Coordination, not Control, is Central to Movement

Fred Cummins

School of Computer Science and Informatics
University College Dublin
fred.cummins@ucd.ie `fred.cummins@ucd.ie`
`http://pworldrworld.com/fred`

Abstract. The notion of the control of action is contrasted with that of coordination. In coordinated action, many parts of the body (or bodies) come together to act as if they served a specific purpose, recognizable as a behavioral goal. Such simpler domains of yoked components are called coordinative structures. Examples are given of the harnessing of components into coordinative structures. In the first case, known as synchronous speech, two speakers are subsumed within a single dyadic domain of organization that exists for as long as the speakers speak in synchrony. In the second case, a time-varying set of articulators work collaboratively in generating natural and fluent movement in accordance with a behavioral goal consisting of a desired utterance. In the latter case, we introduce a new model, extending the venerable task dynamic model familiar to students of articulatory phonology. In the new embodied task dynamic model, precise gestural timing arises, not from computation and control, but from considerations of optimality in movement. A candidate function for optimization combines terms derived from the estimation of articulatory effort, perceptual clarity, and speech rate. Both of these examples illustrate a methodological advantage of dynamical models that demand that the modeler first identify both components and system boundaries as they occur within the context of a specific behavioral goal. This contrasts with many approaches within computational cognitive science.

Key words: speech production, motor control, coordination, dynamical systems, embodiment, autonomy

1 Introduction

In the study of human behavior, and its relation to brain activity, misunderstandings abound. The phrase “motor control” invites a particularly pernicious misreading that is likely to trap the unwary into a cartoon vision of the brain as a master puppeteer, simultaneously sending “control signals” down neural wires, that have “motor action” or behavior, as their end product. This interpretation is a grotesque distortion of the role of the brain. It fundamentally mischaracterizes its place in the systematic organization of behavior. The puppeteer that lurks behind a simplistic interpretation of the notion of “control” is a homuncular invention whose mysterious agency appears to be made manifest through the inert machinery of the body.

Unfortunately, something like this myth informs most accounts of overt behavior, particularly those in which behavior is seen as the output of a system of three parts: *perception*, considered as input, *cognition* as the substantial middle, and *action* or overt behavior as the final product. This is the familiar architecture of conventional cognitive psychology, and it has, for many, the status of orthodoxy. It is perhaps unsurprising that those who regard their inquiry as directed towards the cognitive heart of this hypothetical system should expend little effort in questioning the role of its peripheral twins, perception and action. Yet it seems, to this author, that the study of the origin and form of action has languished disproportionately on the fringes of cognitive science (and cognitive psychology in particular), and that a fundamentally different account of the role of the brain is becoming available within what we might term a post-cognitivist framework [19, 14, 28, 7, 6].

An alternative to the notion of *control* in describing movement for goal directed behavior is available, and this is *coordination* [14]. Two brief accounts of coordination during speaking will be presented here. Each of them makes use, not of information processing concepts, but of the vocabulary of Dynamic Systems Theory. The first looks at coordination across multiple individuals, and the second looks at coordination among a constantly changing set of body parts within an individual. In each case there is a system that is understood to constitute a well-formed domain in which the constituent parts exhibit lawful interrelations. In neither case is this system a perception-cognition-action unit, and in neither case is the notion of control of use in teasing out the lawfulness we observe within the system. Both bodies of work may be of interest to speech scientists, for whom the specific tasks involved are of obvious relevance. It is to be hoped that the modeling issues that arise, and the implications of adopting a dynamical perspective in understanding action will be of interest to a wider set of researchers within the emerging post-cognitivist framework.

2 Coordination and Coordinative Structures

As long ago as 1930, the Russian physiologist Nikolai Bernstein observed a curious and telling characteristic of the skilled movements of blacksmiths as they repeatedly hit an anvil with a hammer [1, 16] (Fig. 1). He recorded movement at the shoulder, elbow, hand, and at the point of contact between hammer and anvil. Variability from blow to blow was minimized at the point of contact, and not at any of the biomechanical joints. This seems appropriate for skilled action, as the behavioral goal that finds expression here is best expressed at that point, while there are many potential configurations of the limb segments that can give rise to equally accurate hammer blows. But it raises huge problems for any account of the movement as arising from central control directed from brain towards the periphery. If the brain were issuing control signals, and we make the further not unreasonable assumption that any biological process is attended by some non-zero noise level, then noise or error introduced at the shoulder joint would appear as additive noise at the elbow joint, and error from both shoulder and elbow should appear together with wrist noise at the wrist joint. In short, in a multi-link system, distal errors are predicted to be larger than proximal errors. Even if some error correction were possible at the elbow to compensate for shoulder error, it is inconceivable that the point of minimum variation should be the point of contact between hammer and anvil, where direct intervention by the brain is impossible in principle. This argument holds true whether the controlled variables are taken to be joint angles, torques, muscle lengths, or any other candidate.

Fig. 1. Movement variability in this skilled action is minimized at the point where hammer meets anvil, and is greater at the shoulder, elbow, and wrist joints.

This example illustrates a well-known, but perplexing, characteristic of skilled action: although the actor is possessed of a hugely flexible biomechanical system with an innumerable number of potential degrees of freedom, when this system is engaged in the pursuit of a specific behavioral goal, it behaves *as if* it were a much simpler system, with vastly fewer degrees of freedom, in which all the parts work cooperatively towards the attainment of that goal. Perturbation or error at one point of the system is smoothly and rapidly compensated at another part. The rapidity and specificity of compensation appears to rule out any processing architecture that consists of an executive centre that is informed of distal errors and that computes appropriate compensatory changes to control signals. For example, Kelso et al. found that a perturbation to the jaw during production of the final consonant of either /bæz/ or /bæb/ generated compensatory movement that was specific to the articulatory goal currently constraining movement [15]. When the target was /z/, there was an almost immediate response of the tongue, while when the goal was /b/, the response was seen in the movement of the two lips. Crucially, the goal-specific response kicked in approximately 20 ms after the perturbation, which is sufficiently rapid to rule out an account couched in terms of monitoring and error-correcting by an executive controller.

These kind of observations have led to the notion of a *coordinative structure*, which is a task-specific functional linkage of specific body parts, such that they act as a unit in the achievement of the behavioral goal [27]. Another term frequently employed is *synergy* [16]. On this view, parts of the body flexibly partake in a series of organizational forms that are defined by the behavioral goals. We might speak of the emergence, maintenance, and dissolution of special-purpose kicking machines, scratching machines, speaking machines, throwing machines, etc. Skilled action is found to organize the many parts of the body so that they act with common purpose. The effective complexity of the biomechanical system partaking in the action is greatly reduced once a clear and practiced behavioral goal is established.

When body parts become coordinated in this fashion, they constitute a domain of organization within which the component parts are lawfully related. Change in the state of any one component is not entirely independent of change in the state of any other. Components may themselves be complex entities that can be decomposed into sub-constituents that are coordinated to bring about the component-level behavior, and no single level of behavioral description can claim to be privileged. For example, Kelso and colleagues have long studied the form of coordination exhibited when two effectors (fingers, hands, etc) are constrained to oscillate with a common frequency [14]. Given the behavioral goal provided by the task description, only two forms of stable coordination are observed, one in which the effectors oscillate with common phase (synchrony) and one in which they are half a cycle out of step with each other (syncopation). Many features of this system, including rate-dependent multi-stability, phase transitions, critical fluctuations, hysteresis, etc, have been modeled using the Haken-Kelso-Bunz (HKB) model [12]. Details of the model are not relevant here, but the structure of the model provides an insight into how dynamical models might approach the systematic simplification that is evident in skilled action. First, the behavior of individual components is characterized. In the present case, that amounts to describing each effector as a self-sustaining oscillator. Then the behavior of the components in the service of the task is described. The lack of independence between the effectors ensures that this system-level description is simpler (of lower dimension) than a full description of the components and their mutual interactions. In the HKB case, the model then provides a formal account of how the higher dimensional component description collapses to the simpler system-level description.

The system being described here is not a whole person. It is the two effector system, which is considered as a whole—a domain of organization constituted by the coordinated movement of two components. In developing a dynamical account of observed phenomena, the identification of the domain of the model is an important first step, formalized as the definition of the *state* of the system. The selection of appropriate state variables allows the modeler to potentially identify lawful domains of organization that transcend the somewhat arbitrary boundaries separating brains, bodies and environments. This is clearly illustrated in the interpersonal coordination documented in Schmidt, Carello and Turvey (1990). In that study, the task of oscillating two effectors at the same frequency was distributed across two individuals, seated, each wagging their lower leg. The same hallmarks of differential pattern stability were found in this scenario, in that there were two and only two stable patterns of coordination (synchronous/syncopated), there was a greater stability of the synchronous pattern and a strong tendency for the syncopated pattern to transit to the synchronous one in a rate-dependent manner. Of course these two individuals are not obligatorily coordinated. Each is free to stand up and go about their individual lives. But *in the context of the behavioral task*, they behave (or more accurately, the system comprising their two legs behaves) as as a simpler system with few degrees of freedom.

3 Synchronous Speech

Something very similar is seen in the coordination displayed by subjects within the synchronous speech experimental situation. Synchronous speaking is a behavioral task in which a pair of subjects are given a novel text, they familiarize themselves briefly with the text, and then they read the text together, in approximate synchrony, on a cue from the experimenter [8]. It is distinguished from the related notion of choral speaking, in that the text is new and is thus not produced with the exaggerated prosody familiar from the group recitation of oaths, prayers, etc. Typically, subjects are very good at this task, and without any practice, they maintain a tight synchrony with lags (temporal offset between their speech streams) of no more than 40 ms on average [9]. This is comparable to an asynchrony of no more than a single frame of video. Perhaps remarkably, practice does not generate much improvement; the synchronization task appears to tap into a natural facility for synchronization with another person, despite the complexity of the task of speaking.

How should we view the sustained exhibition of very tight synchrony between two speakers (Fig. 2)? One way, and that most readily at hand within most current approaches, is to view each speaker as an independent system. A speech production system within each individual is held responsible for the planning and execution of movement. To this picture, we would have to add a perceptual component that monitors the speech being produced by the other, that compares one production with the other, and that makes corresponding adjustments. This unwieldy picture appears obligatory if we treat the brain as controlling puppeteer, and if we view the people involved as perception-then-cognition-then-action systems. Construing the synchronous speaking situation

Fig. 2. As modelers, we have the freedom to chose to regard a pair of synchronous speakers as a single system, or as two distinct and interacting systems.

in this way, we would expect to see evidence of drift between speakers and compensatory error correction at time lags that would allow for re-planning and alteration of control parameters. As that is not what we observe experimentally, one might instead regard one speaker as providing a lead, to which the other responds. This would suggest that somewhat stable performance might be found with a fixed lag between speakers. This is never the case. For each speaker, the inherent variability that attends normal unconstrained speaking would be augmented by the additional requirement of attempting to match the timing of the co-speaker. Variability would therefore be predicted to increase. It does not. The difficulty in adequately describing (modeling) the situation stems from the prior commitment to the locus of control as lying within an individual.

If we instead view the two speakers as enslaved components within a single overarching system, each of them both driven and driving the behavior of the system as a whole, our expectations would be rather different. Where we know speech production to be inherently complex and variable, even within the speech of a single individual, we would expect a simplification, or reduction in variability while the speakers are behaving as components within a superordinate system. This is, indeed, what we find [10, 11]. Variability in segment duration, in pitch movement, and in pause behavior are all reliably found to be reduced in synchronous speech, as compared to speech that is not so constrained. If the components of the system are mutually correcting, just as in the coordinated movement of body parts within a skilled individual, we would expect no clear leader-follower behavior, and this is, in fact, what we find.

A blacksmith's arm is perfectly capable of wielding a violin bow, of scratching a blacksmith's chest, or of shaking the hand of the fishmonger. But when the blacksmith pursues a well-defined behavioral goal, demanding skilled (and hence unreflective) action, the arm acts as part of an overall domain of organization that is brought into existence in pursuit of just that goal, and that ceases to exist once the blacksmith turns to other tasks. So too, in speaking synchronously, each speaker temporarily becomes part of a larger organizational domain, and the speaker acts as if he were a component within the larger system. Tellingly, we find that when one speaker makes an error, the usual result is for the entire fragile coordinative system to fall apart, each speaker recovers full autonomy, and the speech of each is immediately and completely decoupled from that of the other. The error destroys the boundary conditions imposed by the common behavioral goal.

If this account of coordination, within and across individuals, appears counter-intuitive, it is probably because the conviction is rooted very firmly that behind every lawful intentional action must lurk a controller. How else, we might reasonably ask, is one possible movement selected out of many alternatives? How, indeed, are we to account for volition in action and our sense of being the authors of our own lives? Such existential qualms are probably not warranted here, and some of the unease may be vanquished by recognizing that the type of model being developed within a dynamical framework is fundamentally different from that within a perception-cognition-action, or cognitivist, framework[1]. The dynamic account of action suggests that the fluid movement typical of skilled coordination derives its form from the lawful constraints operative on the many components that contribute to the overall behavior. An account is then required of the nature of these constraints. Why is that we observe one form of movement and not another? This question is particularly vexing as the behavioral goal that underlies the temporary organization of parts into a single-purpose domain of organization does not, by itself, contain any specification for how that goal is to be achieved. In reaching to scratch my nose, there are many possible trajectories my hand and arm could take, and one of these actually happens, without any sense of pondering, selection, or doubt.

An important part of the answer is to look more closely at the specification of the behavioral goals, and to see to what extent they might serve to differentiate among possible forms of movement. Given specific goals, some forms of movement may be *optimal* in a strict and quantitative sense, and optimality criteria may be the best candidates for formal expression of the constraints that are operative. For example, it has been demonstrated that gait selection in horse locomotion makes sense when considered in light of energetic requirements [13]. For each of the three gaits studied, walking, trotting, and galloping, the metabolic cost varied with rate. This allowed identification of rates for each gait at which energetic costs are minimized. Horses observed in the paddock adopting these gaits spontaneously did so at rates that are, in fact, approximately optimal. Locomotion is a form of action that has been shaped both phylogenetically over many millennia, and ontogenetically through constant practice. It seems highly plausible that the resulting action is constrained to be optimal with respect to many potential criteria. Analysis of bone strain as a function of speed leads to similar conclusions as the analysis of metabolic cost as indexed by oxygen consumption [13, 2].

---

[1] One way of describing the difference between the modeling approaches is available in the Aristotelian distinction between *efficient* cause, which comfortably accommodates the notion of a controller, and *formal* cause, which describes lawful domains of organization without the need to commit to any such central executive.

We turn now to another characterization of speech coordination within a model, but framed at quite a different level. We demonstrate that optimality criteria may provide some explanatory power in interpreting the form of observed movement, and may help to bridge the gap between accounts couched in terms of control or in terms of coordination by reducing the need for explicit intervention by a controller.

4 Embodied Task Dynamics

When we speak, the set of goals we embody can be described in many ways: transmission of a sequence of words, effecting a response in the other, making a particular kind of sound, etc. One particularly informative way of describing speech production is as a sequence of articulatory gestures produced in parallel streams, where the gestures correspond to primitive units of combination within a phonology. This is the basic premise of Articulatory Phonology [3], and has provided a powerful explanatory framework for understanding both categorical and gradient features found in articulation [5]. Fig. 3 shows a partial gestural score (by analogy with a musical score) for the utterance /pan/. Each row (or 'tier') is associated with a (vocal) tract variable. These are linguistically significant dimensions of variation. Note that the times of individual gesture onsets and offsets do not necessarily align across tiers, as gestures are not simply coextensive with phonemes. Thus velum lowering precedes tongue tip movement for the /n/, as is well known from phonetic data. Each tract variable is represented as a simple mass-spring dynamical system for which a target equilibrium position is provided during periods in which the gesture is active, as specified in the score. During periods of activation, tract variables move smoothly towards their targets, and when the associated gesture is no longer active, they relax to a neutral position. It is then necessary to map from the space of tract variables (which are all independent of one another and thus context-free) to the space of articulators, where multiple tract variables may compete for influence over a specific articulator. For example, the jaw is crucially involved in three of the four tract variable movements shown, and for a period, there are conflicting influences on the jaw, pushing it lower for the /a/ target, and raising it for /n/.

The tricky business of mapping from tract variables to articulators within an articulatory synthesis system is provided by the task dynamic model, originally introduced in [21] to account for limb movement, and later extended to the speech domain in [22]. Task dynamics uses techniques from linear algebra to uncover the optimal mapping from tract variable motions (arising from the mass-spring dynamics) to model articulator motions. In this way, the nice, smooth motion resulting from relatively simple dynamical systems (tract variables) can be manifested in the more complex space of a model vocal tract. Together, articulatory phonology and its task dynamic implementation have been very successful at accounting for a wide range of linguistic phenomena, and at linking observed movement to underlying discrete behavioral goals through a principled and explicit mapping [4, 5].

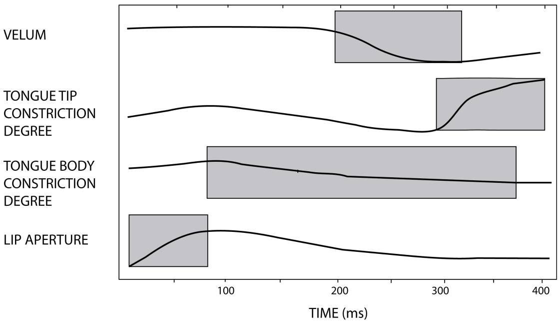

Fig. 3. Partial gestural score for the utterance /pan/. Blocks represent activation intervals. Curves represent tract variable movement.

A gestural score contains information about both sequential order, and precise timing. It can thus be interpreted as a sophisticated kind of control algorithm, with task dynamics providing the means by which its control variables are made to do real work. Getting the timing right within the gestural score is thus a difficult problem, and one of fundamental importance. Past approaches have sought to learn appropriate timing from articulatory data using neural networks [22], using an extra layer of planning oscillators [18], and by introducing a degree of flexibility in relative timing through so-called phase windows [20].

In a recent development of the task dynamic model, Simko and Cummins [25] sought to divide the control task represented in the gestural score into two distinct parts. The specification of serial order among gestures is specified independently of the timing among gestures. A modified form of task dynamics then makes it possible to express constraints under which the speech is produced, and these constraints, in turn, allow us to distinguish forms of movement that are more or less efficient, and thus to find an optimal gestural score that satisfies the overt goal (the gesture sequence) and that reflects the constraints under which speech is produced. Fig. 4 shows a simple gestural score before and after optimization. The top panel (before) specifies only the linear order of the sequence /abi/, while precise timings have emerged after a process of automatic optimization in the lower panel. In the radically simplified vocal tract model employed so far, we model only the consonants /p/ and /b/[2] and the vowels /a/ and /i/.

Optimization is based upon a cost function with three weighted components: $C = \alpha_E + \alpha_P + \alpha_D$. Collectively, these express the high-level constraints under which speech is produced. The first two components, α_E and α_P serve to establish a trade off between ease of articulation and communicative effectiveness. α_E is a quantification of the effort expended in executing a series of movements, while α_P is a cost that rises if articulation is sloppy, or the speech is hard to parse by a listener. Together, the relative magnitude of these weights serves to locate

---

[2] Without a glottal model, there is no meaningful distinction within the model between these stops and their voiced counterparts, /b/ and /d/.

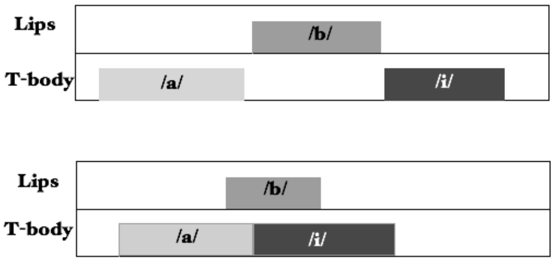

Fig. 4. Gestural score before and after optimization.

a given production as lying on a specific point on a scale from hypo-articulation to hyper-articulation [17]. The third component, α_D places a relative cost on executing an utterance quickly. Full details of the model are presented in [24, 26] and a summary overview can be found in [25].

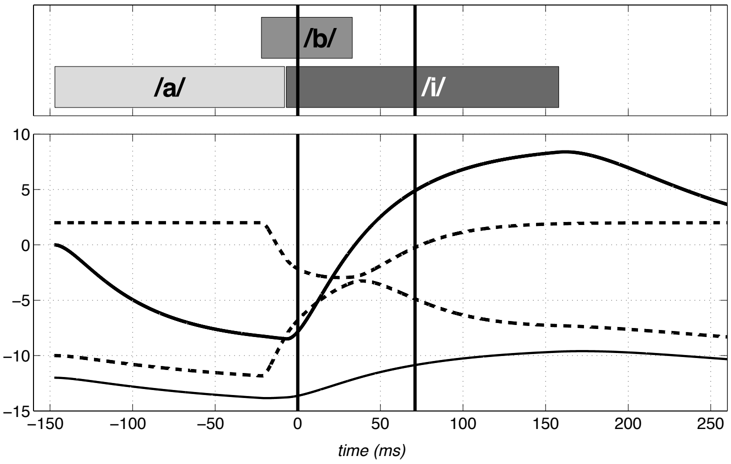

Fig. 5. Sample movement traces after optimization. Heavy solid line: tongue body; Light solid line: jaw; Dashed lines: lips. Vertical lines demarcate the interval of consonant closure.

Fig. 5 illustrates the output of the model. Movement traces are shown for jaw (light solid line), tongue body (heavy solid) and lips (dashed). The vertical lines demarcate the interval of consonant closure. Movement of the lips within that interval arises from soft tissue compression. This score, and the associated movement traces, is derived fully automatically from the sequence specification /abi/ and specific choices of weights for the three elements in the cost function,

C. The resulting form of movement is fluent, and to date, has matched published details of articulatory movement with a high degree of accuracy [26]. Further development, including extension to a full vocal tract model, will be guided by precise articulatory data.

The gestural score of the original articulatory phonology/task dynamics model is a form of control algorithm that specifies both sequence and precise timing. In the more recent embodied model, these jobs are separated. The input to the model (the residual control element) comprises just the behavioral goal expressed as a sequence of gestures, without explicit timing information. To these is added the specification of constraint under which speech is produced (the cost function weights), and this collectively allows computation of the optimal form of movement. This is no longer a control algorithm, but rather an account of the form of observed movement couched in terms of discrete behavioral goals along with high-level intentional constraints (weights).

5 Dynamics and Autonomy

In each of the two examples presented herein, an attempt is made to understand the form of observed behavior, while remaining somewhat agnostic as to the underlying (efficient) causes. We observe tight synchrony among speakers and wish to find the best characterization thereof that can account for reduced variability, sustained synchrony without leaders, and the fragility of the cooperative behavior when errors creep in. The conceptual tools of dynamical systems, and in particularly their application to understanding coordination, provide an appropriate vocabulary and instrumentarium for capturing this regularity. They do so by taking the identification of the system to be modeled as a critical part of scientific inquiry. Rather than assuming that the sole domain of autonomy in behavior is the individual person, the dynamical approach here posits the temporary capture of two subjects within a superordinate dyadic domain.

In the embodied task dynamic model, the set of entities that exhibit mutual coordination changes over time. The boundaries of the set are determined by the gestural score, which in turn arises from a minimal set of discrete behavioral goals, and some context-specific constraints. Collectively, these serve to identify an optimal form of movement, where ‘optimality’ has a precise and quantifiable meaning. Again, the domain within which lawfulness and constraint operate is not give *a priori*, but is rather dependent on time-varying behavioral goals and the context within which behavior occurs.

The brief accounts provided here to not do justice to either experiment or model. For those details, the reader is encouraged to seek out the primary publications referenced herein. The goal here has been to show how the adoption of a dynamical perspective can help in understanding the structure and lawfulness of behavior, but that this approach also demands an openness with respect to the system being modeled. Any dynamic modeling must first be explicit about the set of state variables to be considered. This initial choice positively encourages the creation of models that cut across the somewhat arbitrary boundaries that separate brain, body, and environment, and gives the modeler pause for thought about the domain being studied: is it a person, a well-defined subset of the person, a person plus a tool, a set of persons? The right answer will, in all cases, be an empirical issue, and will depend on where lawfulness is observed. This flexibility to identify lawfulness at many different levels may ultimately encourage us to identify time-varying domains within which components exhibit interdependent behavior that is both constitutive of, and constrained by, the system level organization.

A final word is appropriate to fend off an inevitable potential point of confusion. The notion of autonomy within dynamical modeling is clearly separate from the concept of agency. The autonomous domain is precisely that set of variables that exhibits a lawful set of relations among its components. To identify, e.g. a dyad, as an autonomous domain is to make an observation about the structure and form of their collective behavior. It does not say anything at all about the more mysterious notion of agency. Even though individuals may act as relatively simple components within an overall pattern, this does not rob them of their intrinsic autonomy. This is clear when we realize that participants in a Mexican wave are acting as simple components in a collective organization, without any sacrifice of their ability to later leave the game and go buy a hot dog.

References

1. N. A. Bernstein. A new method of mirror cylographie and its application towards the study of labor movements during work on a workbench. *Hygiene, Safety and Pathology lf Labor*, 5:3–9 and 6:3–11, 1930. In Russian. Cited in Latash (2008).
2. A. A. Biewener and C. R. Taylor. Bone strain: a determinant of gait and speed? *The Journal of Experimental Biology*, 123:383–400, July 1986.
3. Catherine P. Browman and Louis Goldstein. Articulatory phonology: An overview. *Phonetica*, 49:155–180, 1992.
4. Catherine P. Browman and Louis Goldstein. Dynamics and articulatory phonology. In R. F. Port and T. van Gelder, editors, *Mind as Motion*, chapter 7, pages 175–193. MIT Press, Cambridge, MA, 1995.
5. C.P. Browman and L. Goldstein. Articulatory gestures as phonological units. *Phonology*, 6(02):201–251, 2008.
6. P. Calvo and A. Gomila. *Handbook of Cognitive Science: An Embodied Approach.* Elsevier Science Ltd, 2008.
7. A. Chemero and M. Silberstein. After the Philosophy of Mind: Replacing Scholasticism with Science. *Philosophy of Science*, 75:1–27, 2008.
8. F. Cummins. On synchronous speech. *Acoustics Research Letters Online*, 3(1):7–11, 2002.
9. F. Cummins. Practice and performance in speech produced synchronously. *Journal of Phonetics*, 31(2):139–148, April 2003.
10. Fred Cummins. Synchronization among speakers reduces macroscopic temporal variability. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, pages 304–309, 2004.
11. Fred Cummins. Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37(1):16–28, January 2009.

12. H. Haken, J. A. S. Kelso, and H. Bunz. A theoretical model of phase transitions in human hand movement. *Biological Cybernetics*, 51:347–356, 1985.
13. D.F. Hoyt and C.R. Taylor. Gait and the energetics of locomotion in horses. *Nature*, 292(5820):239–240, 1981.
14. J. A. Scott Kelso. *Dynamic Patterns.* MIT Press, Cambridge, MA, 1995.
15. J.S. Kelso, B. Tuller, E. Vatikiotis-Bateson, and C.A. Fowler. Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10(6):812–832, 1984.
16. M.L. Latash. *Synergy.* Oxford University Press, USA, 2008.
17. Björn Lindblom. Explaining phonetic variation: a sketch of the H&H theory. In William J. Hardcastle and A. Marchal, editors, *Speech Production and Speech Modelling*, pages 403–439. Kluwer Academic, Dordrecht, 1990.
18. H. Nam and E. Saltzman. A competitive, coupled oscillator model of syllable structure. In *International Conference on Phonetic Sciences*, 2003.
19. Robert Port and Timothy van Gelder, editors. *Mind as Motion: Explorations in the Dynamics of Cognition.* Bradford Books/MIT Press, Cambridge, MA, 1995.
20. E. Saltzman and D. Byrd. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, 19(4):499–526, 2000.
21. Elliot Saltzman and J. A. S. Kelso. Skilled actions: A task dynamic approach. *Psychological Review*, 94:84–106, 1987.
22. Elliot Saltzman and Kevin Munhall. A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1:333–382, 1989.
23. R. C. Schmidt, C. Carello, and M. T. Turvey. Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology. Human Perception and Performance*, 16(2):227–247, May 1990.
24. Juraj Simko and Fred Cummins. Embodied task dynamics. *Psychological Review*, 2009. In press.
25. Juraj Simko and Fred Cummins. Sequencing of articulatory gestures using cost optimization. In *Proceedings of INTERSPEECH 2009*, Brighton, U.K., 2009.
26. Juraj Simko and Fred Cummins. Sequencing and optimization within an embodied task dynamic model. *Cognitive Science*, 2010. In revision.
27. B. Tuller, MT Turvey, and H.L. Fitch. The Bernstein perspective: II. The concept of muscle linkage or coordinative structure. *Human Motor Behavior: An Introduction*, pages 253–270, 1982.
28. B. Wallace, A. Ross, J. Davies, and T. Anderson. *The Mind, the Body and the World. Psychology after Cognitivism.* Imprint Academic, Charlottesville, VA, 2007.