INSTITUT DE LA COMMUNICATION PARLÉE

Les
Cahiers
de
l'I.C.P.

SOMMAIRE

NUMÉRO 4 1998

Special issue

B

Bulletin de la Communication Parlée N° 4

1119

Tending the garden or the plant?

ELLIOT SALTZMAN[*] [**] & DANI BYRD[*]

[*]*Haskins Laboratories, 270 Crown Street, New Haven, CT 06511 USA,*
[**]*Department of Physical Therapy, Boston University, Boston, MA USA.*

1. Overview

The target paper of this commentary – «The Equilibrium Point Hypothesis and Its Application to Speech Motor Control» – by Perrier, Ostry, and Laboissière (hereafter POL) presents a sophisticated application of the λ-version of the Equilibrium Point Hypothesis to issues of linguistic concern in the control and coordination of speech articulators. This is done in terms of an elegant model of jaw biomechanics and neuromotor control. Their synthesis of the EP Hypothesis and jaw model will continue to encourage the consideration of biomechanics in the interpretation of articulatory kinematics. Our reflections, outlined below, generally concern the appropriateness of micro- versus macro- levels of abstraction when considering the manifestation of linguistic intentions in speech motor control. Specifically, we address 1) POL's conception of planning and invariance; 2) the significance of the linguistic task in understanding speech motor control; and 3) areas of future development for this and other speech production models in the areas of intra- and inter-articulator timing.

2. Planning and invariance

POL distinguish between inter-articulator coarticulation, which is «planned, » and intra-articulator coarticulation which is unplanned, i. e., due simply to the interaction of central commands and peripheral biomechanics. By planned coarticulation POL seem to mean the existence of different articulatory targets and/or different temporal coordinations of those targets as a function of linguistic context, for example, the identity of upcoming phonemes. However, this state of affairs also seems to obtain for POL's example of *intra*-articulator coarticulation given in Figure 5. In this figure different planned targets within two jaw kinematic degrees of freedom are used for successive instances of the same phonemes in the [isisa] sequence. Furthermore, the temporal coordination between EP points in the simulations of intra-articulator coarticulation detailed in Ostry, Gribble, and Gracco's (1996) Figure 6 suggest that timing differences are required to simulate the anticipatory patterns seen in actual data. Thus, it seems that planning in POL's terms must also be involved in intra-articulator coarticulation. The implication of this conception of planning is that for POL *linguistic* invariance (i. e.,

Bulletin de la Communication Parlée, n° 4, p. 79-83. (1998)

invariance between phonemic goals and coordinative structures) can not be defined at the level of phonemic targets specified as articulator equilibrium positions or of their timing. For POL, invariant commands refer rather to the *articulatory* invariance between a specified EP command and an intended articulatory position. Finally, work detailed in Fowler and Saltzman (1993) casts some doubt on the likelihood of context-specific, planned variations in the temporal extent of anticipatory *inter*-articulator coarticulation. In summary, it is not clear where linguistic invariance actually is realized within the POL conception of speech motor control or even, in fact, if this issue is of major concern to POL. Can (invariant) linguistic units be mapped onto invariant control structures in POL's model; and, if not, where is the appeal of this approach in comparison to models (e. g. Saltzman & Munhall, 1989; Bailly et al., 1991; Vatikiotis-Bateson et al., 1993; Guenther, 1995) that can simulate these coarticulatory behaviors using invariant control structures at a task-level?

3. Levels of abstraction

POL rightly remark that the task-dynamic model of speech production (Saltzman, 1986; Saltzman & Munhall, 1989) has « no physiological mechanism underlying the control » of the articulators and that it gives no account « of either the inertial properties or the muscle mechanical properties » of the articulators. This is because the goals of the task-dynamic model are defined at a more abstract, functional level of description than the λ-model described by POL. That is, our research goal has been an understanding not of how the articulatory plant responds to motor commands but rather how linguistic intents are made manifest in articulation. While we of course ultimately look forward to incorporating a realistic biomechanical model, we made the choice to focus on the more abstract level of control in the development of our research program since we believe that biomechanical specifics are of secondary importance in understanding the coordinative processes that shape linguistically significant changes of the vocal tract over time. In other words we posit that invariant central control signals are defined in terms of linguistic tasks and are transformed into contextually-varying muscle commands, possibly of the λ-shift type. What is *necessary* is that the system accomodates whatever biomechanics that the motor periphery offers – in the spirit of Bernstein (1967/1984) who posited that the highest degree of skill is one in which the active forces supplied by muscles complement the passive forces provided by intrinsic biomechanics and the external environment in order to produce the total force required to perform the task at hand. Speech is clearly such a highly skilled behavior.

4. Timing : state and parameter dynamics

In a dynamical framework of skilled behavior, ideally one would like to elucidate not only the system's state dynamics but also its parameter dynamics. State dynamics refers to the shaping of motion patterns in the system's state variables such as position and velocity; parameter dynamics refers to shaping of the changes in the system's parameter values such as equilibrium position (or target) and co-contraction (or stiffness). (See Farmer, 1990; Saltzman & Munhall, 1992; for more complete accounts of state- and parameter-dynamics and, additionally, graph-dynamics). One reservation we have regarding the target article is its abandonment of the overall dynamic strategy that made Feldman's original work (e. g. 1966) on the λ-model so compelling : namely, the focus on describing an underlying autonomous dynamical system that gives rise to the patterning of a limb's kinematic state variables during the course of skilled positioning and rhythmic tasks. This dynamic approach has not been extended from state- to parameter-dynamics. When POL address the patterning of the system's parameters, a dynamic approach is abandoned in favor of the concatenation of piecewise-constant-velocity trajectories for equilibrium position and co-contraction.

Admittedly, we and our colleagues are likewise guilty. Until recently we have used rule-generated gestural scores to describe the manner in which each gesture in an utterance is actively incorporated into the ongoing spatiotemporal control of vocal tract shape. The trajectory of each gesture's *activation* coordinate defines a forcing function specific to the gesture and acts to insert the gestural parameters into the interarticulatory dynamical system defined by the tract-variable and model articulator coordinates. Additionally, the activation function gates the components of the forward kinematic model (from model articulators to tract variables) associated with the gesture into the overall forward and inverse kinematic computations. However, in recent work (Saltzman, 1995; Rubin, Saltzman et al., 1996; Saltzman et al., in press(a)) we have begun to explore the dynamics that give rise to the activation functions. A recurrent, sequential network architecture (Jordan, 1986, 1990, 1992) provides a dynamical shaping of gestural activation trajectories in a manner that generates appropriately constrained anticipatory coarticulation. In this network an output unit exists for each gesture in a sequence with the values of the output units corresponding to gestural activations. The temporal coordination among the network's output units, that is, gestural activations, then becomes an implicit consequence of the network architecture and the values of its connection weights.

4.1 *Articulatory consequences of parameter trajectory shapes*

In early task-dynamic work, activation trajectories were defined for the sake of simplicity as step functions switching discretely between values of zero (the gesture has no influence on tract shape) and one (the gesture has maximal influence on tract shape). However, it has been noted by POL and others for quite some time that a simple step-function activation waveshape is an oversimplification (e. g., Coker 1976; Bullock & Grossberg, 1988; Kröger et al., 1995; Ostry et al., 1996). We have begun to explore the consequences of non-step-function activation waveshapes on articulator kinematics and their implications for explaining individual differences in gestural velocity profiles (Byrd & Saltzman, in press). Similar to POL's EP and co-contraction command trajectories, our gestural activation trajectories have gradual rise and fall intervals; specifically, our simulations use a half-cosine rise and fall waveshape. Interestingly, we found that phrase-initial gestural lengthening can be modeled using variations of gestural natural frequency (stiffness) and activation rise-time. Furthermore, speakers adopt different strategies for producing the lengthening, with corresponding differences in the shapes of their velocity profiles. The velocity profiles of one of our speakers became more and more positively skewed as lengthening increased. We were able to model this individual's pattern with a lowered natural frequency and no change in the activation wave's rise time (i. e., the frequency of the half-cosine). Other speakers produced the gestural lengthening at phrase boundaries while maintaining a relatively constant proportional time to peak velocity. We were able to model this pattern with a lowered gestural stiffness and a concomitantly lengthened rise-time – specifically, however, the gestural natural frequency and the frequency of the cosine rise-time were constrained to be the same in order to maintain the nearly constant skewing of the velocity profile.

4.2 Phase-resetting

Our final reflection concerns a crucial aspect of speech timing which lacks an account in all current speech production models – namely, the fact that the temporal structure of central commands governing inter-gestural timing can be altered by events at the motor, and possibly auditory, periphery. Phase-resetting experiments of Saltzman and colleagues (Saltzman et al., 1995; Saltzman et al., in press(b)) have shown that central control of intergestural timing is sensitive to and can be changed by mechanical perturbations delivered to the articulators during both repetitive and naturalistic speech. Such resetting occurs only when the perturbation is delivered within a « sensitive phase » of the cycle during which it opposes the just-initiated, actively controlled gesture. These results imply that the central clock is a dynamical system that not only governs the parameter trajectories that drive the articulators, but is also modulated by peripheral feedback regarding articulatory state. Any model of speech production that posits preplanned parameter (e. g., EP and co-contraction) trajectories will not be able to account for such findings.

One other finding from the phase-resetting experiments suggests the need for a more abstract, or task-level, specification into which individual articulatory movements cohere. Perturbations delivered during a /... pæpæpæ... / sequence induce systematic steady-state shifts in the timing between successive bilabial closing and laryngeal devoicing gestures for /p/. Significantly, while shifts in the *relative* phasing of these gestures exist, they are nearly an order of magnitude smaller. (See Saltzman et al., in press(a), for a more complete exposition.) That is, the individual temporal shifts of the bilabial and laryngeal gestures are large compared to the relative temporal shift between these gestures, and the lips and larynx appeared to be phase-advanced as a relatively coherent unit. Thus, these results imply that intergestural temporal cohesion is greater within segments than between segments as has been hypothesized by Byrd (1996); Löfqvist (1991); Nittrouer et al. (1988) and, Saltzman & Munhall (1989). Such cohesion underscores the necessity for identifying a dynamics, effectively a *parameter* dynamics, of intergestural timing.

5. Conclusion

We have outlined above some comments regarding POL's sophisticated application of the EP Hypothesis to issues in speech motor control. Of special interest to us is the general question of level of description. We would be thrilled to find a resolution which incorporates an explicit understanding of the plant, such as POL have begun to provide, with an understanding of the broader linguistic context, « garden » if you will, in which the plant is but one member of a complex ecosystem with its own higher-level dynamics.

Acknowledgement

This work was supported by NIH grants DC-00016, DC-00121, OC-03663 and DC-03172.

References

Bailly, G., Laboissière, R., & Schwartz, J. L. (1991). Formant trajectories as audible gestures : An alternative for speech synthesis. *Journal of Phonetics, 19,* 9-23.

Bernstein, N. A. (1967). *The coordination and regulation of movements.* London : Pergamon. Reprinted in H. T. A. Whiting (Ed). (1984) *Human motor actions : Bernstein reassessed.* New York : North-Holland.

Bullock, D., & Grossberg, S. (1988). Neural dynamics of planned arm movements : Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review, 95,* 49-90.

Byrd, D. (1996). A phase window framework for articulatory timing. *Phonology, 13,* 139-169.

Byrd, D., & Saltzman, E. (in press). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics.*

Coker, C. H. (1976). A model of articulatory dynamics and control. *Proceedings of the IEEE, 64,* 452-460.

Farmer, J. D. (1990). A Rosetta Stone for connectionism. *Physica D, 42,* 153-187.

Feldman, A. (1966). Functional tuning of the nervous system with control of movement or maintenance of a steady posture – II. Controllable parameters of the muscle. *Biophysics, 11,* 565-578.

Fowler, C., & Saltzman, E. (1993) Coordination and coarticulation in speech production. *Language and Speech, 36,* 171-195.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review, 102,* 594-621.

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. (1990). Motor learning and the degrees of freedom problem. In M. Jeannerod (Ed), *Attention and Performance XIII.* Hillsdale (NJ) : Erlbaum.

Jordan, M. I. (1992). Constrained supervised learning. *Journal of Mathematical Psychology, 36,* 396-425.

Kröger, B., Schröder, G., & Opgen-Rhein, C. (1995). A gesture-based dynamic model describing articulatory movement data. *Journal of the Acoustical Society of America, 98* (4), 1878-1889.

Löfqvist, A. (1991). Proportional timing in speech motor control. *Journal of Phonetics, 19,* 343-350.

Nittrouer, S., Munhall, K. G., Kelso, J. A. S., Tuller, B., & Harris, K. S. (1988). Patterns of interarticulator phasing and their relation to linguistic structure. *Journal of the Acoustical Society of America, 84,* 1653-1661.

Ostry, D. J., Gribble, P., & Gracco, V. L. (1996). Coarticulation of jaw movements in speech production : Is context sensitivity in speech kinematics centrally planned? *The Journal of Neuroscience, 16* (4), 1570-1579.

Perrier, P., Ostry, D. J., & Laboissière, R. (1996). The equilibrium point hypothesis and its application to speech motor control. *Journal of Speech and Hearing Research, 39.*

Rubin, P. E., Saltzman, E., McGowan, R., Goldstein, L., Tiede, M., & Browman, C. (1996). CASY and extensions to the task-dynamics model. *Proceedings of the First ESCA Tutorial and Research Workshop on Speech Production Modeling,* 125-128. Autrans, France.

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

Saltzman, E. (1995). Intergestural timing in speech production : Data and modeling. In K. Elenius, & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences,* 84-91.

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

Saltzman, E. L., & Munhall, K. G. (1992). Skill Acquisition and Development : The roles of state –, parameter –, and graph-dynamics. *Journal of Motor Behavior, 24, 1,* 49-57.

Saltzman, E., Löfqvist, A., Kinsella-Shaw, J., Kay, B., & Rubin, P. (1995). On the dynamics of temporal patterning in speech. In F. Bell-Berti, & L. Raphael, (Eds.), *Studies in speech production : A Festschrift for Katherine Safford Harris.* Woodbury, New York : American Institute of Physics.

Saltzman, E., Löfqvist, A., Mitra, S. (in press(a)). «Clocks» and «glue» – global timing and intergestural cohesion. In M. Broe and J. Pierrehumbert (Eds), Papers in Laboratory V. CUP.

Saltzman, E., Löfqvist, A., Kay, B., Kinsella-Shaw, J., and Rubin, P. (in press(b)). Dynamics of intergestural timing : A perturbation study of lip-larynx coordination. Exp. Br. Res.

Vatikiotis-Bateson, E., Hirayama, M., Wada, Y., & Kawato, M. (1993). Generating articulator motion from muscle activity using artificial neural networks. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (RILP), 27,* 67-77.