



Periodic and aperiodic synchronization in skilled action

Fred Cummins*

UCD School of Computer Science and Informatics, University College Dublin, Dublin, Ireland

Edited by:

Leonhard Schilbach,
Max-Planck-Institute for Neurological
Research, Germany

Reviewed by:

Scott Kelso, Florida Atlantic
University, USA
Tobias Schlicht, Ruhr-Universität
Bochum, Germany

***Correspondence:**

Fred Cummins, UCD School of
Computer Science and Informatics,
University College Dublin, Belfield,
Dublin 4, Ireland.
e-mail: fred.cummins@ucd.ie

Synchronized action is considered as a manifestation of shared skill. Most synchronized behaviors in humans and other animals are based on periodic repetition. Aperiodic synchronization of complex action is found in the experimental task of synchronous speaking, in which naive subjects read a common text in lock step. The demonstration of synchronized behavior without a periodic basis is presented as a challenge for theoretical understanding. A unified treatment of periodic and aperiodic synchronization is suggested by replacing the sequential processing model of cognitivist approaches with the more local notion of a task-specific sensorimotor coordination. On this view, skilled action is the imposition of constraints on the co-variation of movement and sensory flux such that the boundary conditions that define the skill are met. This non-cognitivist approach originates in the work of John Dewey. It allows a unification of the treatment of sensorimotor synchronization in simple rhythmic behavior and in complex skilled behavior and it suggests that skill sharing is a uniquely human trait of considerable import.

Keywords: synchronization, speech, rhythm, skill, aperiodicity

How can we know the dancer from the dance? (W. B. Yeats)

1. INTRODUCTION

Synchronized action, or doing the same thing at the same time, is a very specific form of interpersonal coordination. Indeed, the difference between synchronized, and merely coordinated, action will not be categorical, but will depend rather on the precision and granularity with which we choose to define the “thing” or action undertaken. This is conveniently illustrated by considering dancers in a line dance – where the actions of each are, to a great degree – identical, and dancers in a tango pair, where each dancer has a rather different role from the other, but their joint action remains highly coordinated. Clearly, if two or more people can synchronize their actions, they share a specific skill. In what follows, the issue of what it is to share a skill will find some elaboration through study of synchronization in both simple and complex cases.

Mathematically, synchronization may be understood as a generic dynamical¹ process whereby two or more oscillatory systems interact, such that their combined state admits of a simpler (lower-dimensional) description than the mere enumeration of the states of the individual components, especially when they couple such that their cycles adopt a fixed phase relation (Pikovsky et al., 2001). The space of possible stable relations two or more systems can adopt will depend on both the form and strength of their interaction, and also on the intrinsic dynamics of each system considered autonomously. The more general notion of coordination is similarly captured by description of two or more interacting dynamical systems whose effective joint state space when doing some specific task is of lower dimension than the mere conjunction of the state spaces of the component systems (Haken et al., 1985; Kelso, 1995).

¹A good introduction to basic dynamical concepts as applied within cognitive science can be found in Norton (1995).

The mathematics of synchronization has developed enormously since its first application in the physical sciences. When we wish to employ this theoretical arsenal in the service of describing synchronization phenomena in complex animate behavior, some thought is in order. The mathematical notion of an oscillator, and the pre-theoretical notion of an act (the “thing” implied in the phrase “doing the same thing”), can only be aligned if some care is taken to circumscribe the domain of observation, and the interpretation of those observations within a sufficiently rigorous framework. This paper will attempt to tease out issues of how such dynamical concepts can properly be applied to understand animate behavior.

A canonical and (perhaps deceptively) simple example is provided in describing the coordinated movement of a group of dancers (Cummins, 2009a). The domain of observation is the dance, measurements are spatio-temporal measurements, e.g., of limb/torso position and velocity, and the periodic nature of the behavior allows a straightforward definition of phase, which is essential to the quantitative description of synchronization. “Phase” here will mean the *relative* time of an event with respect to some containing and repeating unit, which in turn is defined by what it is we understand the dance to be.

The ability to synchronize with an external signal, across a range of tempi, when engaging in repetitive behavior is known as sensorimotor synchronization (Merker et al., 2009). It appears to be a particularly human ability and is thus deserving of much attention. Synchronization with a periodic referent is most commonly found in one of two forms: in phase (matched down beats), or anti-phase (syncopated). In what follows, I will augment the discussion of simple sensorimotor synchronization with an instance of complex synchronized behavior that clearly does not have a periodic basis. This is the remarkable feat of two speakers speaking in synchrony. The manifest absence of a periodic basis for synchronization in this case presents a challenge to any explanatory framework that

sees periodicity as the *sine qua non* of synchronization. The hope is that rising to this challenge will help to refine our vocabulary for discussing synchronized behavior in a range of cases, and allow us to come at such behavior afresh, with an eye on the definition of skill, and how skills might be shared.

I will argue that non-periodic synchronization motivates a framework for understanding perception and action that departs fundamentally from cognitivist, information-processing models. I advocate discarding the notion of the cognitive system as a sequential processing chain with perceptual input leading ultimately to action output, and replacing this view with an alternative in which behavior is described with respect to a transient and task-specific domain. For the coordinated action of a single individual, identifying a given domain is tantamount to addressing the challenge above of delimiting the “thing” that actors may perform synchronously, or, indeed, of defining what is meant by the “dance.” The approach extends naturally to domains that cut across boundaries of nervous system, body and world, and specifically to domains defined over multiple, interacting, individuals. This alternative framework has much in common with work in coordination dynamics (Kelso, 1995), ecological psychology, and in theories of enactive perception, and extends back to earlier suggestions arising in the work of John Dewey.

2. PERIODIC AND APERIODIC SYNCHRONIZATION

2.1. SENSORIMOTOR SYNCHRONIZATION AS A MARK OF THE HUMAN

Synchronous behavior is often observed in the animal kingdom. Some of the better studied examples include the synchronous flashing of fireflies in both Asia (Hanson et al., 1971) and America (Copeland and Moiseff, 1994), claw waving in fiddler crab courtship (Backwell et al., 1998), and chirping in katydids (Sismondo, 1990). Weaker forms of synchronization, in which local periodic movements become mutually coordinated, without rigid frequency and phase locking, are commonplace in schooling, shoaling, and flocking behaviors.

What appears to be unique to humans, is the ability to synchronize with an exogenous periodic signal at a wide variety of tempi (Merker et al., 2009). Despite occasional prohibition, music and dance occur, and exhibit tempo variation, in every human culture, bearing witness to our ability to mutually entrain our movements, given a periodic reference signal. Although we can synchronize to a range of tempi, a strong preference for a tempo with a fundamental beat period of approximately 500 ms is well documented (MacDougall and Moore, 2005). A similar ability is not to be found in either apes or monkeys. A limited ability to track tempo changes while moving in time with music has been documented in a cockatoo (Patel et al., 2009), and some insects have been found to slightly adjust the frequency of their behavior to match an external referent (Ermentrout, 1991), but these interesting cases notwithstanding, flexible sensorimotor synchronization appears to be a particularly human ability.

Accounts of how this behavior arises are often couched within a cognitivist vocabulary that unquestioningly presumes the locus of agency, and hence the causal origin of behavior, to be the cognitive system of an individual, conceived of as a computational system operating on perceptual input, and producing behavior in

the form of movement as output. This is the conventional framework of cognitive psychology, and provides the setting for virtually all of computational cognitive neuroscience. Within a cognitivist framework, the task of perceiving temporal structure, and of regulating timing in movement is usually assigned to a timekeeping component, such as the central clock in the influential timing model of Wing and Kristofferson (1973). The central timekeeper provides an amodal timing reference that can be availed of by perceptual processes and motor processes alike.

We here adopt instead a dynamical perspective. Dynamical modeling of behavior can be happily agnostic about the locus of agency, and focus instead on domains within which lawfulness may be found in the spatio-temporal change over time of observables. These domains may be transient in nature, they may cut across the boundaries of nervous system, bodies, and environments, and they may be defined over multiple individuals as well as within a single organism. For many of these reasons, dynamical systems theory is rapidly becoming the lingua franca of post-cognitive approaches to cognitive science, with emphasis on embodiment, enaction, and the ecological embedding of organisms in environments (Kelso, 1995; Port and van Gelder, 1995; Stewart et al., 2011).

Synchronization is, of course, a dynamic phenomenon and is most naturally described using the toolbox of dynamical concepts. That can be done within a cognitivist framework as well, simply by considering the organism to instantiate one system, a periodic referent, such as a music signal or another dancer, to constitute another, and to ensure that the two systems interact. However, there are deeper reasons for adopting a non-cognitivist dynamical stance in addressing coordination and synchronization. These are motivated by the need to overcome the Cartesian presumptions that separate the living subject from the physical and social world in which they are embedded. But first it will be necessary to motivate that by turning to a puzzling case of synchronization without any underlying periodicity.

2.2. APERIODIC SYNCHRONIZATION IN SYNCHRONOUS READING: A PUZZLE

When two people are presented with a novel text and asked to read the text in synchrony with one another, they typically have no difficulty in doing so, even though the task might appear *prima facie* to be very complex. The constituents of speech, whether they be considered at the phoneme, syllable, or phrasal level, exhibit great temporal variability, even within the speech of a single individual. This inherent variability contributes to the context-sensitivity and expressiveness of the spoken word, and helps to make the voice such an exquisitely plastic and communicative medium. In a synchronous speech experiment, subjects are presented with a text they have not seen before, are allowed to read it silently once, and are then required to read in synchrony after the experimenter's signal. The mean asynchrony that is found when the two parallel speech waveforms are compared is approximately 40 ms on average, rising to 60 ms at the start of phrases (Cummins, 2003). This constitutes a very tight synchronization in which subjects diverge by approximately a single frame of a video movie. Remarkably, practice at the task does not lead to markedly better performance. Rather, most subjects find the task to be relatively easy, and they do not require a lengthy period of acclimatization to either the

task conditions, or to a specific co-speaker. It is never the case that one speaker consistently leads and the other lags behind. Rather, the speech of the two speakers seems to fuse, with only minimal leading or lagging, and no consistent leader (Cummins, 2009b). It may well be that one speaker is dominant over the other, but this may be manifested in the establishment of a joint tempo closer to the endogenous tempo of one speaker rather than the other, or it may be manifested in differential speech volume, but it is not evident in the relative timing of the two speakers.

Synchronous speech is highly constrained in an experimental setting, and is thus to be distinguished from choral speech, familiar from group recitations, oaths, prayers, etc. Choral speech typically involves over-rehearsed texts recited in large groups with heavily stylized prosody. Synchronous speech, in contrast, is often virtually indistinguishable from normal read speech, as long as no speaker makes a speech error.

Errors, when they arise, reveal the yoked, or coupled, nature of the system. One frequently observed consequence of an error by one speaker is the abrupt and simultaneous cessation of speech by both speakers – a form of speech error not observed under conventional conditions of speaking as an individual. This form of catastrophic failure is indicative of a strong constraint obtaining between the speakers, making them non-independent. A physical analogy can help here. In a three-legged race, the contestants are required to run, and to coordinate that running with their partner. The link between them is a physical one, as their legs are tied together, thereby enforcing a strong constraint on their movement. This inhibits, but does not prevent, running. Rather, running of a somewhat constrained sort is possible, if the participants coordinate successfully. If one makes an error, however, a frequent consequence is the failure of the entire coordinative ensemble, as both participants hit the ground in ungainly fashion. So too, in a synchronous speaking scenario, one can imagine a strong constraint yoking the two participants together, and bringing the joint speaking activity to a halt as one makes an error. This analogy will be fleshed out and made rather more explicit in the following section, where we will be particularly concerned with characterizing the nature of the constraint linking the two speakers.

Synchronous speaking poses an interesting puzzle. Despite pre-theoretical notions of “rhythm” in speech, there is no strict period to speech (Dauer, 1983). While there is a documented tendency for alternating stresses (in English) to form quasi-regular series, this tendency is continually frustrated and usurped by the vagaries of syntax and lexical choice (Classé, 1939; Lehiste, 1977). There is thus nothing in the speech signal that can act as a periodic referent. How, then, is such exquisite synchronization possible, despite the heterogeneous nature of the signal? In what follows, I suggest that we can recast our account of skilled action in a manner quite different from the cognitivist account, and that doing so opens up a novel space of potential accounts of coordinated behavior across individuals more generally, and provides an interesting alternative view of just what sensorimotor synchronization is and why it might be important. This dynamical perspective also has consequences for the kind of activity we might expect to find in brains as people engage in coordinated and skilled behavior in real time (see, for example, Tognoli et al., 2007; Dumas et al., 2010).

3. OF PERCEPTION, ACTION, AND SKILL

What is the difference between a gesture, such as an obscene hand movement, and an involuntary twitch, such as a tic? The first is interpretable, or meaningful, if placed in the context of human-human interaction. The second does not admit of any more elaborate interpretation. To one who can not avail of the framework of human communication, there is no obvious difference in the simple observation of bodily movement in both cases. Likewise, the difference between a scrawl and a hand-written character is not evident in the ink marks on a page, but in the framework within which they are interpreted. If there is an alphabetic framework with cursive writing conventions, we may interpret one as a character.

Intentional action is distinguishable from mere movement, not by differences in raw movement, but by the framework within which those movements are interpreted. The definition of the framework is a matter in part of convention and shared understanding. Writing systems and obscene gestures are not natural kinds, but human conventions.

We can draw a similar distinction between sensory flux² and perception. A given pattern of stimulation, be it luminance variation or sound pressure variation, is not intrinsically meaningful. It becomes meaningful to the extent that it can be interpreted by the subject based upon the subject's knowledge of the world, of sensorimotor contingencies, of perils and opportunities, of affordances. In short, the subject must make use of sensory variation within some interpretive framework.

When the head turns, there is a corresponding change in the sensory variation at the retina. This may be cast within an action framework (the subject is looking toward something) or a perception framework (the subject sees something), but both interpretations go beyond the raw data of co-variation of movement and sensory flux. The degree to which we can interpret such movement with attendant variation as an intentional act will depend upon the observer and the framework within which the observation is made. The same head movement and attendant retinal flux might be described as “looking,” or as “having one's attention captured,” or as “checking to see if the coast is clear,” or as any of an infinity of other interpretive frameworks. Describing it as constitutive of perception, or of willful action, will be licensed by selecting an appropriate framing context (Kelso, 1981).

The view that perception and action are distinguishable from sensory flux and mere movement by virtue of interpretation within an organizing domain was first laid out by Dewey (1896). He there took exception to the notion of the reflex arc, which was emerging as a unifying concept in nineteenth Century psychology (Dewey, 1896). Dewey describes the limitations of the notion of viewing the organism as a one-way processing system, with stimulus/perception as the input, and movement/action as the output. In his critique, he foreshadows both the behaviorist and the cognitivist viewpoints, emphasizing their similarity and the way in which the strict sequencing of perception and action creates an artificial separation of organism and world. This

²I use the term sensory flux to describe physical, chemical, and energetic change at a receptor surface. No experiential attribute, conscious sensation, or qualia is meant.

critique, now well over 100 years old, succinctly expresses many of the reservations about the cognitivist orthodoxy that now find expression in embodied and enactive approaches to being (Hurley, 2001; Stewart et al., 2011). Dewey laments the characterization of the relationship between organism and environment as a linear throughput thus:

The sensory stimulus is one thing, the central activity, standing for the idea, is another thing, and the motor discharge, standing for the act proper is a third. As a result, the reflex arc is not a comprehensive, or organic unity, but a patchwork of disjointed parts, a mechanical conjunction of unallied processes. . .What shall we term that which is not sensation-followed-by-idea-followed-by-movement, but which is primary. . .Stated on the physiological side, this reality may most conveniently be termed coördination.

(Dewey, 1896, p. 358)

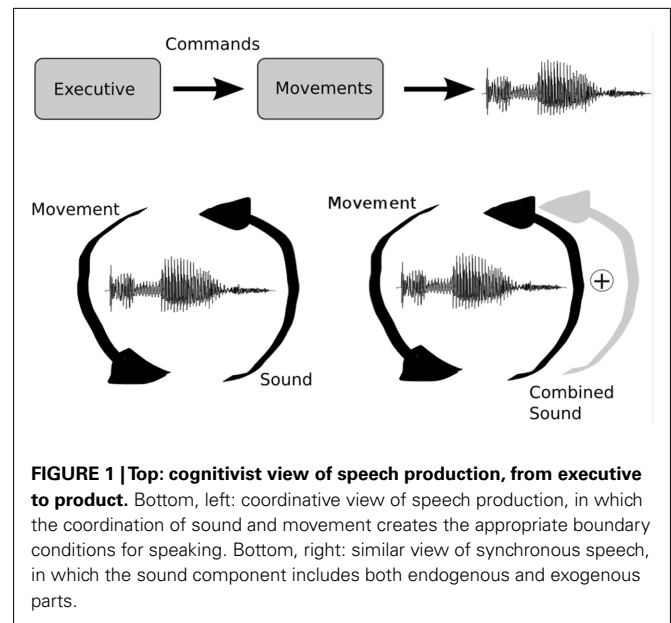
For Dewey, the sensorimotor coordination is the overarching domain within which we can make sense of perception and of action, but each must be interpreted with respect to the coordination:

In other words, the real beginning is with the act of seeing; it is looking, and not a sensation of light. The sensory quale gives the value of the act, just as the movement furnishes its mechanism and control, but both sensation and movement lie inside, not outside the act.

(Dewey, 1896, p. 358/359)

We can contrast a cognitivist and a dynamical view of what it is to speak. On the former, cognitivist, view, there is a speech production system and a speech perception system. These may share representational resources (Lieberman and Mattingly, 1985; Goldstein et al., 2006), but their functional separation is a starting point for developing an account of what speech is. Within the speech production system so conceived, an intention is formed by a notional executive, that becomes the basis for commands emanating from the center to the periphery, resulting in movement, that in turn results in speech sound. This efficient causal chain constitutes the speech production system. This chain is illustrated in the top row of **Figure 1**.

Alternatively, in the spirit of Dewey, one might suggest that speech might be understood as skilled action in which both the movement and the sound are subsumed and interpretable within an overarching coordinative framework, which is the skilled act of speaking. The present author has been speaking for almost 50 years. In that time, *every single utterance* (conceived of as movement) was accompanied (not followed) by speech sounds. Speech sounds, and speech articulations co-occur; they are not in sequence. The coordinative framework proposed here suggests that that skill is manifested in the co-registration of movement and sound. To be a skilled speaker is to align movement and sound such that speech results. As with moves within a tango dance, or obscene gestures, speech sounds are not natural kinds, and we can not look for purely naturalistic criteria for distinguishing speech from non-speech. Rather, speech is a skill, and the nature of that skill emerges within a group of speakers, in the collective acceptance and preference for some sequences of sounds over others.



When a baby babbles, toying with the various combinatorial and coordinative possibilities afforded by its vocal tract, it is learning that certain sounds and certain movements go together, not that some movements are followed by specific sounds. When it then alights upon successful coordinations that produce speech-like syllables, it has brought into being a higher-order coordinative domain defined and constituted by the mutual relations of sound and movement. The criteria that distinguish felicitous coordinations from mere uninterpretable babbling will be found both in the infant's own view of similarity to speech sounds occurring in its environment, and in the differential reaction of its environment to the speech it utters, reinforcing some coordinations, and ignoring others. On this view, to be a skilled speaker is to exhibit mastery over the sensorimotor contingencies of speaking, keeping the mutual relations of movement and sound such that the criteria of skilled performance are met. It is the nature of the skill that provides the boundary conditions that serve to delimit the skill from non-skill, and the boundaries separating speech from non-speech will become more clearly defined with practice, though they remain at all times somewhat plastic and sensitive to the contingencies of the context within which speech occurs.

This second view of what the act of speaking is, is illustrated in the bottom left of **Figure 1**. The skilled action of speaking is seen as a tight synergistic alignment of movement and sound, with neither one being temporally ordered before the other. Before progressing to consider the more complex case of synchronous speech, the coordinative view of skilled action being developed here may serve to shed light upon some other known characteristics of speech production. Articulatory movements and speech sounds always co-occur, and skilled speaking is the felicitous creation and maintenance of this tight reciprocal coupling of the two. It is immediately apparent then why delayed auditory feedback (DAF) should be so destructive of fluent speech production (Yates, 1963). By time-shifting the sound component, the very conditions that instantiate the act of speaking are removed. It is just

as if one were to time-delay the movements of one dancer in a tango-dancing pair. Disastrous miscoordination and a fall would surely result.

We might note that speaking without making use of the speech sound is possible, and can itself be considered a skill, as shown by the relative inefficacy of DAF in disrupting the speech of experienced polyglot simultaneous interpreters (Fabbro and Daro, 1995). This might be compared to the skill of touching one's nose with closed eyes. In both cases, the sensorimotor coordination is constituted by a suitably constrained relationship between movement and proprioception, instead of between movement and sound/vision (+ proprioception). From this perspective, it may seem less mysterious that playing loud noise at a stutterer can help to overcome stuttering: it effectively changes the task of speaking from the mutual coordination of movement and sound, to the coordination of movement and proprioception. This is, of course, by no means a full explanation, but rather a descriptive framework within which the properties of sensorimotor coordination may admit of novel forms of description.

It is precisely this shift in explanatory scaffolding, or framework, that will now be employed in reconsideration of synchronous speech, and subsequently, of sensorimotor synchronization.

4. SYNCHRONIZATION RECONSIDERED

The bottom right panel of **Figure 1** provides a way of understanding synchronization among speakers, without making reference to a periodic framework. In this view, the speaking act still consists of the alignment of movement and sound, but the sound employed is now a fusion or superposition of two distinct signals: An endogenously generated signal, as before, and an exogenously generated signal, that stems from the co-speaker. These two signals together constitute the sensory arc of the sensorimotor coordination that is speaking. In this respect, the co-speaker, through their speech, co-creates the framework that instantiates the act of speaking of the first speaker, or, to put it slightly differently, the speech of the other becomes formally incorporated into the condition of speaking of the first.

An analogy of a three-legged race was provided before. Under the present approach, the constraint, or tie, that binds the two speakers consists in the fact that each of them integrates the speech of the other within their respective sensorimotor coordination that is the act of speaking. It has been demonstrated that, although speakers can synchronize to some extent with a recording of another person, and the degree of synchrony is greater if the recording is of someone who was recorded while themselves speaking synchronously, there is still an advantage for the situation in which two speakers interact in real time (Cummins, 2009b). This makes sense if each speaker is, in a very real sense, constraining the other in this way.

Interestingly, speaking in unison with others is also an effective way of overcoming stuttering in many cases (Kalinowski and Saltuklaroglu, 2003). A conventional account of this would claim that the stutterer is imitating the speech of co-speakers, perhaps facilitated by the mirror neuron system, or similar. The account that suggests itself in the present context is simpler. The speech of others is playing a role in stabilizing the coordinative domain that is constitutive of the act of speaking. When speaking alone, the

stutterer is not successful at establishing the required coordination of movement and sound, but by augmenting the endogenous signal with an intact exogenous signal, speech is made possible.

Instead of providing an account of synchronized behavior in which periodicity provides a system-external temporal referent with respect to which actions might be timed, the present view suggests that the common skill that two competent speakers have acquired forms the basis for the phenomenon of synchronization. On this view, to have the skill of speaking is precisely the capacity to co-align movement and sound in a manner that is sufficient to be accepted as speech by both the speaker and her environment. But that alone is not enough to account for the synchronization observed. To this we must add the remarkable ability of replacing the endogenously produced sound with a composite signal that is a fusion of an endogenous and an exogenous sound. That is, indeed, a significant achievement, but it is by no means without precedent.

For this is surely precisely what is demonstrated in the simpler case of sensorimotor synchronization. Let us now consider what it is to bob one's head, torso, or foot along with some music. Any oscillatory movement of the body generates sensory variation that is periodic. This is entirely modality independent, so that proprioceptive, visual, and auditory variation all provide the same information. If one regards bobbing/tapping/bouncing as an extremely rudimentary skill (shared by humans and almost all animals), then the exercise of that skill lies precisely in constraining movement such that the attendant sensory flux varies in periodic fashion, or, equivalently, such that the rate of change is approximately constant. What humans seem to have added, somewhere in the time since our evolutionary path last diverged from that of the other great apes, is the remarkable ability to employ a fusion of an endogenously generated sensory flux and an exogenously generated pattern of variation, so that they collectively function as the sensory arc of the sensorimotor coordination that is rhythmic bobbing. In this way, a group of people dancing to music will be bound together in their movement, not only through the music, but through the visual registration of phase in the movements of the others. Rhythm, I have claimed elsewhere, may fruitfully be defined as an affordance for the entrainment of movement (Cummins, 2009a). This notion can now be fleshed out through the unifying concept of the sensorimotor coordination, and it becomes immediately apparent why rhythm should be such a potent force in getting people to move together.

5. DISCUSSION

This article offered synchronization among concurrent speakers as a puzzle. It is a puzzle, because the phenomenon of synchronization in behavior is typically addressed using a modeling framework based on oscillatory processes. Periodicity is a striking facet of most behaviors that we might describe as synchronized, and the mathematics of interacting oscillators provides a convenient and powerful framework for quantitative and qualitative modeling. And yet this approach seems to rule out any account of aperiodic synchronization, generating a strict division between two kinds of phenomena, which otherwise seem to have much in common. It also offers no help at all in understanding how we might presume to model aperiodic synchronization.

One-way to re-unite both periodic and aperiodic synchronized behaviors is made possible by recognizing that in each case, movement and sensory change are highly constrained by the nature of the task. The task of rhythmic bobbing or toe tapping is so simple as to barely merit the application of the term “skill,” for simple periodic oscillation is an ability we share with almost all living creatures. But if we cast it as a skill, defined by the constraint of movement such that it accompanies a constant rate of change in sensory flux, it then becomes apparent that our species-specific ability to track the fluctuating tempo of an external source of rhythm is more than an idiosyncratic party trick. When people gather to make music together, or to dance in groups, each participant is creating and maintaining a sensorimotor coordination in which one arc is generated collectively. This ability to share a skill is what we have, and the jellyfish does not.

Sensorimotor synchronization allows the tight synchronization of action in marching, rowing, and in collective heaving or pulling, as in a tug of war. In all of these instances, simple periodicity in the behavior ensures that the auditory information shared among participants aligns with visual and proprioceptive information, providing robust multimodal support for action timing. In synchronous speaking, we see that this ability to collectively bring about synchronous skilled behavior can extend to complex, time varying behavior as well, just in case the individuals share a sufficiently constrained definition of the skill in question. This is, perhaps uniquely, the case for speech.

5.1. COMMON CODE AND IDEOMOTOR THEORIES

The linear sequence of the cognitivist orthodoxy that starts with sensation/perception and results in intentional action has been under attack from many quarters. One of the oldest and most severe criticisms comes from the recognition that perception and action are inextricably linked, and must be understood together. William James observed that:

Every representation of a movement awakens in some degree the actual movement which is its object; and awakens it in a maximum degree whenever it is not kept from doing so by an antagonistic representation present simultaneously in the mind

(James, 1890, vol. II, p. 526).

This view has found significant elaboration in the common code theory of Prinz (1984, 2005). On this view, action and perception are inextricably linked because they make use of a common representational substratum. The perceptual effects of actions give rise to the neural representation of those actions, with the result that observation of action activates the same representations that would be employed in carrying out the same acts. This theoretical approach is buttressed by many experimental observations of the effect of action on perception (Kilner et al., 2003; Wilson and Knoblich, 2005). A rather specific form of common code theory has been mooted in the speech domain with the proposition that speech perception and speech production employ common representations (Lieberman and Mattingly, 1985). An even closer relation between the production and perception of speech is implied by the findings of Fadiga et al. (2002) that listening to speech can be shown to elicit subliminal activation of the articulators that would

be used to produce that speech. This finding, obtained using transcranial magnetic stimulation, suggests that the listener literally resonates to the speech being heard.

The momentum accorded to theories of common representation has increased enormously since the discovery of so-called mirror neurons in monkey brains (Rizzolatti and Craighero, 2004). These neurons, as is all too well known, are found to fire preferentially both in carrying out specific goal-directed actions, such as grasping, and in observing someone else carrying out the same action. The import of this discovery will depend greatly on the view of perception and action adopted.

On the cognitivist view for which perception and action are separate, this is indeed extremely strong evidence that a common representational basis might play a role in linking the two distinct functions. Common representation has been mooted as a basis for understanding the actions of others, for empathy, and for more besides (Gallese, 2001).

On the view being pursued here, the discovery raises important questions about how a goal is to be defined, and what the circumstances are under which equivalence appears in neural activity when acting or observing action. Movement and sensory flux are uninterpretable without a superordinate domain with reference to which they can be understood. Likewise, neural activation will be uninterpretable without specification of the setting under which it is observed, and the formal inclusion of that specification in the apportioning of functional relevance to the activity. The importance of Dewey's approach lies in the freeing of functional explanation from the limits of a single domain, the cognitive system, with respect to which function is defined. Instead, a plurality of domains is established, each task-specific and context bound.

We are a long way from having any such account of neural activity. The present thesis suggests that the development of a theory of goals and skill in individual and in joint action is a precondition to understanding nervous system activity in intentional action.

5.2. COORDINATION DYNAMICS

The sensorimotor coordination posited by Dewey bears many similarities to the notion of a coordinative structure (Kelso et al., 1980; Kugler et al., 1980; or synergy Latash, 2008), as this concept is applied in the study of skilled action. An early and hugely significant observation was made by the Russian physiologist, Nikolai Bernstein, who found that skilled blacksmiths, when striking an anvil repeatedly, generated movement in which variability was minimized at the point of contact between anvil and hammer (Bernstein, 1967; Latash, 2008). In a link segment effector system, this observation rules out a puppeteer role for the central nervous system, as any errors or noise introduced closer to the center, say at the shoulder, would be amplified and added to further out, at elbow, wrist, and especially at the distal point of contact between anvil and hammer. Rather, he observed, the entire body and tool assemblage functioned as though it were a task-specific device, with vastly fewer degrees of freedom than the sum of its several components. When engaged in the task, a perturbation introduced at any one point was smoothly compensated for by other elements within the overall coordinative domain. Likewise, in speech production, it has been shown that a perturbation to the jaw administered in unpredictable fashion, will generate an almost

immediate compensatory response that is specific to the speech production goals existing at the time of the perturbation (Kelso et al., 1984). A downward thrust to the jaw during the second /b/ segment of /bab/ produced an almost immediate compensatory response in the upper lip, while an identical perturbation during /z/ production in /baz/ resulted in a compensatory response in the tongue body, appropriate for forming a /z/ articulation.

A coordinative structure, sometimes also called a synergy, is a task-specific, flexibly assembled system comprising parts of the body (or body + tools) that function together in the service of a well-defined behavioral goal. The concept is critical in understanding the role of the brain in movement, which is not, as popularly assumed, that of controlling the muscles individually, but rather of contributing to the constraining of movement in a task-specific fashion. The novel perspective provided by Dewey's notion of a sensorimotor coordination is to show how this coordinative domain is constituted, not just by structured movement, but by constraining the relation between movement and the attendant sensory changes.

5.3. GIBSON, ECOLOGICAL APPROACHES, AND SKILLED ACTION

Gibson (1966, 1979) developed a thoroughly radical stance in trying to understand both perception and action. Instead of worrying about what was going on in the head of a subject, Gibson looked at the lawfulness that inheres in the co-variation of movement and sensory flux within a specific environment. For very simple behavioral goals, this lawfulness can be approached analytically, as shown by the work of Lee and Reddish (1981). In the seminal work that introduced the concept of the tau variable to the visual control of action, they noted that the rate of expansion of the pattern of light and dark on the retina of an organism approaching a fixed surface is directly proportional to the time to contact between the two parties. This "information" lies in the mutual relation of organism and world, and does not require or benefit from the addition of notional cognitive mechanisms for extracting and processing the information.

Gibsonian approaches tease out the lawfulness that inheres in the embedding of an active organism in its environment. In this way, they provide an invaluable basis for developing an account of how movement and sensory flux co-vary in a task-specific manner. Gibsonian approaches run into explanatory limits, however, as the behavior to be characterized becomes more complex. Even in the canonical case of a diving gannet, the interpretation of expansion of the pattern of light and dark on the retina as information is only licensed by a framing understanding of the organism as an entity for whom a future collision is of systemic importance. For more complex behaviors, the informational relevance of any variable we care to measure will only be meaningful if interpreted within a context that defines and delimits the behavior itself. Speech is only speech by virtue of the conventions and habits of an entire speech community, and a full account of speaking will have to refer to such conventions if the skill is to be fully described.

5.4. THEORIES OF ENACTION

The coordinative focus adopted herein has been informed, in part, by the emerging enactive approach to mind and life, originating

in the biological theories of Maturana and Varela (1980), formulated in *The Embodied Mind* (Varela et al., 1991), and expanded upon more recently in Thompson (2007) and Stewart et al. (2011). Within this somewhat heterogeneous body of work, an emphasis is placed on the mutuality that exists between organism and world. If we regard an organism as an autonomous entity with a great number of constitutive degrees of freedom, then it becomes clear that such a system can potentially take part in coordinations that span the borders between an individual body and its environment. Indeed, the active embodied interaction of an organism with its surround is seen as the very basis for the emergence of mind.

Within the enactive tradition, the same principles of autonomy and coordination can be applied from the level of the single cell, through multicellular organisms, up to the level of social organization. For social interaction, De Jaegher and Di Paolo (2007) have taken one of the explanatory pillars of enaction, sense-making, and extended it to what they call "participatory sense-making," by which they mean the collective establishment and maintenance of supra-personal domains of relative autonomy, the interactional domains, that arise in the reciprocal coordination of two or more individuals. Appealing to the dynamical concepts of coupling and entrainment, they propose the following definition of social interaction:

Social interaction is the regulated coupling between at least two autonomous agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the autonomy of the agents involved (though the latter's scope can be augmented or reduced).

(De Jaegher and Di Paolo, 2007, p. 8)

The present contribution extends this notion directly, by proposing that the delimitation of the constraints that define a task may allow us to apply the dynamical concepts of coupling among individuals with somewhat more rigor. In common with the participatory sense-making notion, I have here emphasized the mutuality that exists, in a very literal sense, when two skilled actors synchronize, temporarily bringing into being a higher-level domain of coordination that is the dyad.

6. CONCLUSION

The puzzle posed by synchronized speaking lies in the absence of a periodic referent suitable for generating temporal expectations in concurrent speakers. This apparent anomaly has motivated a fundamental reconsideration of the descriptive framework within which skilled action is viewed. The sensorimotor coordination, first suggested by Dewey (1896), has been employed to frame an account of skilled action that permits a unified treatment of periodic and aperiodic synchronization. On this account, skilled action lies in constraining the co-occurrence of movement and sensory variation such that the criteria that serve to define a skill are met. Given this framing, synchronization can be described as skilled action in which the sensory arc of a sensorimotor coordination is collectively constituted. In this way, the constraint can be made explicit that yokes together the actions of people moving collectively with common goal.

The approach may, perhaps, seem to be question begging, placing as it does most of the structure and complexity of behavior on the framing constraints that define a skill. But unless we take the objectionable step of claiming that behaviors are natural kinds, this shift in focus seems necessary. The boundaries of speech are to be found in the conventions and practices of a community, not in the structure of the speech signal considered in isolation.

The emphasis on the context and constraints within which behavior happens may serve as one more contribution to a growing body of work that plays down hypothetical internal processes assumed to underlie behavior, emphasizing instead the totality of observables, from neural activity, through movement, to the context in which behavior is observed.

The framework advocated here could be seen as a radical break from information-processing accounts of cognitive functioning. Alternatively, it could be regarded as nothing more than a worked example of empirical behavioral research that does what most of us do most of the time: look for local solutions to local problems. By positing that movement and attendant sensory variation are

interpretable only within an appropriate frame of reference, we are doing nothing more outlandish than pointing out that kicking the ball into the net only constitutes a goal if it happens during a game of football. The principal contrast with cognitivist psychological theories lies in the context-specificity of any interpretation of movement, or of nervous system activity. Dynamical, coordinative approaches to understanding behavior do not refer to a single monolithic cognitive system. Rather, lawfulness may be found in task-specific contexts, within domains constituted by part of a person, a whole person, a person with a tool, or groups of people. The domains within which lawfulness is found will be emergent, and will be characterized by the generic dynamical principles of self-organization in complex systems.

ACKNOWLEDGMENTS

Thanks to Marek McGann, Sile O'Modhrain, Scott Kelso, and an anonymous reviewer for insightful feedback on earlier drafts and for the sustained discussion required to tease out the concepts presented herein.

REFERENCES

- Backwell, P., Jennions, M., Passmore, N., and Christy, J. (1998). Synchronized courtship in fiddler crabs. *Nature* 391, 31–32.
- Bernstein, N. (1967). *The Coordination and Regulation of Movements*. London: Pergamon Press.
- Classé, A. (1939). *The Rhythm of English Prose*. Oxford: Basil Blackwell.
- Copeland, J., and Moiseff, A. (1994). The occurrence of synchrony in the North American firefly *Photinus carolinus* (Coleoptera: Lampyridae). *J. Insect Behav.* 8, 381–394.
- Cummins, F. (2003). Practice and performance in speech produced synchronously. *J. Phon.* 31, 139–148.
- Cummins, F. (2009a). Rhythm as an affordance for the entrainment of movement. *Phonetica* 66, 15–28.
- Cummins, F. (2009b). Rhythm as entrainment: the case of synchronous speech. *J. Phon.* 37, 16–28.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *J. Phon.* 11, 51–62.
- De Jaegher, H., and Di Paolo, E. (2007). Participatory sense-making. *Phenomenol. Cogn. Sci.* 6, 485–507.
- Dewey, J. (1896). The reflex arc concept in psychology. *Psychol. Rev.* 3, 357.
- Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., and Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS ONE* 5, e12166. doi:10.1371/journal.pone.0012166
- Ermentrout, B. (1991). An adaptive model for synchrony in the firefly *Pteroptyx malaccae*. *J. Math. Biol.* 29, 571–585.
- Fabbro, F., and Daro, V. (1995). Delayed auditory feedback in polyglot simultaneous interpreters. *Brain Lang.* 48, 309–319.
- Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402.
- Gallese, V. (2001). The shared manifold hypothesis. From mirror neurons to empathy. *J. Conscious. Stud.* 8, 33–50.
- Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton Mifflin.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.
- Goldstein, L., Byrd, D., and Saltzman, E. (2006). “The role of vocal tract gestural action units in understanding the evolution of phonology,” in *Action to Language via the Mirror Neuron System*, ed. M. A. Arbib (Cambridge: Cambridge University Press), 215–249.
- Haken, H., Kelso, J. A. S., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movement. *Biol. Cybern.* 51, 347–356.
- Hanson, F., Case, J., Buck, E., and Buck, J. (1971). Synchrony and flash entrainment in a New Guinea firefly. *Science* 174, 161.
- Hurley, S. (2001). Perception and action: alternative views. *Synthese* 129, 3–40.
- James, W. (1890). *The Principles of Psychology*. New York: Henry Holt and Co.
- Kalinowski, J., and Saltuklaroglu, T. (2003). Choral speech: the amelioration of stuttering via imitation and the mirror neuronal system. *Neurosci. Biobehav. Rev.* 27, 339–347.
- Kelso, J., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 812.
- Kelso, J. A. S. (1981). “Contrasting perspectives on order and regulation in movement,” in *Attention and Performance, IX*, eds A. Baddeley and J. Long (Hillsdale, NJ: Erlbaum).
- Kelso, J. A. S. (1995). *Dynamic Patterns*. Cambridge, MA: MIT Press.
- Kelso, J. A. S., Holt, K. G., Kugler, P. N., and Turvey, M. (1980). “On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence,” in *Tutorials in Motor Behavior*, eds G. Stelmach and J. Requin (Amsterdam: North-Holland), 49–70.
- Kilner, J., Paulignan, Y., and Blakemore, S. (2003). An interference effect of observed biological movement on action. *Curr. Biol.* 13, 522–525.
- Kugler, P. N., Kelso, J. A. S., and Turvey, M. T. (1980). “On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence,” in *Tutorials in Motor Behavior*, eds G. Stelmach and J. Requin (Amsterdam: North-Holland), 1–40.
- Latash, M. (2008). *Synergy*. Oxford: Oxford University Press.
- Lee, D. N., and Reddish, P. E. (1981). Plummeting gannets: a paradigm of ecological optics. *Nature* 293, 293–294.
- Lehiste, I. (1977). Isochrony reconsidered. *J. Phon.* 5, 253–263.
- Lieberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36.
- MacDougall, H., and Moore, S. (2005). Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *J. Appl. Physiol.* 99, 1164.
- Maturana, H., and Varela, F. (1980). *Autopoiesis and Cognition: The Realization of the Living*, Vol. 42, New York: Springer.
- Merker, B., Madison, G., and Eckerdall, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex* 45, 4–17.
- Norton, A. (1995). “Dynamics: an introduction,” in *Mind as Motion: Explorations in the Dynamics of Cognition*, Chap. 1, eds R. F. Port and T. van Gelder (Cambridge, MA: Bradford Books/MIT Press), 45–68.
- Patel, A., Iversen, J., Bregman, M., and Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19, 827–830.
- Pikovsky, A., Rosenblum, M., and Kurths, J. (2001). *Synchronization: A Universal Concept in Nonlinear Sciences. Number 12 in Cambridge Nonlinear Science Series*. Cambridge: Cambridge University Press.
- Port, R., and van Gelder, T. (eds). (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: Bradford Books/MIT Press.
- Prinz, W. (1984). “Modes of linkage between perception and action,” in *Cognition and Motor Processes*, eds W. Prinz, A. F. Sanders, and D. A. Allport (New York: Springer), 185–193.

- Prinz, W. (2005). *An Ideomotor Approach to Imitation*. Cambridge, MA: The MIT Press.
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
- Sismondo, E. (1990). Synchronous, alternating, and phase-locked stridulation by a tropical katydid. *Science* 249, 55.
- Stewart, J., Gapenne, O., and Di Paolo, E. (2011). *Enaction: Toward a New Paradigm for Cognitive Science*. Cambridge, MA: The MIT Press.
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Belknap Press.
- Tognoli, E., Lagarde, J., DeGuzman, G., and Kelso, J. (2007). The phi complex as a neuromarker of human social coordination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8190.
- Varela, F., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: The MIT Press.
- Wilson, M., and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychol. Bull.* 131, 460.
- Wing, A. M., and Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Percept. Psychophys.* 14, 5–12.
- Yates, A. (1963). Delayed auditory feedback. *Psychol. Bull.* 60, 213.
- Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 18 October 2011; accepted: 12 December 2011; published online: 30 December 2011.
- Citation: Cummins F (2011) Periodic and aperiodic synchronization in skilled action. *Front. Hum. Neurosci.* 5:170. doi: 10.3389/fnhum.2011.00170
- Copyright © 2011 Cummins. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.