . . not only non-ecological but naïve in the light of the enormous range of phonetic variation observed in natural speech". KST present data and theory emanating from research that *exploits* systematically manipulated phonetic variation—as well as a number of other factors known to increase speaker-listener variability—in order to discover invariances. That theme is at the very heart of our research program, and has been for many years. It is *Lindblom & MacNeilage* that are non-ecological in the following sense: they have taken KST out of context and in doing so interpreted omissions to fatal flaws in our approach. The reader, we trust, will be more objective and do the necessary scholarship.

## 6. On intentional contents

Let us discuss briefly certain aspects of intentionality (origins, contents) that were raised by some of the commentators within the context of coordinative structures. We can quickly dispatch *Lindblom & MacNeilage*'s question, "Where do intentions come from?" here. This is like asking a cosmologist: Where does the Universe come from? The answer is that no one knows, and the level of abstraction necessary to even approach the problem is difficult to imagine. Our bet is that an intention, in some sense of the word, is an abstraction of the system's (cooperative) dynamics, somewhat akin, perhaps, to an order parameter (cf. KST, Section 4). This is obviously only a place-marker statement on a topic that goes far beyond the time and space restrictions of the present reply. But let us discuss briefly, how the dynamical approach, and the ecological program in general, particularly as elaborated by Turvey and colleagues (e.g. Kugler & Turvey, in press; Turvey, in press; Turvey *et al.*, 1981) might deal with the problem.

Remember that the thrust of the dynamical approach is to provide a law-based account of movement control and coordination. A law-based explanation (Bunge, 1977; Turvey *et al.*, 1981) involves the laws themselves, of course, but also the *circumstances* (initial conditions, boundary conditions, constraints) that must be brought to bear in order to account best for which *actual* states are realized. As Iberall & Soodak (1978, in press) emphasize, in many physical and biological systems there is great diversity of form, even though science itself rests on the existence of a small set of principles of universal applicability. Thus, though the laws are very general, they receive their expression in a vast number of diverse, local morphologies. None of this is news to the physicist, biologist, or even linguist, though the great diversity can serve to obscure the underlying principles (see, e.g., d'Arcy Thompson, 1917/1947; Stevens, 1974). Diversity and variation, of course, are just other ways of recognizing the diversity of *circumstances* that apply to natural law. For the problems of interest to the phonetician or the behavioral scientist, these circumstances are not only (if at all) structural (cf. *Ostry*). Rather, *intentional contents* constitute a major component of the circumstances that are appended to laws. Intentions, then, within the dynamical approach, can be viewed as exceptional boundary conditions on natural law (e.g., Kugler & Turvey, in press; Turvey *et al.*, 1981; Yates, 1978). Indeed, a way of interpreting our experiments in speech motor control (e.g. Kelso, Tuller & Fowler, 1982; Kelso, Tuller, V.-Bateson & Fowler, 1984) is that the intentional content, e.g. "produce the word 'bob' ", is the boundary condition that harnesses a lawful co-operativity among components.

An example due to Iberall & Soodak (in press) may help clarify this proposal. The key aspect of their description is that there are always three "levels" necessary to create (self) organization. At the lowest level, there are "atomisms" whose particular structure depends on the system being studied (e.g. molecules in a liquid). There is a constraint imposed from above (e.g. the walls in a pipe); and inbetween there is a co-operativity (e.g. an attractor for the flow field). The basic mechanism for the emergence of a co-operativity is that of a generalized Reynolds number, that is, an interplay between the forces sweeping into the system (e.g. inertia), and the forces "holding" (as it were) the system together (e.g. viscosity). As this ratio approaches a critical value, an instability occurs and the emergence of a new patterned form occurs (for example, see Kelso & Tuller, 1984*a*; KST, Section 4).

Our focus has been upon the co-operative level, established by special boundary conditions above (i.e. intentional contents), and the atomisms below (i.e. a system of individual articulators). *A coordinative structure analysis, therefore, requires all three levels: the individual atomisms and their characteristics that are organized into an ensemble, the ensemble operation itself, and the constraint or ordering from above.* Note that in this view, intentional contents are no longer the sole province of "mental states" inside the head (cf. *Lindblom & MacNeilage*). Moreover, in the present scheme, the study of intentionality does not necessarily begin at the level of the *individual* "mind". Rather, intentional contents may themselves arise within larger contextual settings. The onus, therefore, is on science to elaborate the necessary and sufficient conditions for the emergence of intentions, which goes far beyond the analysis of mental and/or brain states. This problem of how constraints (or in the synergetic language adopted in KST, *order parameters*) arise is not limited to the field of linguistics or cognitive science: it has been a persistent focus of physical analysis (e.g. Pattee, 1973).

In Section 4 of KST we emphasized the emergence of new (or different) forms of motion, seen at the ensemble, co-operative level, that correspond to certain kinds of phonetic change. This was an explicit example of how so-called "linguistic elements" could be derived from the (self-organized) cohesive workings among articulators. The theoretical point for speech production (note that we did not say "for linguistic analysis") is that segments, phonemes etc. do not have *explicit* status as constructs in the theory. Our position is thus similar to one taken by Lindblom *et al.* (1984), namely that "they (features, segments, allophonic rules) are, in fact, *imputed* [emphasis added] to our simulated speaker the moment we subject the derived syllables to 'a conventional linguistic analysis'" (p. 200; i.e. their Dr. Jekyll version, see footnote 1).

Our colleague *Fowler* wants us to "come around" to recognize the reality of phonological segments, which for her are "literally uttered". She objects, she says, along with *Lindblom & MacNeilage*, to the notion that an explanation of linguistic units must await a dynamic account of talking (but see footnote 1). Of course, it is the (for us) Mr. Hyde version of *Lindblom & MacNeilage* with which she is in agreement. Alternatively, witness MacNeilage on Lindblom:

> Lindblom's work . . . is representative of a small group of functionalists who *attempt to derive linguistic elements and processes deductively from an extralinguistic base* [emphasis added] in contrast to the more formal approach within linguistics, in which linguistic elements and processes are postulated axiomatically. [MacNeilage, 1983, p. ix]

Or Lindblom *et al.* (1984):

> Currently overly formal approaches to phonology underestimate the role of performance constraints in the formation of sound patterns as well as neglecting the exploitation of the concept of self-organization. [p. 200]

and

> Our description is a phonology without "features", "segments", and "rules" which . . . also manages to dispense with mental blueprints. [Lindblom *et al.*, p. 200]

*Fowler* would presumably object to this (for us, Dr. Jekyll) version of *Lindblom & MacNeilage* (*avec* Studdert-Kennedy) as well. We do not really have any problem in accepting with Fowler that segments are psychologically real. However, we believe that, in the case of speech production, it is necessary to understand how a system of vocal-tract articulators is controlled and coordinated so that sounds are appropriately structured for a perceiver. If *Fowler* could tell us how the concept of segment informs us about the coherent operation of the articulatory system, then we would, in all likelihood, "come around". Unfortunately our own suspicion is that the notion of segment will prove insufficient to capture much about the dynamics of talking. Indeed, we question whether an understanding of the latter and the consequences it produces for a perceptual system needs the concept of "segment" at all.

## 7. Reply to some individual comments

### 7.1. *Fujimura*

*Fujimura* provides the readership with a useful discussion of linearly coupled oscillators (swings). Of course, a main result of our experimental work is that a non-linear oscillator model is needed to guarantee the stable relationship between a movement's amplitude and duration as rate is increased (Kay, Kelso & Saltzman, in preparation; Kelso & Kay, in press). Kay *et al.* show that this function (i.e. its slope and intercept) can be nicely accounted for by a non-linear, limit-cycle oscillator in which only one parameter, linear stiffness, is adjusted to scale movement frequency (see Kelso & Kay, in press, Figure 5). Another main result, discussed in KST (Section 4) is that phase between coupled articulators may not change demonstrably within a wide range of movement frequencies (though it may undergo small fluctuations, see Kelso & Scholz, 1985). Then, at a critical frequency value, the system will jump to a new phase. As Haken, Kelso & Bunz (1985) show, *nonlinearly* coupled systems are needed for such novelty (i.e. a new ordering between the components) to occur. Thus, our modeling approach accommodates a spatiotemporal pattern's stability as well as the instabilities that lead to new pattern formation. All this has been described in a mathematically rigorous fashion (cf. *Fujimura*'s "intriguing" possibilities).

*Fujimura* also provides us with his brief philosophy of modeling and offers some interesting suggestions regarding his iceberg model. His point regarding multiple descriptions is well taken. As Kelso & Kay (in press) note, one of the investigator's tasks in the dynamical approach is to become familiar with the behavioral characteristics of various classes of dynamical systems and to obtain data addressing their similarities and differences. There are difficulties, however. Dynamics typically starts with a set of equations and evaluates their solutions under various conditions, such as changes in parameters and initial conditions. Non-linear dynamical systems, however, generally defy exact solutions and only approximate (via numerical methods) and/or qualitative solutions are usually possible. In speech production we are faced with an even more difficult problem: given a solution, a particular sequence of spatiotemporal articulatory events produced by a speaker in a communicative environment, what kinds of equations could produce this particular solution? This is where dynamical analogy seems crucial: some insight is needed into the similarity between the real event and something we know —such as a non-linear oscillator. Then, if we adapt the thing we know, at least a qualitative model of the data becomes possible. But the question is: From whence cometh the insight?

We have one or two small bones of contention with *Fujimura* which seem worth making. The first is his reference to our phase stability results as "informal observations" and a misunderstanding about the data's source ("Let us assume here that the word is repeated many times"). The data, tabled in KST (Table 1) are actually a complete reanalysis of Tuller & Kelso (1984). In that experiment, four subjects produced 10–12 tokens of each of eight utterance types under four stress and speaking rate combinations. *Fujimura* notes that "the value of [our] claim must depend on the correctness of factual observations". We agree that the facts are important to establish and provide statistical support for our position. With reference to the nine individual trajectories shown in *Fujimura*'s Figure 1, we are forced to repeat what we said to *Lubker*: a rigorous analysis of the data is essential before they can afford serious comparison with other potentially competing views.

Finally, we disagree with the criticism that "our observations are limited to quite artificial situations of speech production", that they are somehow not "natural" (see also *Remez* on *Fowler* and vice versa). In other sciences it is always necessary to simplify material, to delimit problems, to work under somewhat artificial conditions (Stetson, 1951). Research performed within a restricted context is only a problem if, and only if, the dynamics of the phenomenon we want to understand are destroyed in the process of reduction. So-called naturalistic renditions of speech have not told us much (yet) about the speech production process in spite of efforts in the past by experimental phoneticians (see Stetson, 1928/1951, for a review). Perhaps they will as technology and ingenuity make the speech production system more accessible to observation. But gadgeteering and data collecting will not a science make. Principles will still be needed to rationalize the facts and to know what facts to look for (see also footnote 2).

## 7.2. *Grillner*

We have no argument with *Grillner* when he defines the CPG in neuroanatomical terms, i.e. the spinal cord contains a network of nerve cells which when active will activate the different motoneurones and muscles in a specific sequence. And when he acknowledges difficulties with how the network works on a neurone by neurone basis, we empathize (cf. *Lindblom & MacNeilage*'s criticism that a hard-wired unit of action is a strawman). As long as Grillner stays *within* a level of description that he is clearly comfortable with, there is no problem. The neurophysiologist has her/his own language and conventions for describing cells and cellular networks, just as the chemist has for molecules and the physicist for atomic particles. *Grillner* believes, however, that he can explain locomotion and even perception (Grillner, personal communication) via neural networks alone, in a purely reductionist fashion.[5] That is, for *Grillner*, there is an ontological priority of the neural description over others. We believe this is wrong: living organisms can be described as ensembles of subatomic particles, atoms, molecules, cells, cellular "networks", neuromuscular components and so on, but none of these is logically prior to the other. A living organism is all of these at one and the same time.

It is when the CPG is applied to higher-level behavioral properties that difficulties and confusions emerge (e.g. see Grillner's 1985 account of wiggling the big toe). Many in *Grillner*'s field, for example, refer to locomotor CPGs, swimming CPGs, flying CPGs, etc. For us, there is no more a "locomotor" CPG than there is a "selfish" (or homosexual or angry) gene (see Lewontin, Rose & Kamin, 1984). Locomotion and selfishness are macroscopic, behavioral attributes of complex organisms, not CPGs and genes.

We agree with *Grillner* when he says that "terms like CPG or central programs have in general got compromised by having been used within several fields with different meanings". However, we are not sure that he himself is beyond reproach in this matter. Our own detailed concerns about programs and CPGs in the movement literature have been expressed before (for a recent synopsis see Kelso & Scholz, 1985). A main problem is that the "program" metaphor has taken on a real identity and its source (the man-made digital computer) tends to be ignored. Thus, Grillner (1985) in addition to defining a CPG (as he did in his commentary) also says that (p. 144):

> the spinal cord is programmed to counteract the impediment by gating the reflex effects to different muscle groups depending on the part of the step cycle in which the perturbation occurs.

But though we know, according to *Grillner*, that the spinal cord contains CPGs, how do we know that the spinal cord is also "programmed"? And why should the "program" metaphor be appropriate in this context? Is this not a sleight of hand, albeit couched in neurophysiological language? And is this not the kind of "hazardous" step beyond the known facts about CPGs that *Grillner* warns us about?

## 7.3. *Kent*

*Kent* warns us that much is to be lost in our understanding of speech if we simply label it a co-operative process. We agree. Any label such as central pattern generator (*Grillner*) or coordinative structure (*Lindblom & MacNeilage*) is only a start to scientific investigation. We are sorry that *Kent* chose not to comment on Section 4 of KST in which some of the basic rudiments of self-organized, co-operative phenomena are presented and references to explicit examples in the movement literature given. Our approach is to ask: What are the necessary and sufficient conditions for pattern generation to occur, not to simply label it? *Kent* seems to have (temporarily) ignored our earlier efforts to identify coordinative structures with (self-organized) dissipative structures (Kugler, Kelso & Turvey, 1980, 1982) both in his commentary here and in his own recent efforts to link speech acquisition to self-organizing processes (Kent, 1985). Perhaps we are at fault in not making the linkages among concepts as clearly as we should. We would have thought that other, almost synonymous terms, such as non-equilibrium phase transitions, cooperative phenomena, etc., commonly used by workers in this field, would have been known to a proponent of self-organization such as *Kent*. Self-organization *is* a co-operative process; far from being a mere label, it can be explicitly described using the conceptual and mathematical tools of synergetics (cf. KST Section 4 and references therein).

On the other hand, *Kent* is a kindred spirit as far as his dissatisfaction with segmentation in speech is concerned (see Section 6 above). Moreover, his concern for ontogeny and the emergence of articulatory patterns that can structure sound in intelligible ways is one that we also share (see, e.g., Kelso & Clark, 1982). *Kent*'s discussion of velocity profiles in childrens' articulations may pertain to a distinction that we have drawn previously between control and coordination (e.g. Kugler *et al.*, 1980). However, we stress that an understanding of both the metrical (control) and the structural (coordinative) properties of speech can be enhanced by examining the *relations* among kinematic variables, rather than a single kinematic variable. Thus, the intimate relationship between an articulator's velocity and displacement, as observed on the phase portrait, can specify the abstract dynamic (control) parameter of stiffness (e.g. Cooke, 1980; Kelso *et al.*, 1985; Ostry & Munhall, 1985). Similarly, as KST emphasize, the phase stability observed in adult speech is informative about the coupling (or coordination) among articulators. Phase is thus a parameter that captures the ordering relations among components (see KST's discussion of order parameters in Section 4). The dynamical perspective, as *Löfqvist* correctly notes, has stability and change as topics of major concern, topics that are dear to the hearts of developmentalists. We encourage those who study childrens' speech patterns to examine the geometry of the speech system's dynamics as tools to aid their understanding of speech motor development. Time, far from being out of favor with us as *Kent* contends, is *implicitly* given in the evolving geometry of the system's phase plane trajectories. However, it does not, as KST point out, appear *explicitly* in the description. Rather, as Kelso and Tuller (1985*a*, in press) emphasize, a given dynamical system—at whatever stage of its development—generates its own intrinsic time *according to its constitutive parameters*. The latter will depend on a variety of factors, e.g. growth, the maturation of the nervous system, skill level, language experiences, etc. But in response to *Kent* and other commentators, the point is not that we have eliminated time. Time is indeed with us, only not in the way it is traditionally defined, i.e. as conventional or mechanical time (in seconds, hours, etc.) imposed on a system *regardless* of its particular dynamics.

## References

Abbs, J. A., Shaiman, S., Gracco, V. L. & Cole, K. J. (1985) Task dependent sensorimotor actions are inherent in speech motor programs, *Society of Neurosciences Abstracts*, 76.

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

Bizzi, E. & Abend, W. (1982) Posture control and trajectory formation in single and multiple joint arm movements. In: J. E. Desmedt (ed.), *Brain and spinal mechanisms of movement control in man*. New York: Raven Press.

Bizzi, E., Accornero, N., Chapple, W. & Hogan, N. (1982) Arm trajectory formation in monkeys, *Experimental Brain Research*, **46**, 139–143.

Bunge, M. (1977) Levels and reductions, *American Journal of Physiology*, **233**, R75–R82.

Dennett, D. C. (1978) *Brainstorms: philosophical essays on mind and psychology*. Montgomery, VT: Bradford Books.

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

Eccles, J. C. (1985) Complementarity gives a new insight into the mind–brain problem. In: H. Haken (ed.), *Complex systems: operational approaches in neurobiology, physics, and computers*. Heidelberg: Springer-Verlag.

Eckmann, J.-P. (1981) Roads to turbulence in dissipative dynamical systems, *Reviews of Modern Physics*, **53**, 643–654.

Evarts, E. V. (1971) Feedback and corollary discharge: A merging of the concepts. In: *Central control of movement. Neurosciences Research Program Bulletin*, **9**, 86–112.

Feigenbaum, M. J. (1980) Universial behavior in nonlinear systems, *Los Alamos Science*, **1**, 4–27.

Gibson, J. J. (1979) *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Grillner, S. (1985) Personal communication to J. A. S. Kelso, May (with permission).

Grillner, S. (1985) Neurobiological bases of rhythmic motor acts in vertebrates, *Science*, **228**, 143–149.

Haken, H. (ed.) (1985) *Complex systems: operational approaches in neurobiology, physics, and computers*. Heidelberg: Springer-Verlag.

Harris, K. S., Tuller, B. & Kelso, J. A. S. (1986) Temporal invariance in the production of speech. In: J. S. Perkell & D. H. Klatt (editors), *Invariance and variability of speech processes*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Helleman, R. H. G. (1980) Self-generated chaotic behavior in nonlinear mechanics. In: E. C. D. Cohen (ed.), *Fundamental problems in statistical mechanics*, Vol. 5, pp. 165–233. Amsterdam: North-Holland.

Iberall, A. S. & Soodak, H. (1978) Physical basis for complex systems—some propositions relating levels of organization, *Collective Phenomena*, **3**, 9–24.

Iberall, A. S. & Soodak, H. (in press) Towards a physics of self-organization. In: F. E. Yates (ed.), *Self-organizing systems: the emergence of order*. New York: Plenum Press.

Kay, B. A., Kelso, J. A. S., Saltzman, E. L. & Schöner, G. (submitted) The space–time behaviour of single and bimanual rhythmical movements: data and model.

Kelso, J. A. S. (1982) The process approach to understanding human motor behavior. In: J. A. S. Kelso (ed.), *Human motor behavior: an introduction*, pp. 3–19. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Kelso, J. A. S. & Clark, J. E. (eds.) *The development of human movement coordination and control*. New York: John Wiley.

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

Kelso, J. A. S., Holt, K. G. & Flatt, A. E. (1980) The role of proprioception in the perception and control of human movement: toward a theoretical reassessment, *Perception & Psychophysics*, **28**, 45–52.

Kelso, J. A. S. & Kay, B. A. (in press) Information and control: a macroscopic analysis of perception-action coupling. In: H. Heuer & A. F. Sanders (eds.), *Tutorials in perception and action*. Amsterdam: North-Holland.

Kelso, J. A. S. & Stelmach, G. E. (1976) Central and peripheral mechanisms in movement. In: G. E. Stelmach (ed.), *Motor control: issues and trends*. New York: Academic Press.

Kelso, J. A. S. & Tuller, B. (1981) Toward a theory of apractic syndromes. *Brain and Language* **12**, 224–245.

Kelso, J. A. S. & Tuller, B. (1983) "Compensatory articulation" under conditions of reduced afferent information: a dynamic formulation, *Journal of Speech and Hearing Research*, **26**, 217–224.

Kent, R. D. (1985) Developing and disordered speech: strategies for organization. In: J. L. Lauter (ed.), *Proceedings of the Conference on the Planning and Production of Speech in Normal and Hearing-impaired individuals. ASHA Reports*, **15**, 29–37.

Kugler, P. N. & Turvey, M. T. (in press) *Information, natural law, and the self-assembly of rhythmic movement: theoretical and experimental investigations*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Landauer, R. (1981) Nonlinearity, multi-stability, and fluctuations: reviewing the reviewers, *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, **10**, R107–R113.

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

Lewontin, R. C., Rose, S. P. R. & Kamin, A. J. (1984) *Not in our genes*. New York: Pantheon Press.

Lichtenberg, A. J. & Lieberman, M. A. (1983) *Regular and stochastic motion*. New York: Springer-Verlag.

Lindblom, B., MacNeilage, P. & Studdert-Kennedy, M. (1984) Self-organizing processes and the explanation of phonological universals. In: B. Butterworth, B. Comrie & O. Dahl (eds.), *Explanations for language Universals*. The Hague: Mouton.

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

May, R. (1976) Simple mathematical models with very complicated dynamics, *Nature*, **261**, 459–467.

Munhall, K. G. & Kelso, J. A. S. (1985) The role of similarity analysis in movement, *Journal of Motor Behavior*, **17**, 493–498.

Ostry, D. J. & Munhall, K. (1985) Control of rate and duration of speech movements, *Journal of the Acoustical Society of America*, **77**, 640–648.

Pattee, H. H. (1973) Physical problems of the origin of natural controls. In: A. Locker (ed.), *Biogenesis, evolution, homeostatis*. Heidelberg: Springer-Verlag.

Planck, M. (1933) *Where is science going*? (1981 edn.) Woodbridge, CT: Oxbow Press.

Rosen, R. (1985) *Anticipatory systems*. Oxford: Pergamon Press.

Shaw, R. E. (1981) Modeling chaotic systems. In: H. Haken (ed.), *Order and chaos in nature*. Heidelberg: Springer-Verlag.

Sommerhoff, G. (1969) The abstract characteristics of living systems. In: F. E. Emery (ed.), *Systems thinking*. Hammondsworth, England: Penguin.

Stevens, P. S. (1974) *Patterns in nature*. Boston: Little Brown.

Thompson, D. W. (1917/1942) *On growth and form*. Cambridge: Cambridge University Press.

Turvey, M. T. (in press) Intentionality: A problem of multiple reference frames, specificational information and extraordinary boundary conditions on natural law, *The Behavioral and Brain Sciences*.

Turvey, M. T., Shaw, R. E., Reed, E. S. & Mace, W. M. (1981) Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981) *Cognition*, **9**, 237–304.

Tye, N., Zimmermann, G. & Kelso, J. A. S. (1983) "Compensatory articulation" in normal and hearing-impaired speakers: a cinefluorographic study, *Journal of Phonetics*, **11**, 101–115.

Warren, W. H., Jr. & Shaw, R. E. (eds.) (1985) *Persistence and change: Proceedings of the First International Conference on Event Perception*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Wightman, A. S. (1985) Regular and chaotic motions in dynamical systems: Introduction to the problems. In: G. Velo & A. S. Wightman (eds.), *Regular and chaotic motions in dynamic systems*. New York: Plenum.

Yates, F. E. (1978) Complexity and the limits to knowledge, *American Journal of Physiology*, **235**, R201–R204.

Zimmermann, G. N., Kelso, J. A. S. & Lander, L. (1980) Articulatory behavior pre and post full-mouth tooth extraction and alveoloplasty: a cinefluorographic study, *Journal of Speech and Hearing Research*, **23**, 630–645.

## Appendix

In modeling a discrete bilabial closure (unperturbed or perturbed), the damped second-order dynamics specified in the task space give rise to an evolving pattern of state-dependent "forces" exerted on an effector-independent constriction along the axes of constriction location ($t_1$) and constriction degree ($t_2$). The equations of motion for this task-dynamic regime are described in matrix notation as follows (note that in these and following equations, a superscript T denotes the vector or matrix *transpose* operation):

$$M_T\ddot{\mathbf{t}} + B_T\dot{\mathbf{t}} + K_T\mathbf{t} = \mathbf{0}, \tag{1}$$

where

$$\mathbf{t} = (t_1, t_2)^T;$$

$$\dot{\mathbf{t}}, \ddot{\mathbf{t}} = \text{the first and second derivatives of } \mathbf{t} \text{ with respect to time;}$$

$$M_T = \begin{bmatrix} m_{T_1} & 0 \\ 0 & m_{T_2} \end{bmatrix} \quad B_T = \begin{bmatrix} b_{T_1} & 0 \\ 0 & b_{T_2} \end{bmatrix} \quad K_T = \begin{bmatrix} k_{T_1} & 0 \\ 0 & k_{T_2} \end{bmatrix};$$

$$m_{T_1}, m_{T_1} = \text{inertial coefficients;}$$

$$b_{T_1}, b_{T_2} = \text{damping coefficients; and}$$

$$k_{T_1}, k_{T_2} = \text{stiffness coefficients.}$$

Equation (1) describes a linear, uncoupled set of task-space equations, whose terms are defined in units of force, and whose dynamic parameters (i.e., $M_T$, $B_T$, $K_T$) are constant.

Secondly, the task-space constriction is identified with the effector-specific bilabial constriction, and the task-space dynamic system is transformed kinematically into a two-dimensional body-space system governing movement of the bilabial constriction along the tract variables of lip protrusion ($z_1$) and lip aperture ($z_2$). Both lip protrusion and lip aperture are defined as functions of the current locations in jaw-space coordinates ($x_1$, $x_2$; see Figure 1(b) in KST) of the upper and lower lips and teeth. The resulting set of linear tract-variable equations of motion for the bilabial constriction may be rearranged and defined in matrix form as follows:

$$M_B\ddot{\mathbf{z}} \;=\; -B_B\dot{\mathbf{z}} - K_B\Delta\mathbf{z}, \tag{2}$$

where

$M_B \;=\; M_T$, where $M_T$ = task-space inertial matrix;
$B_B \;=\; B_T$, where $B_T$ = task-space damping matrix;
$K_B \;=\; K_T$, where $K_T$ = task-space stiffness matrix; and
$\Delta\mathbf{z} \;=\; \mathbf{z} - \mathbf{z}_0$, where $\mathbf{z} = (z_1, z_2)^T$, the current position vector of the bilabial constriction, and $\mathbf{z}_0 = (z_{01}, z_{02})^T$, the target or rest position vector for the bilabial constriction.

As with the task-space equations, the terms of (2) are defined in force units and the resultant set of tract-variable dynamic parameters ($M_B$, $B_B$, $K_B$, and $\mathbf{z}_0$) is constant. In (2), the right-hand side represents the tract-variable total force vector ($\mathbf{F}$); the first and second terms on the right-hand side represent the force vector components due to system damping ($\mathbf{F}_d$) and stiffness ($\mathbf{F}_s$), respectively.

Thirdly, the tract-variable dynamic system is transformed into a four-dimensional "model" articulator space where the moving segments have length but are massless (see Figure 2(c) in KST). Like the transformation from task space to body space, this transformation is a strictly kinematic one (since the segments have no mass) and involves only the substitution of variables defined in one coordinate system for variables defined in another coordinate system. As illustrated in Figure 2(c) in KST, this corresponds to expressing tract variables ($\mathbf{z}$, $\dot{\mathbf{z}}$, $\ddot{\mathbf{z}}$) as functions of model articulator variables ($\boldsymbol{\phi}$, $\dot{\boldsymbol{\phi}}$, $\ddot{\boldsymbol{\phi}}$; where $\boldsymbol{\phi} = [\phi_1, \phi_2, \phi_3, \phi_4]^T$, and $\phi_1$ = lip protrusion (LX1), $\phi_2$ = jaw angle (J), $\phi_3$ = upper lip vertical (ULX2), and $\phi_4$ = lower lip vertical (LLX2). For scaling purposes in the simulation program, commensurability between linear and angular units of measurement was obtained using radians for angular units and decimeters (1 dm = 100 mm) for linear units. The tract variables of equation (2) are transformed into the model-articulator variables using the following kinematic relationships:

$$\mathbf{z} \;=\; \mathbf{z}(\boldsymbol{\phi}) \tag{3a}$$

$$\dot{\mathbf{z}} \;=\; J(\boldsymbol{\phi})\dot{\boldsymbol{\phi}} \tag{3b}$$

$$\ddot{\mathbf{z}} \;=\; J(\boldsymbol{\phi})\ddot{\boldsymbol{\phi}} + \dot{J}(\boldsymbol{\phi}, \dot{\boldsymbol{\phi}})\dot{\boldsymbol{\phi}}, \tag{3c}$$

where

$\mathbf{z}(\boldsymbol{\phi}) \;=\; (z_1(\boldsymbol{\phi}), z_2(\boldsymbol{\phi}))^T$, the current tract-variable position vector expressed as a function of the current model articulator configuration. These functions are specific to the particular geometry assumed for the set of model articulators used to simulate speech gestures or produce speech acoustics via articulatory synthesis. In our simulations, the $\mathbf{z}(\boldsymbol{\phi})$ functions are defined according to the geometry of the simplified articulatory degrees of freedom in the Haskins Laboratories' articulatory speech synthesizer (Rubin, Baer & Mermelstein, 1981).

$J(\phi)$ = the *Jacobian* transformation matrix whose elements $J_{ij}$ are partial derivatives, $\partial z_i/\partial \phi_j$, evaluated at the current $\phi$; and

$\dot{J}(\phi, \dot{\phi})$ = $(\mathrm{d}J(\phi)/\mathrm{d}t)$, a matrix resulting from differentiating the elements of $J(\phi)$ with respect to time. The elements of $\dot{J}$ are functions of both the current $\phi$ and $\dot{\phi}$.

The elements of $J$ and $\dot{J}$ thus reflect the geometrical relationships among motions of the (simplified) model speech articulators (four degrees of freedom) and motions of the corresponding tract variables (two degrees of freedom). Using the kinematic relationships in equation (3), the model effector system's equation of motion is as follows:

$$M_B J\ddot{\phi} = -B_B J\dot{\phi} - K_B \Delta \mathbf{z}(\phi) - M_B \dot{J}\dot{\phi}, \tag{4}$$

where

$M_B$, $B_B$, $K_B$ are the same constant matrices used in Equation (2); and

$\Delta \mathbf{z}(\phi)$ = $\mathbf{z}(\phi) - \mathbf{z}_0$, where $\mathbf{z}_0$ = the same constant vector used in equation (2). It should be noted that since $\Delta \mathbf{z}$ in equations (2) and (4) is *not* assumed to be "small", a differential approximation $\mathrm{d}\mathbf{z} = J(\phi)\mathrm{d}\phi$ is not justified and, therefore, equation (3a) was used instead for the kinematic displacement transformation into model articulator variables.

The terms of (4) are still defined in units of linear force, however. They may be rewritten in units of linear ($\ddot{\phi}_1$, $\ddot{\phi}_3$, $\ddot{\phi}_4$) and angular ($\ddot{\phi}_2$) acceleration (note that the angular acceleration terms in the jaw's motion equation may be multiplied by a unit scaling factor, however, to ensure dimensional homogeneity (linear units) along all articulatory degrees of freedom):

$$\ddot{\phi} = -J^* M_B^{-1} B_B J\dot{\phi} - J^* M_B^{-1} K_B \Delta \mathbf{z}(\phi) - J^* \dot{J}\dot{\phi}. \tag{5}$$

In equation (5), the first and second terms on the right-hand side represent the model articulator acceleration vector components due to system damping ($\ddot{\phi}_d$) and stiffness ($\ddot{\phi}_s$), respectively. The third term on the right-hand side represents an acceleration component vector ($\ddot{\phi}_{vp}$) that is non-linearly proportional to the squares and pairwise products of current articulatory velocity (e.g. $\dot{\phi}_2^2$, $\dot{\phi}_2\dot{\phi}_3$, etc.; see Saltzman, in press, and Saltzman & Kelso, in press, for further details).

Additionally, in (5) $J^*$ is a weighted Jacobian pseudo inverse (e.g. Benati, Gaglio, Morasso, Tagliasco & Zaccaria, 1980; Klein & Huang, 1983; Whitney, 1972). It is used because there are a great number of model articulator variables than tract variables for this task. More specifically, $J^* = A^{-1}J^T(JA^{-1}J^T)^{-1}$, where $A$ is a positive definite *articulatory weighting* matrix, the elements of which are constant during a given gesture. Using $J^*$ provides a unique, optimal least squares solution for the differential transformation from tract variables to model articulator variables that is weighted according to the pattern of elements in the $A$ matrix. $A$'s elements reflect task-specific constraints on the relative motions of the articulators during a given gesture. In current modeling, the $A$ matrix is defined to be of diagonal form (i.e. where element $a_{ii}$ is associated with articulator $\phi_i$), and a given set of articulator weights will constrain motion of a given model articulator in direct proportion to the magnitude of the corresponding weighting element relative to the remaining weighting elements.