SPEECH MOTOR CONTROL AND STUTTERING
Peters, Hulstijn and Starkweather, editors.

808



Chapter 3

# THE TASK DYNAMIC MODEL IN SPEECH PRODUCTION

*Elliot Saltzman*

The communicative act of speaking entails the precise, spatiotemporal structuring of activity in the articulatory and phonatory apparatus, a multi-degree-of freedom dynamical system whose components must be coordinated over time in order to appropriately structure sound for a listener. This paper focusses on the gestural control schemes that underlie the spatiotemporal patterning of speech, within the framework of the task-dynamic model of speech production. This model represents an attempt to reconcile the linguistic hypothesis that speech involves an underlying sequencing of abstract, context-independent units, with the empirical observation of context-dependent interleaving of articulatory movements. Currently, the task-dynamic model provides an intrinsically dynamical account of interarticulator coordinative processes during unperturbed and mechanically perturbed gestures, as well as during intervals of gestural coproduction. Work is in progress to incorporate a serial dynamics of intergestural coordination based on the dynamics of recurrent connectionist networks. Such dynamics will be used to intrinsically shape patterns of relative timing among the gestures themselves, in a hybrid (task + serial) dynamical model of speech production. Finally, implications of the present theoretical framwork for understanding certain aspects of stuttering behavior are discussed.

A common intuition concerning skilled activities is that there is an underlying invariance of control despite observed surface variations in performance. This intuition applies equally to ordinary activities, such as walking, writing, and talking, and to extraordinary activities, such as dancing, sculpting, and singing. Although a given type of action is never performed the same way twice, one is convinced that there is something invariant that underlies the organization of each varied performance. In this chapter, an approach to these issues is discussed whose aim is to capture in a single framework both the variable and invariant aspects of skilled actions. The framework has been called a task-dynamic one for very simple reasons. The term task is used since the approach was designed to account for performances of tasks in the real world; the term dynamic is used since the goal is to provide a dynamical account of the forces that give rise to a movement's observable kinematic patterns, rather than simply provide a kinematic redescription of the performance in terms of, for example, a motor template. A dynamical account is preferable, since it provides a unified account not only of a movement's form, but also of the stability of the form, and of the lawful warping of the form that

occurs with changes along performance dimensions such as rate and emphasis.

The specific focus of this chapter is on the control of the speech articulators. There are three main topics: a) interarticulator coordination within the time span of single speech gestures, e.g., the coordination between lips and jaw during bilabial closure; b) intergestural blending when several gestures overlap in time (co-production) and the gestures share articulators in common, e.g., the blending in a /VbV/ sequence in which the period of bilabial control overlaps those of the flanking vowels, and vocalic and consonantal influences are blended in controlling the shared jaw; and c) intergestural patterns of relative timing or phasing, e.g., the relative timing between bilabial closing-opening and laryngeal abduction-adduction for /p/, or between succesive bilabial gestures in a /bVb/ sequence. Finally, implications of the present dynamical framework for understanding the origins of stuttering will be presented in an admittedly speculative manner.

## INTERARTICULATOR COORDINATION

One important phenomenon displayed during skilled actions is motor equivalence (Hebb, 1949; Lashley, 1930), whereby the action system spontaneously reorganizes itself en route to the task goal when faced with a disruption or perturbation. If one route is blocked, the system finds an alternative route to the goal. Examples of this sort of behavior in speech production have been provided by experiments in which mechanical pertubations are unexpectedly delivered to the lower lip or jaw during ongoing speech gestures (Abbs & Gracco, 1983; Folkins & Abbs, 1975; Kelso, Tuller, Vatikiotis-Bateson & Fowler, 1984; Munhall & Kelso, 1985; Munhall, Löfqvist & Kelso, 1986; Shaiman, 1989). In one representative study (Kelso et al., 1984), movements of the upper lip, lower lip, and jaw were measured optoelectronically, and tongue raising was assessed by electromyographic monitoring of genioglossus activity. Subjects were required to repeat either the phrases "It's a /bæb/ again" or "It's a /bæz/ again". On a low percentage of trials, a downward torque perturbation was delivered to the jaw during the jaw raising gesture associated with either the second /b/ in /bæb/ or the /z/ in /bæz/. It was found that for the /bæb/ trials, in which lip (but not tongue) action is crucial in creating the final /b/ closure, there was extra compensatory lowering of the upper lip but normal activity of the tongue, relative to the nonperturbed control trials. Conversely, for the /bæz/ trials, in which tongue (but not lip) action is crucial in forming the alveolar constriction, there was extra genioglossus activity but normal activity of the upper lip, relative to the controls. Furthermore, the corrective actions were rapid, in that the delay from onset of the perturbation to onset of the anatomically remote, compensatory responses were approximately 20-30 ms.

The speed and specificity of remote compensation in such experiments support several conclusions regarding processes of interarticulator coordation and control. The speed of response implies that there is an automatic reflex-like organization involved (with a 20-30 ms loop time), but the specificity of response implies that

such an input-output mapping cannot be hard- wired. Rather, there must exist a task- or gesturally- specific, selective pattern of gating or coupling among the component articulators that is specific to the gesture being produced.

What kind of dynamics could give rise to these remote compensatory patterns? One possibilty, following from dynamical mass-spring accounts of single degree-of-freedom limb positioning tasks (e.g. Kelso, 1977; Polit & Bizzi, 1978; Cooke, 1980; Schmidt & McGown, 1980), is that each speech articulator has its own gesture-specific target or equilibrium position (e.g. separate target positions for the lips and jaw, as in Lindblom, 1967). Such an account is inadequate, however, since according to this scenario the unperturbed articulators would attain their individual targets, but the perturbed articulator would not. This would result in the higher-order constriction goal (e.g. bilabial closure) not being met. In the Kelso, et al. (1984) experiment, the constriction goals were attained in the perturbed trials, but with different articulator positions than those observed in the unperturbed control trials. The task-dynamic model of gestural control provides for such specificity and flexibility of articulatory behavior. This work has been the result of collaboration with a number of colleagues over the past several years (Browman & Goldstein, 1986, in press; Browman, Goldstein, Kelso, Rubin & Saltzman, 1984; Browman, Goldstein, Saltzman & Smith, 1986; Kelso, Saltzman & Tuller, 1986a, 1986b; Kelso, Vatikiotis-Bateson, Saltzman & Kay, 1985; Saltzman, 1986; Saltzman, Goldstein, Browman & Rubin, 1988a, 1988b; Saltzman & Kelso, 1987; Saltzman & Munhall, 1989; Saltzman, Rubin, Goldstein & Browman, 1987; for recent reviews and critiques, see also Hawkins, in press; and Jordan & Rosenbaum, 1989). We begin by considering the coordinative processes involved in task- dynamic simulations of a single, temporally isolated bilabial closing gesture.

One of the major tasks of speech production is the creation and release of constrictions of differing degree in different regions of the vocal tract. In simulating isolated gestures, invariant control structures are defined in the task-dynamic model that give rise to constriction-specific and contextually variable patterns of articulator movements. This is accomplished in two basic steps. The first is to define an invariant dynamical control regime at an abstract, task-specific level of system description; and the second step is to use these invariant dynamics to generate the contextual variation observed for articulatory trajectories. The first step entails defining time-invariant dynamics at the level of tract-variable coordinates. For bilabial gestures, the two tract variables used are lip aperture and lip protrusion[1]. Lip aperture defines the degree of bilabial constriction, and is defined by the vertical distance between the upper and lower lips; lip protrusion defines the location of bilabial constriction, and is defined by the horizontal distance between the (yoked) upper and lower lips and the upper and lower front teeth, respectively. Gestural dynamics are assumed in the model to be those of a damped, second-order dynamical system, analogous to a damped mass- spring. Such point attractor dynamics result in gestural primitives that are defined as discrete gestures, e.g., bilabial closing, laryngeal abduction, velic lowering, etc. The tract-variable equation of motion is time-invariant, since its parameters, such as target

position, stiffness,and damping, are constant over the course of a single isolated gesture.

However, a tract-variable description does not contain any of the articulatory details required for an adequate simulation of speech production. Thus, the second step is to transform the tract-variable dynamical system in an explicitly articulatory set of coordinates. For bilabial gestures, the set of model articulators are jaw angle, and the (independent) vertical positions and (yoked) horizontal positions of the upper and lower lips relative to the upper and lower teeth, respectively. These model articulators are defined in strictly kinematic terms, however, in that they have lengths but no masses. Thus, the coordinate tranformation from tract-variables to model articulators is a strictly kinematic one. Since there are more articulators than tract-variables, this transformation is also indeterminate (one-to-many), and the effector system is considered technically to be redundant (the issue of redundancy in the speech production system is also addressed in the work presented by Neilson & Neilson, 1991). The model-articulator dynamical system that results from this coordinate transformation is not time-invariant, however, in that its parameters are no longer constant, even over the time course of a single isolated gesture. Rather, the parameter values change as a function of both the ongoing posture of the articulators during the gesture and the set of constant parameters defined at the tract-variable level. Significantly, the dynamical system at the articulatory level does not simply generate independent time functions that force or drive each articulator in an independent manner. Rather, these dynamics define a gesture-specific and posture-specific set of coupling or gating functions among the articulators, that "create" an articulatory synergy or coordinative structure. Thus, these functions define a task-specific cooperativity among the articulators that allow them to flexibly and adaptively attain speech-relevant goals, and captures the essence of what is meant by "coordination" (see also Turvey, 1990).

These results are exactly the type required to formalize our previously described intuitions concerning invariance and variability in skilled activity. That is, in the task-dynamic model, invariance exists at the level of abstract, tract-variable dynamics, and this invariance is used in a rigorous way to generate the context-sensitive, kinematic variability observed at the articulatory level. In addition, the model operates in an autonomous manner, in that unperturbed movement trajectories unfold as implicit consequences of the system's dynamics, rather than as explicit consequences of kinematically specified trajectory plans or templates. Further, remote compensation to simulated perturbation is immediate, in the sense that compensation occurs without explicit error-detection, replanning, or reparameterizing procedures (Saltzman, 1986; Kelso et al., 1986a, 1986b).

## MULTIPLE GESTURES: CO-PRODUCTION AND INTERGESTURAL BLENDING

Speech utterances contain more than one type of gesture. Consequently, the inventory of gestures was expanded to include the sets of tract-variables and model articulators illustrated in Figure 1.

| Tract variables | | Model articulators |
|---|---|---|
| **LP** | lip protrusion | upper & lower lips |
| **LA** | lip aperture | upper & lower lips, jaw |
| **TDCL** | tongue dorsum constrict location | tongue body, jaw |
| **TDCD** | tongue dorsum constrict degree | tongue body, jaw |
| **LTH** | lower tooth height | jaw |
| **TTCL** | tongue tip constrict location | tongue tip,body, jaw |
| **TTCD** | tongue tip constrict degree | tongue tip,body, jaw |
| **VEL** | velic aperture | velum |
| **GLO** | glottal aperture | glottis |

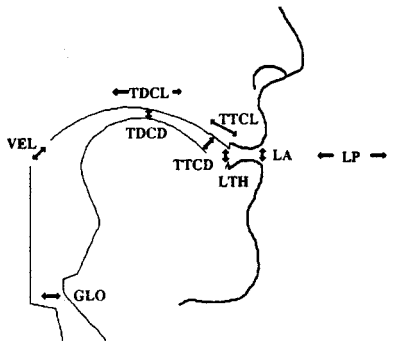


Figure 1. Top: Table showing the relationship between tract-variables and model articulators; Bottom: Schematic midsagittal vocal tract volume, with tractvariable degrees of freedom indicated by arrows.

With this expanded gestural repertoire, the simulation of more realistic utterances could begin in earnest. In particular, in order for the simulations to be realistic, they must capture the phenomenon of co-production (e.g. Bell-Berti & Harris, 1981; Fowler, 1977, 1980; Harris, 1984; Keating, 1985; Kent & Minifie, 1977; Ohman, 1966, 1967; Perkell, 1969; Sussman, MacNeilage & Hanson, 1973). Co-production refers to the articulatory and/or acoustic consequences of temporal overlap in gestural activity associated with nearby phonemes or segments in a given utterance. During periods of co-production, the influences of the temporally overlapping gestures on articulatory movements are blended. For example, blending influences have been observed in the production of VCV sequences by several researchers (e.g. Ohman, 1966, 1967; Sussman, et al., 1973). In such cases, the articulator movement associated with forming the constriction for the medial consonant is influenced by the identities of the flanking vowels. The extent of blending depends upon the degree of spatial overlap existing among the consonantal and vocalic gestures, i.e., the extent to which these gestures share articulators in common.

In current task-dynamic simulations, it is assumed that vowels are produced using the tract-variables of tongue-dorsum constriction location and degree, and the articulator set of jaw angle and tongue body motion relative to the jaw. Thus, a continuum of supralaryngeal spatial overlap can be identified, dependent upon

the identity of the medial consonant. If the consonant is /h/, no overlap occurs, and there is no blending. If the consonant is a bilabial, with an articulator set consisting of the jaw and lips, then spatial overlap occurs at the shared jaw. If the consonant is an alveolar, with an articulator set defined by the tongue tip, tongue body, and jaw, then spatial overlap occurs at the shared tongue body and jaw, and there is still flexibility left in the system since the overlap is not total. Ohman (1967) showed that, in such cases, both the degree and location of the tongue-tip constriction are not affected by the identity of the flanking vowels, although the position of the tongue body is affected in a vowel-specific manner. Finally, if the medial consonant is a velar, with an articulator set defined by the tongue body and jaw, the spatial overlap is total, and the system loses flexibility in its goal-seeking behavior. Thus, in this instance, Ohman showed that the degree of tongue-dorsum constriction was not affected by the flanking vowels' identity, but the location of the constriction was shifted according to the vowel context.

In order to explain how the task-dynamic model replicates such findings, the means by which the model generates gestural sequences will first be described in more detail. The main additional construct to be considered is the set of gestural activation coordinates. Each gesture is associated with its own activation variable, and the set of activation variables defines a third coordinate system of the model. (Recall that tract-variables and model articulator variables comprise the first two coordinate systems). Figure 2 illustrates the "anatomical" relationships that exist

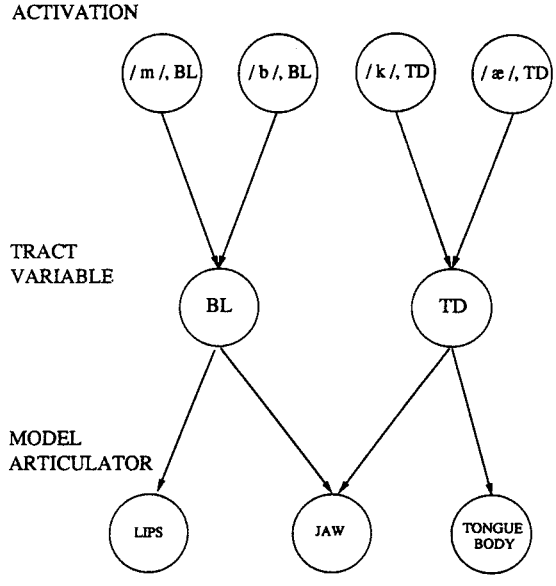


Figure 2. Example of the "anatomical" relationships defined among model-articulator, tract-variable and activation coordinate systems. BL and TD denote tract-variables associated with bilabial and tongue-dorsum constrictions, respectively. Gestures at the activation level are labeled in terms of both linguistic identity (e.g., /k/) and tract-variable affiliation (e.g., TD).

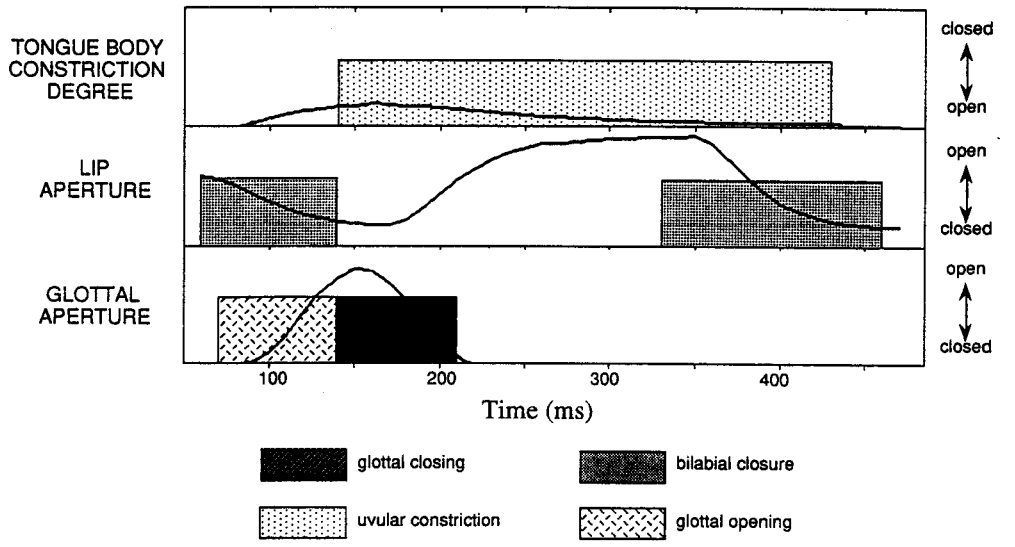


Figure 3. Gestural score for the simulated sequence /pʌb/. Filled boxes denote intervals of gestural activation. Box heights are either 0 (no activation) or 1 (full activation). The waveform lines denote tract-variable trajectories produced during the simulation.

among all three coordinate systems. Gestural activation refers to the strength with which a gesture "attempts" to shape vocal tract movement at a given point in time.

Currently, periods of gestural activation are represented as step functions, normalized from zero (the gesture is "off" or inactive) to one (the gesture is "on" or maximally active). Relative timing of activation intervals for a given utterance is specified with reference to a corresponding gestural score. The gestural score for the utterance "pʌb" is illustrated in Figure 3.

Currently, the relative timing or phasing among an utterance's gestural activation intervals is not shaped according to an implicit, underlying dynamics. Rather, each utterance's gestural score is specified either "by hand", or according to the linguistic gestural model of Browman and Goldstein (1986; in press). This model embodies the rules of their articulatory phonology, and incorporates knowledge of contrastive gestures in English and how these gestures cohere in larger units.

Using these methods, the task-dynamic model produces the type of co-production and blending results described above (see Figure 4). Figure 4c shows the simulated sagittal outlines of the vocal tract for steady-state productions of the vowels /i/ and /æ/. Figure 4a shows the simulation frames containing the tongue-tip's first contact with the upper tract wall for the symmetric VCV sequences /idi/ and /ædæ/. As occurs with actual data, the simulations show that both constriction location and degree for the alveolar are attained identically across differing vowel height contexts. These results are due to the manner in which the model handles temporal overlap of gestures defined along different tract variables. In these instances, the corresponding articulator sets share some, but not all, articulators in common. Using the same relative timing patterns as for the alveolar,

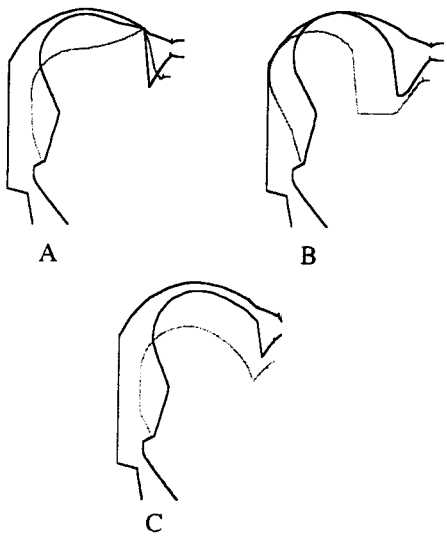


Figure 4. Simulated vocal tract shapes. A. First contact of tongue-tip and upper tract wall during symmetric vowel-alveolar-vowel sequences; B. First contact of tongue-dorsum and upper tract wall during symmetric vowel-velar-vowel sequences; C. Corresponding steady-state vowel productions.

simulations of the sequences /igi/ and /ægæ/ (see Figure 4b) also mirror actual data, in that the degree of tongue-dorsum constriction for the velar is identical across vowel contexts, even though the constriction location is altered by the flanking vowels. These results are due to the manner in which the model handles temporal overlap of gestures defined along the same tract-variables where, by definition, there is total spatial overlap of the articulator sets involved (see Saltzman & Munhall, 1989; for further details of these coproduction and blending processes).

Evidence consistent with such processes of within-tract- variable blending have been reported by Laver (1980), in an experiment on laboratory-induced errors in vowel production. In this experiment, blended vowel forms intermediate between canonical forms were produced, supporting Laver's hypothesis that speakers were "indecisive" at a level of neuromuscular pattern selection or activation, and consequently issued commands appropriate for the simultaneous, blended production of different vowels.

## INTERGESTURAL TIMING: SERIAL DYNAMICS

In the current task-dynamic model, the spatiotemporal patterns of speech may be viewed as emergent behaviors that are implicit in a dynamical system with two functionally distinct but interacting levels (see Figure 5).

The interarticulator level is defined according to both tract-variable (e.g. lip aperture and protrusion) and model articulator (e.g. lips and jaw) coordinates; the intergestural level is defined according to the set of activation coordinates. The tract-variable and model articulator coordinates of each gesture specify,

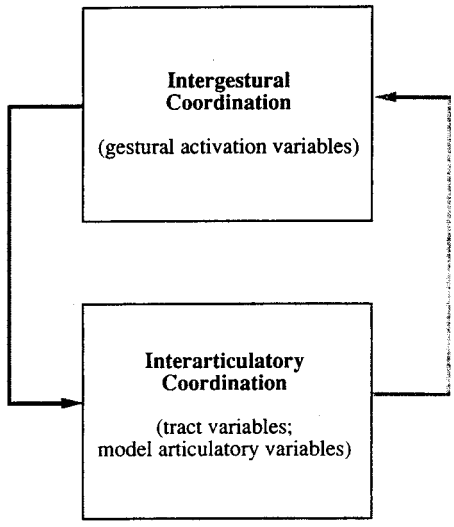


Figure 5. Schematic illustration of the two-level dynamical model for speech production, with associated coordinate systems indicated. The darker arrow from the intergestural to the interarticulator level denotes the feedforward flow of gestural activation. The lighter arrow indicates feedback of ongoing tract-variable and model articulator state information to the intergestural level.

respectively, the particular vocal-tract constriction (e.g. bilabial) and articulatory components whose behaviors are affected directly by the gesture's activation; each gesture's activation coordinate reflects the ongoing strength of that gesture's control in shaping the vocal tract. Invariant gestural units are posited in the form of context-independent sets of dynamical parameters (e.g. lip protrusion target, stiffness, and damping coefficients), and are associated with corresponding subsets of the model's coordinates. Each gestural unit's influence over the vocal tract waxes and wanes according to the activations of the units. Variability emerges in the unfolding of articulatory movements as a result of both the utterance-specific interleaving of gestures and the accompanying processes of coproduction and blending. Significantly, the model functions in exactly the same way during simulations of unperturbed, mechanically perturbed, and coproduced speech gestures.

As discussed above (in the section entitled "Multiple gestures: co-production and intergestural blending"), however, patterns of intergestural relative timing are currently specified explicitly according to rule-generated or manually generated gestural scores. Task-dynamics currently contains no intrinsic dynamics for the timing or phasing of gestural activation intervals that is comparable to the dynamics employed for interarticulator coordination. What is needed, therefore, is a dynamic of intergestural coordination that will implicitly shape or "grow" the gestural scores in an autonomous manner. Work is currently in progress to incorporate such a serial dynamic of intergestural phasing based on the dynamics of recurrent connectionist networks. In particular, the sequential network of Jordan (1986; 1990; in press) will be used to intrinsically shape patterns of relative timing

among the gestural activation intervals themselves, in a hybrid (task + serial) dynamical model of speech production. (see Saltzman & Munhall, 1989; for further details on serial dynamics, and on the bidirectionality of coupling required between the intergestural and interarticulator levels).

In the hybrid network, each member of the sequential network's output nodes corresponds to a given gestural activation coordinate. The values of the output nodes range from zero (the corresponding gesture is inactive) to one (the gesture is fully activated). However, these values vary continuously within this range, thereby generalizing the possible waveshapes for gestural activation beyond the step functions used in the current model. The functioning of these gestural output nodes is consistent with recent electromyographic data reported by Mowrey and MacKay (in press) on sublexical speech errors elicited during the production of tongue twisters. For example, during intended productions of "Bob flew by Bligh Bay", EMG output from the lingual tranversus/verticalis complex (T/V) was monitored. In trials that were normal both acoustically and electromyographically, T/V activity occurred in two distinct bursts accompanying the production of the lateral /l/s in "flew" and "Bligh". For trials that were anomalous both acoustically and electromyographically, error patterns were of two types: a) intrusion errors, that consisted of increased T/V activity during the intended production of "Bay" (the error sounded like "Blay"); and b) exchange errors, that consisted of both a decrease of T/V activity for "Bligh", and T/V intrusion for "Bay" and sometimes "by" (the errors sounded like "Bigh", "Blay", and "bly"). Significantly, T/V errors were not constrained to be all or none. Rather, they were graded continuously between levels appropriate for normal and fully anomalous tokens. (Some of these graded T/V error patterns were auditorily undetectable, even by trained phoneticians). This gradation of electromyographic activity is clearly consistent with the continuously graded activation levels proposed for the set of gestural output units in the hybrid dynamical model.

The sequential network proposed for shaping patterns of intergestural relative timing is also, in fact, virtually identical in architecture and functioning to the network used by Gary Dell (1991) to simulate speech errors. This is no coincidence, since both are variants of the sequential net designed by Jordan (1986; 1990; in press). The output nodes of Dell's model correspond to activations of individual phonemes. The network produces speech errors with the same distribution of phonological characteristics that is seen in actual data. In human data, these error patterns have been taken as evidence of an explicit, rule-based distinction between an utterance's phonological frame (e.g., onset-rhyme structure) and its phonological content. However, Dell's network reproduces these human error patterns without encoding or training the frame- content structure explicitly into the net. Rather, the human-like error patterns emerge as implicit functions of the network's dynamics and of the statistical regularities in the language inputs used in training the net to "speak". Rule-like structural effects are produced without reference to explicitly incorporated rules.

Hearing Dell's presentation gave me the feeling of working with a group of

friends on a large jigsaw puzzle. Typically, as time progresses several large subsections are constructed independently, until at some point it becomes clear that the subsections can be joined together. The final outputs of Dell's model are phonemic activations; the initial inputs of the task- dynamic model are gestural activations. Although a significant gap still exists in a dynamical framework between phonemic and gestural representations (see Saltzman & Munhall, 1989;, for a discussion of this problem), it seems clear at this point that these separate models can be joined together, offering the exciting possibility of a seamless union of "higher level" linguistic/ phonological and "lower level" sensorimotor/phonetic processes.

## CONCLUSION: A DYNAMICAL METAPHOR FOR STUTTERING

The task-dynamic model in its present form cannot stutter. However, several of the participants at this conference presented dynamical models that could display stuttering-like behaviors: the modified Hopfield network of Braamhof, Wieneke, Coolen and Janssen; the sequential network of Dell; the neuronal network of McClean; the control theoretic models of Neilson and Neilson; and of Nudelman, Herbrich, Hoyt and Rosenfield; and the audio-phonatory coupling model of Kalveram; and of Jäncke. How is one to interpret these simulations? One hypothesis, consistent with that of Nudelman, et al. (1991), is to consider stuttering to be a bifurcation phenomenon, a qualitatively different dynamical pattern produced by the speech production system when the system crosses over into a particular region of its space of control parameters. Stuttering is the "normal" behavior for the system in this region, although its effects on the speech patterns intended by the speaker are clearly disruptive and potentially devastating. In this admittedly metaphorical framework, some questions of obvious concern for speech scientists and clinicians alike are: Along what dimensions is the parameter space defined? What factors govern transitions into the stuttering region? What factors enhance the frequency with which such transitions are made? Finally, how can these transitions be minimized or eliminated?

Bifurcation phenomena are common in the nonlinear dynamics literature, and are illustrated by the example of Rayleigh-Benard convection. Rayleigh-Benard convection (e.g. Baker & Gollub, 1990; Thompson & Stewart, 1986) occurs when a thin layer of fluid such as silicone oil is placed in a container between two thermally conducting, horizontal plates. Temperature gradients are created across the fluid by heating the bottom plate and cooling the top plate. When the temperature difference between the two plates is small, heat flows from bottom to top by conduction, and the fluid is still. However, when the temperature difference reaches a critical value, convection occurs and the fluid begins to circulate. The motion takes the form of a parallel series of rotating, sausage-like rolls. Beyond a second critical temperature value, the rolls or "convection cells" begin to wobble as well as rotate; and the flow patterns become chaotic or turbulent at even higher critical values. In this example, the control parameter is the applied temperature

difference, and the critical values mark the points of system bifurcation as well as partition the parameter space into regions associated with qualitatively distinct patterns of fluid flow. Like many other movement forms that appear in flowing media, such as vortices, eddies, and whirlpools in a stream (where the driving gradient is typically gravitational rather than thermal), the Rayleigh-Benard forms disappear when the gradient disappears. The movement forms only exist during appropriately constrained flows of the system.

This example is instructive when stuttering is viewed from a dynamical perspective. If the transition to stuttering is indeed analogous to a bifurcation phenomenon, then the way to avoid acute instances of stuttering is to avoid crossing over into the stuttering region of the relevant parameter space. This may be difficult for speakers who are predisposed to more frequently visit this region. However, it is likely that the situation is even more complex. In the concluding talk of the conference, David Prins (1991) asked why and how stuttering, initially a relatively infrequent phenomenon, becomes a chronic behavior. He hypothesized that speakers' own "corrective reactions" to acute stuttering events contribute to the retention of the stuttering pattern. Somehow, the stutterer's act of attentional or voluntary intervention serves to facilitate future transitions into the stuttering region. The system "learns" to return to or stay in this region more easily.

How can one formulate such predispositions and learning processes within a dynamical framework? A reasonable hypothesis is that these processes are linked to the dynamics that shape movements of the control parameters themselves. In the Rayleigh- Benard example discussed earlier, motion of the control parameter was determined explicitly by experimental manipulation. For human speakers, the origins of the driving influences that shape motion patterns in the parameter space are far less clear. However, one can think of these motions as shaped according to an underlying landscape whose surface is slightly sticky. The landscape consists of hills, valleys, plateaus, and so on, and the current value of the system's parameter set is represented as the location of a small ball on this surface. The ball moves by rolling downhill whenever it can.

Imagine, for the sake of simplicity, that the landscape has two regions, separated by some distance, that correspond to regions of fluent and stuttered speech, respectively. Where the system spends most of its time will depend of the topography of the landscape. For example, if the layout consists of a single basin whose bottom is located within the fluent region (Figure 6A), the system will be stably fluent most of the time. If there is noise in the system that perturbs the ball (parameter values) to move up the walls of the basin, the ball will still tend to slide down and return to the bottom. If the basin is centered within the stuttering region (Figure 6E), however, the speaker will unfortunately stutter in a stable manner most of the time. There are intermediate scenarios as well (Figure 6B - Figure 6D). In fact, one can hypothesize that the landscapes of speakers with predispositions to stutter are more similar in shape to that of Figure 6E than are the landscapes of speakers with no such predispositions. Assuming the existence of perturbing noise in the system, it is clear that it will be more difficult to stay within the fluent region

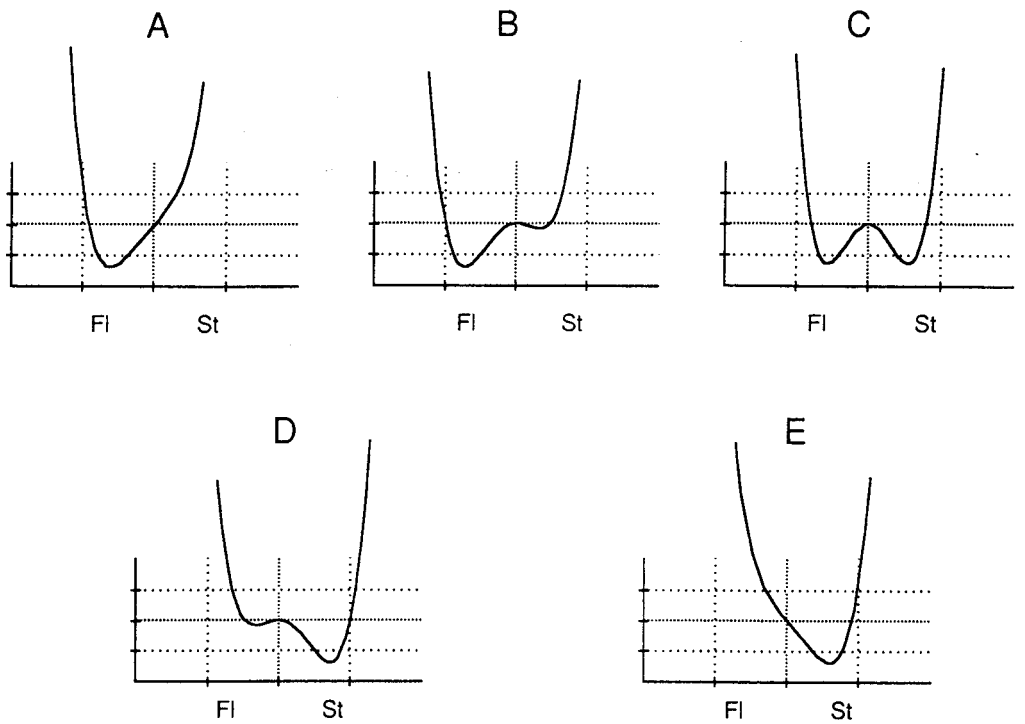


Figure 6. Schematic control landscapes for fluent and stuttered speech. Fluent (F1) en stuttering (St) regions of the hypothetical parameter space are indicated on the horizontal axis, and control "potential" is represented on the vertical axis. A-E illustrate samples along a continuum of landscape topographies. (See text for details).

the closer the landscape is to that of Figure 6E. In such instances, it is more likely that the system will lose its balance and fall into the stuttering region.

Returning to Prins' hypothesis, it is possible that if corrective or attentional processes are recruited by the onset of stuttering, these processes may inadvertently alter the shape of the landscape itself. The landscape will be affected in a plastic manner, such that frequent disruptions act to drive the shape closer to that of Figure 6E, and thereby enhance the system's predisposition to reside in the stuttering region of parameter space. Thus, in this final scenario, the main challenge is to develop a set of intervention techniques that can induce long-lasting counter-changes in the landscape of the parameter space, changes that move the landscape toward a form that is more conductive to fluent speech than stuttering.

## ACKNOWLEDGMENTS

Grant support is acknowledged from NIH Grant NS-13617 (Dynamics of Speech Articulation) and NSF Grant BNS-8520709 (Phonetic Structure Using Articulatory Dynamics) to Haskins Laboratories. I am also grateful to Katherine Harris, Joe Kalinowski, and Philip Rubin for critical reviews of earlier versions of this chapter.

## NOTES

[1] In all simulations to date, constrictions are defined in the sagittal plane of the vocal tract only, and thus are at most two-dimensional. The reason for this is that the simulations use the articulatory geometry represented in the Haskins Laboratories software articulatory synthesizer (Rubin, Baer & Mermelstein, 1981). This synthesizer transforms a given articulatory configuration in the sagittal plane to a sagittal outline of the vocal tract, a three dimensional tube shape, and finally, with the addition of appropriate voice source information, an acoustic waveform.

## REFERENCES

Abbs, J.H. & Gracco, V.L. (1983). Sensorimotor actions in the control of multimovement speech gestures. *Trends in Neurosciences, 6,* 393-395.

Baker, G.L. & Gollub, J.P. (1990). *Chaotic dynamics: An introduction.* New York: Cambridge University Press.

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

Braamhof, M., Wieneke, G., Coolen, T. & Janssen, P. (1991). Can neural networks explain dysfluent speech? In: H.F.M. Peters, W. Hulstijn and C.W. Starkweather (Eds.), *Speech motor control and stuttering.* Amsterdam: Elsevier Science Publishers.

Browman, C.P. & Goldstein, L. (1986). Towards an articulatory phonology. *Phology yearbook, 3,* 219-252.

Browman, C.P. & Goldstein, L. (in press). Tiers in articulatory phonology, with some plications for casual speech. In: J. Kingston and M.E. Beckman (Eds.), *Papers in laboratory phonology: I. between the grammar and the physics of speech.* Cambridge England: Cambridge University Press.

Browman, C.P., Goldstein, L., Kelso, J.A.S., Rubin, P. & Saltzman, E.L. (1984). Articulatory synthesis from underlying dynamics (Abstract). *Journal of the acoustical society of America, 75* (Suppl. 1), S22-S23.

Browman, C.P., Goldstein, L., Saltzman, E.L. & Smith, C. (1986). Gest: A computational model for speech production using dynamically defined articulatory gestures (Absctrat). *Journal of the acoustical society of America, 80* (Suppl.·1), S97.

Cooke, J.D. (1980). The organization of simple skilled movements. In: G.E. Stelmach and J. Requin (Eds.), *Tutorial in motor behavior.* Amsterdam: North Holland.

Dell, G.S. (1991). Connectionist approaches to the production of words. In: H.F.M. Peters, W. Hulstijn and C.W. Starkweather (Eds.), *Speech motor control and stuttering.* Amsterdam: Elsevier Science Publishers.

Folkins, J.W. & Abbs, J.H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of speech and hearing research, 18,* 207-220.

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

Fowler, C.A. (1980). Corarticulation and theories of extrinsic timing control. *Journal of Phonetics, 8,* 118-138.

Harris, K.S. (1984). Coarticulation as a component of articulatory descriptions. In: R.G. Daniloff (Ed.). *Articulation assessment and treatment issues* (pp. 147-167). San Diego, CA: College Hill Press.

Hawkins, S. (in press). An introduction to task dynamics. In: D.R. Ladd & G.J. Docherty (Eds.), *Proceedings of the second conference on laboratory phonology.* Cambridge: Cambridge University Press.

Hebb, D.O. (1949). *The organization of behavior.* New York: Wiley.

Jäncke, L. (1991). The "audio-phonatory coupling" in stuttering and nonstuttering adults: Experimental contributions. In: H.F.M. Peters, W. Hulstijn and C.W. Starkweather (Eds.), *Speech motor control and stuttering.* Amsterdam: Elsevier Science Publishers.

Jordan, M.I. (1986). *Serial order in behavior: A parallel distributed processing approach (Tech. rep. no. 8604)* San Diego: University of California, Institute for cognitive science.

Jordan, M.I. (1980). Motor learning and the degrees of freedom problem. In: M. Jeannerod (Ed.), *Attention and performance XIII.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Jordan, M. (in press). Serial order: A parallel distributed processing approach. In: J.L. Elman and D.E. Rumelhart (Eds.), *Advances in connectionist theory: Speech.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Jordan, M.I. & Rosenbaum, D.A. (1989). In: M.I. Posner (Ed.), *Foundations of cognitive science.* (pp. 727-767). Cambridge, MA: MIT Press.

Kalveram, K.T. (1991). How pathological audio-phonatory coupling induces stuttering: A model of speech flow control. In: H.F.M. Peters, W. Hulstijn and C.W. Starkweather (Eds.), *Speech motor control and stuttering.* Amsterdam: Elsevier Science Publishers.

Keating, P.A. (1985). CV phonology, experimental phonetics and coarticulation. *UCLA working papers in phonetics, 62,* 1-13.

Kelso, J.A.S. (1977). Motor control mechanisms underlying human movement reproduction. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 529-543.

Kelso, J.A.S., Saltzman, E.L. & Tuller, B. (1986a). The dynamical theory in speech production: Data and theory. *Journal of Phonetics, 14,* 29-60.

Kelso, J.A.S., Saltzman, E.L. & Tuller, B. (1986b). Intentional contents, communicative context, and task dynamics: A reply to the commentators. *Journal of Phonetics, 14,* 171-196.

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

Kelso, J.A.S., Vatikiotis-Bateson, E., Saltzman, E.L. & Kay, B.A. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America, 77,* 266-280.

Kent, R.D. & Minifie, F.D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics, 5,* 115-133.

Lashley, K.S. (1930). Basic neural mechanisms in behavior. *Psychological Review, 37,* 1-24.

Laver, J. (1980). Slips of the tongue as neuromuscular evidence for a model of speech production. In: H.W. Dechert and M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler.* The Hague: Mouton.

Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. *Transmission laboratory quarterly progress status report, STL-OPSR-4,* 1-29.

McClean, M.D. (1991). Simulation of input to motoneurons during speech dysfluency using a simple inhibitory neuronal network. In: H.F.M. Peters, W. Hulstijn and C.W. Starkweather (Eds.), *Speech motor control and stuttering.* Amsterdam: Elsevier Science Publishers.

Mowrey, R.A. & MacKay, I.R.A. (in press). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America.*

Munhall, K.G. & Kelso, J.A.S. (1985). Phase dependent sensitivity to perturbation reveals the nature of speech coordinative structures (Absctract). *Journal of the acoustical society of America, 78* (Suppl. 1), S38.

Munhall, K.G., Löfqvist, A. & Kelso, J.A.S. (1986). Laryngeal compensation following sudden oral perturbation (Abstract). *Journal of the acoustical society of America, 80* (Suppl. 1), S109.

Neilson, M.D. & Neilson, P.D. (1991). Adaptive model theory of speech motor control and stuttering. In: H.F.M. Peters, W. Hulstijn and C.W. Starkweather (Eds.), *Speech motor control and stuttering.* Amsterdam: Elsevier Science Publishers.

Nudelman, H.B., Herbrich, K.E., Hoyt, D.B. & Rosenfield, D.B. (1991). A neuroscience approach to stuttering. In: H.F.M. Peters, W. Hulstijn and C.W. Starkweather (Eds.), *Speech motor control and stuttering.* Amsterdam: Elsevier Science Publishers.

Ohman, S.E.G. (1966). Coarticulation in VCV utterances: Spectographics measurements. *Journal of the Acoustical Society of America, 89,* 151-168.

Ohman, S.E.G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America, 41,* 310-320.

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

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

Prins, D. (1991). Theories of stuttering as event and disorder: Implications for speech production processes. In: H.F.M. Peters, W. Hulstijn and C.W. Starkweather (Eds.), *Speech motor control and stuttering.* Amsterdam: Elsevier Science Publishers.

Rubin, P.E., Baer, T. & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America, 70,* 321-328.

Saltzman, E.L. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research, Ser 15,* 129-144.

Saltzman, E.L., Goldstein, L., Browman, C.P. & Rubin, P. (1988a). Dynamics of gestural blending during speech production (Abstract). *Neural Network, 1,* 316.

Saltzman, E.L., Goldstein, L., Browman, C.P. & Rubin, P. (1988b). Modeling speech production using dynamic gestural structures (Abstract). *Journal of the Acoustical Society of America, 84 (suppl.1),* S146.

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

Saltzman, E.L. & Munhall, K.G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1 (4),* 333-382.

Saltzman, E.L., Rubin, P., Goldstein, L. & Browman, C.P. (1987). Task-dynamic modeling of interarticulator coordination (Abstract). *Journal of the Acoustical Society of America, 82 (Suppl. 1),* S15.

Schmidt, R.A. & McGown, C. (1980). Terminal accuracy of unexpectedly loaded rapid movements: Evidence for a mass-spring mechanism in programming. *Journal of Motor Behavior, 12,* 149-161.

Shaiman, S. (1989). Kinematic and electromypgraphic responses to perturbation of the jaw. *Journal of the Acoustical Society of America, 86,* 78-88.

Sussman, H.M., MacNeilage, P.F. & 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.

Thompson, J.M.T. & Stewart, H.B. (1986). *Nonlinear dynamics and chaos: Geometrical methods for engineers and scientists.* New York: Wiley.

Turvey, M.T. (1990). Coordination. *American Psychologist, 45,* 928-953.

# SPEECH MOTOR CONTROL AND STUTTERING

Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering, held in Nijmegen, the Netherlands
June 13 - 16, 1990

*Editors:*

**Herman F.M.Peters, Ph.D.**
University Hospital Nijmegen
Department of Voice and Speech Pathology
Nijmegen, the Netherlands

**Wouter Hulstijn, Ph.D.**
University of Nijmegen
Nijmegen Institute for Cognition Research and Information Technology
Nijmegen, the Netherlands

*and*

**C. Woodruff Starkweather, Ph.D.**
Temple University
Department of Speech
Philadelphia, Pennsylvania, USA

1991

EXCERPTA MEDICA

AMSTERDAM - OXFORD - NEW YORK