4

Synchrony and swing in conversation: coordination, temporal dynamics, and communication

Daniel C. Richardson, Rick Dale, and Kevin Shockley

4.1 Introduction

If you watch footage of a jazz quintet, even with the sound turned off, you will witness synchrony, swing, and coordination at multiple levels. At the time scale of a bar, gestures are locked to the rhythm of the music. At the time scale of a musical phrase, perhaps as part of a breath of air, there are sways of the body. All the while expressions and body language reflect emotive content and nuances of expression. Are the musicians' movements an outcome of the physical act of producing music, and their synchrony an epiphenomenon of following the same score? Or conversely, do the actions coordinate the music at one level, and shape its expression at another?

In conversation too, there are multiple levels of coordination. When two people exchange words, they share a great many things besides. For example, conversants will spontaneously converge upon dialect (Giles 1973), speaking rate (Street 1984), vocal intensity (Natale 1975), and pausing frequency (Capella and Planalp 1981; see Giles *et al.* 1991, for a review). Even without interacting with them, people will spontaneously imitate the speech patterns of others (Goldinger 1998; Shockley *et al.* 2004). But conversational partners do not limit their behavioral coordination to speech. They spontaneously move in synchrony with each other's speech rhythms (Condon 1976; Condon and Ogston 1971; Newtson 1994) and match one another's postures (Condon and Ogston 1966; Kendon 1970; Shockley *et al.* 2003). LaFrance (1982), for example, demonstrated that listeners tend to mirror a speaker's posture whom they find engaging. Imitation can be found throughout human interaction: neonates imitate facial gestures (e.g. Meltzoff and Moore 1983), infants imitate vocalic sounds (Kuhl and Meltzoff 1996), and adults spontaneously imitate facial expressions (McHugo *et al.* 1985; Sebanz and Knoblich, this volume).

In both the case of the musicians and of the conversants we may ask, what is the function of these multiple levels of coordination? How are they organized and whom do they benefit? Our questions focus on the temporal dynamics of conversation. In other words, why is it important that conversants not only do the same thing, but do so at the same time?

Historically, interpersonal coordination has been quantified subjectively (e.g. by hand scoring videotapes of listener movements and hand marking the accompanying speech for its rhythmic properties, Condon and Ogston 1971; cf. Newtson *et al.* 1977, 1987). In this chapter, we describe a recent analytical innovation called recurrence quantification analysis, used extensively in the biological and physical sciences (Webber and Zbilut 2005; Marwan *et al.* 2007). Relatively new to psychologists, this mathematical tool can reveal the characteristics of behavioral coupling. Going beyond subjective analysis of coordination, these tools allow interpersonal coordination to be objectively quantified, while capturing the temporal dynamics of cognition and action in a way that is of increasing interest to cognitive scientists.

In what follows, we first review research showing that within an individual, thinking and action have an interactive, dynamic relationship. This implies that when two or more such individuals who are engaged in conversation, there will be a rich interplay between language processes and outward action. While mental processes may be "private" to each conversant, their diverse and overt behaviors are shared. These shared behaviors exhibit substantial temporal coordination between conversants, and we present recurrence quantification analysis as a means to quantify it. We apply this technique to coordination in two domains: postural sway and eye movements during conversation. We conclude by arguing that these rich patterns of behavioral coordination likely reflect coordination of the underlying cognitive states and processes guiding conversation.

4.2 **Continuity between cognition and action**

Consider the relationship between thinking and action in an individual person. Though there is a venerable opinion that the mind and body are quite distinct things, the past 20 years of cognitive science have in fact shown the opposite. The idea that mental processes are closely related to bodily systems has developed in two schools of thought. The first sees cognition as inherently involving information about the perceptual and motor characteristics of the body (e.g. Ballard *et al.* 1997; Barsalou 1999; Clark A 1997; Dreyfus 1972, 1992; Glenberg and Robertson 2000; Lakoff and Johnson 1999; Rizzolatti *et al.* 1987; Spivey *et al.* in press; Varela *et al.* 1991). The second characterizes cognition in terms of continuous dynamical systems (e.g. Kelso 1995; Port and Van Gelder 1995; Spencer and Schoener 2003; Spivey 2007; Thelen and Smith 1994; Van Orden *et al.* 2003). States of mental processing smoothly transition from one to the next, much like trajectories in a high-dimensional state space of a continuous dynamical system.

Both of these perspectives predict a dynamic interchange between action and cognition. For example, motor outputs do not simply reflect the discrete decisions handed down from cognition: they covary with cognitive processes. The force and velocity of manual responses varies with word frequency in a lexical decision task (Abrams and Balota 1991; Balota and Abrams 1995), and response and stimulus probability in simple reaction-time tasks (Mattes *et al.* 2002; Ulrich *et al.* 1999; see also Osman *et al.* 1986; Balota *et al.* 1989). When reaching for a target object the arm does not always proceed in a ballistic fashion. Graspable distractors around the target can modulate cognitive processes which, in turn, tug at the trajectory of the hand (see also Gentilucci *et al.* 2000; Goodale *et al.* 1986; Sheliga *et al.* 1997; Tipper *et al.* 1997).

By tracking manual output in the form of computer-mouse trajectories, recent work has shown that complex cognitive processes involve continuous temporal dynamics. These tasks have included spoken-word recognition (Spivey *et al.* 2005), sentence processing (Farmer *et al.* 2008), and even reasoning (McKinstry *et al.* 2008). Dale, Kehoe, and Spivey (2007) analyzed computer-mouse trajectories during categorization of animal exemplars. Participants categorized an animal by clicking the mouse on one of two category choices. Mouse-movement trajectories consisted of a movement from the bottom center of the screen, to the correct target on the upper left- or right-hand corners of the screen (then clicking a category label). Target trials used atypical animals (e.g. whale) with an incorrect competitor category that had considerable overlap in terms of semantic features (fish). Though participants responded by clicking the appropriate category (mammal), mouse-movement trajectories exhibited substantial attraction toward the competitor categories.

Recent neurophysiological evidence substantiates a dynamic interchange between cognition and action. The dynamics of action systems, from premotor cortex (see Kalaska *et al.* 1997, for a review) into limb movements (e.g. Tipper *et al.* 1997), seem to be richly intertwined with cognitive processing (see Caminiti *et al.* 1998 and Kalaska *et al.* 1997, for excellent and concise reviews). As evidence for this, motor programs are not simply a collapse of a completed decision process, but rather are continuously updated by the accumulation of a cognitive decision (Gold and Shadlen 2000). Premotor and motor systems for reaching appear to be complex and integrative, and unfold continuously with simultaneous competition among possible responses. In one example, Cisek and Kalaska (2005) tracked nerve cell firing in premotor cortex in a reaching task with two possible choices in different directions. When monkeys were not yet signaled as to which reaching action was needed, a collection of cells maintained a level of activation for both possible reaches. Taken together, these neural and behavioral findings are a compelling demonstration that the continuous dynamics of action contain real-time indices of unfolding cognitive and perceptual processing.

A conversation consists of an elaborate sequence of actions—speaking, gesturing, maintaining the correct body language—which conversants must carefully select and time with respect to one another. The continuous interplay between cognition and action in an individual scales up to a complex and coordinative interplay between cognition-action systems of conversants. In our three laboratories, we have made use of a novel nonlinear analytic technique that can quantify such interpersonal coordination. In the following section, we supply a basic description of how this mathematical technique works.

4.2.1 Recurrence analysis

If cognition and action are interwoven, if decision making is a flux of response planning, how could this be seen in experiments that take the single data point of a reaction time and average across trials and subjects (Carello and Moreno 2005)? In recognition of this limitation, recently psychologists have been using a new analytical tool adapted from the biological and physical sciences. Recurrence analysis[1] is a simple but powerful technique

---

[1] This technique is also referred to as *recurrence quantification analysis, and cross recurrence quantification analysis* for analyzing the coupling between two time series.

that extracts the temporal structure in noisy, coupled dynamical systems. It consists of two basic steps. The first involves calculating the points in time that a system revisits similar states, called recurrences. The second is a quantification of those revisitations.

The first step emerged from a need to provide systematic descriptions of dynamical systems. There exist a variety of analyses that compute parameters characterizing a system's behavior, such as more or less chaotic or deterministic trajectories (e.g. the Lyapunov exponent; see Broer *et al.* 2001, for a recent volume on this and other measures). One limiting factor on these (often simply theoretical) parameters is that they are subject to sometimes unrealistic assumptions regarding the time series available to compute them (e.g. stationarity or extensive length; see Eckmann *et al.* 1987 for a discussion). To overcome such limitations and supplement these measures, Eckmann *et al.* (1987) devised a powerful but simple two-dimensional visualization technique that can also reveal characteristics of a system's dynamics. The technique is free from the assumptions more sophisticated analyses require (see also Webber and Zbilut 1994; Webber and Zbilut 2005). The goal of this technique is simple: to provide a two-dimensional plot whose points represent points in time that a system shows similar patterns of change or movement. These points are called "recurrences" or "recurrent points."

The basics of this first visualization step are quite straightforward. Consider a time series of numeric measurements x_t, with $t = 1,\ldots, N$. An ordered sequence of vectors or "windows" of size m can be constructed from this time series,[2] referred to as the "embedded" time series, $\xi\{x_t\}$

$$\xi\{x_t\} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{N-m+1}\}, \text{ where } \mathbf{x}_i = (x_i, \ldots, x_{i+m-1})$$

As an example, consider the following times series of random integers, and its corresponding embedding when m = 3:

$$x_t = 1, 5, 4, 3, 5, 2, 3\ 1 \ldots$$
$$\xi\{x_t\} = (1\ 5\ 4), (5\ 4\ 3), (4\ 3\ 5), (3\ 5\ 2), \ldots$$

By comparing each pair of vectors in the embedded time series, a recurrence plot (RP) is constructed out of the points (i, j) when the ith and jth indexed values of the embedded time series are sufficiently "close" or similar. An RP is therefore a set of time points (i, j) that visualizes how the dynamical system is revisiting certain paths in the system's trajectory.

$$RP = \{(i, j) \mid d(\mathbf{x}_i, \mathbf{x}_j) < \varepsilon\}, \text{ where } \mathbf{x}_i, \mathbf{x}_j \in \xi\{x_t\}$$

In the above equation, d is a distance measure, for example Euclidean distance, and ε a threshold or "radius" specifying how close two vectors must be to register a point (i, j) to the plot. An RP can have widely varying features depending on its source time series. Figure 4.1 illustrates some plots, revealing what Eckmann *et al.* (1987) originally referred to as differing "textures." Figure 4.1C presents a cross-recurrence plot (CRP), a simple

---

[2] For simplicity, we omit discussion of the additional parameter of lag (see Webber and Zbilut (2005) for an excellent introduction to these methods).

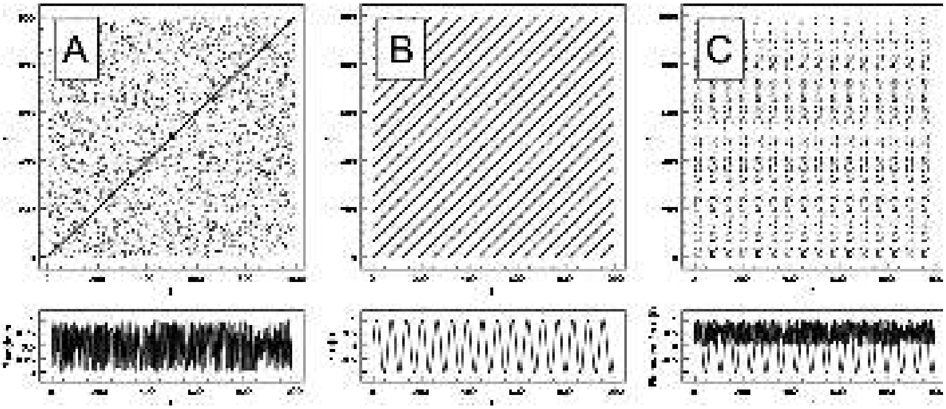

Figure 4.1 Panels A–C show a time series (bottom section) and a RP or CRP based on them (upper section). (A) A RP for a time series of 1000 random numbers between 0 and 1. Points on this plot represent times (i, j) at which the random numbers are revisiting similar patterns (using a window size of m = 3, and threshold of $\varepsilon = 0.1$ (see text for more detail). This RP has a very low percentage recurrence (%REC), at only 0.4%. (B) An RP for a 1000 samples from a sine function. The time series has perfect regularity, and the points therefore line up perfecty along diagonal lines, showing that sequences of vectors continue to revisit similar states. This plot has a greater %REC, at 3.6%. (C) A CRP of the two time series. Any points represent times at which the sine wave and random numbers occupy similar states. The plot not longer exhibits symmetry, and there is little coupling occurring between these two signals (%REC = 0.5%).

extension of the above technique to compare recurrent vectors between two different time series (e.g. Shockley *et al.* 2002; Zbilut *et al.* 1998).

$$\mathrm{CRP} = \{(i, j) \mid d(\mathbf{x}_i,\mathbf{y}_j) < \varepsilon\}, \text{ where } \mathbf{x}_i \in \xi\{x_t\}, \mathbf{y}_i \in \xi\{y_t\}$$

This is the first step of recurrence analysis: compare all patterns (or windows) of change in a time series (or two time series), and draw points on a two-dimensional plot when these patterns of change are similar or near each other. The second step is a quantification of this set of time points: by quantifying the number and nature of recurrence points in a RP or CRP, we can extract measures that illuminate the recurrent structure of the underlying system (in the case of RPs) or the coordination between two systems (in the case of CRPs). For example, Figure 4.1A is a plot of uniform white noise between 0 and 1, and exhibits little structure beyond the line of identity (LOI; where $i = j$, and $d = 0 < \varepsilon$). Figure 4.1B is a plot of a time series drawn from a sine wave function, and contains highly regular structures in the form of diagonal lines (corresponding to the perfectly repeated undulations of the sinusoid). Zbilut and Webber (1992; Webber and Zbilut 1994) devised a supplementary technique called recurrence quantification analysis (RQA) consisting of a suite of measures extracted from RPs. The simplest example is percent recurrence (%REC), the percentage of points registered on the plot. This is computed by dividing the total points by the number of possible points:

$$\| \mathrm{RP} \| \, / \, (N - m + 1)^2.$$

Diagonal structures in a RP are also informative, indicating periods of high regularity where stretches of the time series are recurrent. In Figure 4.1B for example, all the points in the plot fall along diagonal structures, indicating the regularity of the sine wave itself.

This basic process of embedding a time series and subjecting that embedding to analysis is a means of manipulating data sequences used in a number of disciplines. Among others, these include molecular biology (Von Heijne 1987), natural language processing and computational linguistics (see Manning and Schütze 1999, for a review), and physiology (Webber and Zbilut 1994). We employ it here to directly quantify the behavioral coordination—in posture and eye movements—between two people who are communicating or cooperating with each other.

4.2.2 **Postural coordination**

Standing upright may appear at first glance to be a straightforward and, perhaps, uninteresting behavior. However, upright stance is actually a complex pattern of behavior involving continuous movement of the body. This continuous movement obtains from the requirement to balance a large mass (i.e. the body) over a relatively small surface of support (i.e. the feet) by configuring several joints (e.g. knees, ankles, waist, and neck) so as to keep the plumb line from the center of mass of the body within the extents of the feet (i.e. the base of support). Even during quiet stance—standing without performing other behaviors—the configuration of the body (e.g. muscle activations and joint configurations) must be constantly adjusted to accommodate the constantly changing mass distribution of the body (e.g. due to physiological processes such as breathing and heart rhythms or behaviors such as gesturing and reaching). The body sways within a range of approximately 4 cm and in a pattern that is quite irregular and, thus, unpredictable.

Postural sway is typically measured as a time series of the center of pressure using a force platform—a device that measures the forces acting upon it and computes the location of the average of the sum of forces upon a support surface—or by capturing the motion of the approximate center of mass (e.g. the waist) by using motion capture technology. Postural sway dynamics are known to be influenced by suprapostural tasks—tasks that are performed concurrently with standing upright—such as looking (Stoffregen *et al.* 2000) or reaching (Belen'kii *et al.* 1967; Feldman 1966). For example when one is required to focus on something while standing upright, the postural sway pattern tends to become more constrained, which facilitates the ocular stability required to focus (e.g. Stoffregen *et al.* 1999). Of significance to the present discussion, speaking and even breathing have also been shown to influence postural sway dynamics (Conrad and Schonle 1979; Dault *et al.* 2003; Jeong 1991; Rimmer *et al.* 1995; Yardley *et al.* 1999).

Shockley *et al.* (2003) investigated how postural sway activity is influenced by cooperative conversation. They asked standing participants to discuss similar cartoon pictures in order to discover the subtle differences across the two pictures. For example, in one picture pair, each picture had a person, but the person in each picture wore different clothing. Neither participant could see his/her partner's picture, so they had to discover the differences between their respective pictures by talking back and forth. The participants were permitted to look around and gesture freely. However, they were configured to either face each other or face away from each other while they either discussed their pictures with the other participant or with a confederate. In all conditions, each participant performed the task at the same time as the other participant and in the presence of the other participant (regardless of task partner) and the postural sway was always measured for each participant simultaneously (regardless of task partner). At issue was whether between-participant talk and/or between-participant visual contact fostered interpersonal postural coordination.

Shockley *et al.* used cross recurrence analysis to show that both the number of shared postural configurations (%REC, as described in the previous section) and the length of the longest parallel trajectory of the two postural sway patterns (longest diagonal line of recurrence points, MAXLINE) were greater between participants in a pair when the participants were performing the task with each other than when they were each performing the task with a confederate (see Figure 4.2). Surprisingly, it did not matter whether or not

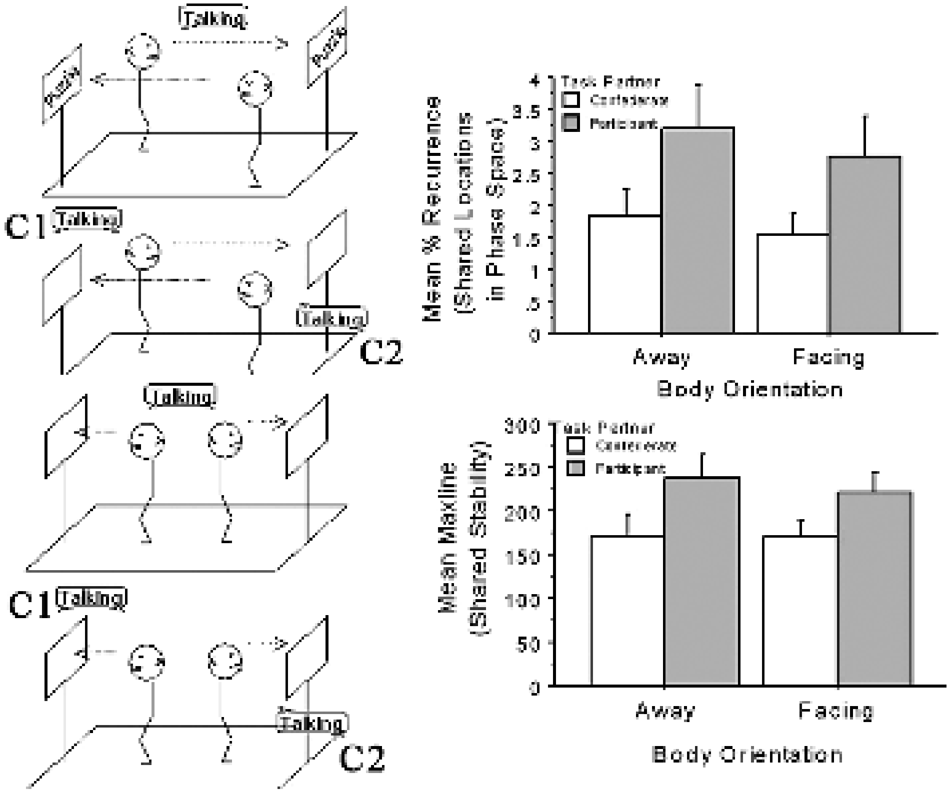

Figure 4.2 (left) Method used by Shockley *et al*. (2003). (right) Mean percent recurrence (%REC) and Maxline for the different experimental conditions. From Mutual interpersonal postural constraints are involved in cooperative conversation, by K Shockley, M-V Santana and C A Fowler 2003 *Journal of Experimental Psychology: Human Perception and Performance*, 29, p. 329 (panel A), 330 (panel C). Copyright 2003 by the American Psychological Association. Adapted with permission.

the two participants could see each other during the task. They concluded that language use serves as a coordination device. However, it raises the question of under what circumstances language use fosters interpersonal coordination. That is, how does discussing the content of two pictures, irrespective of whether one can see his/her partner, serve to entrain postural sway?

One possibility they considered was that participants were, by virtue of the task, required to talk about stimuli that were highly similar. Thus, participants may have produced utterances that had many words in common within a pair. This is significant because prior research has demonstrated that respiration (Conrad and Schonle 1979; Jeong 1991; Rimmer *et al.* 1995) and speaking (Dault *et al.* 2003; Yardley *et al.* 1999) influence postural sway patterns. At issue is whether the postural entrainment observed between conversational partners was mediated by common verbal content within participant pairs. Shockley *et al.* (2003) did not record the conversations and, therefore, could not evaluate this possibility. A phonetic property that could affect postural sway dynamics is stress pattern. For example the word *eth*nic is articulated with greater vocal effort on the first syllable than the second. In contrast, the word, de*serve* is articulated with greater vocal effort on the second syllable than the first. Given that respiratory/articulatory processes have been shown to influence postural sway patterns, one possibility is that utterances that entail greater vocal effort will influence postural sway more so than utterances that entail less vocal effort. Thus, if speakers converge in speaking patterns during cooperative conversation, they may also share the impact of those utterances on postural sway.

Shockley, Baker, MJ Richardson, and Fowler (2007) tested this possibility by having standing participant pairs (see Figure 4.3) either: (1) each utter the same words (S); (2) each utter different words that had the same stress emphasis pattern (DS) (e.g. one said *ethnic* when the other said *ancient*); or (3) each utter different words that had differing stress emphasis patterns (DD) (e.g. one said *ethnic* when the other said *deserve*). They also required participants to utter words either simultaneously or in an alternating fashion as a crude probe into the influence of conversational turn taking. They found no differences for any measures as a function of turn taking. As illustrated in Figure 4.3, however, they found greater shared postural activity when the words spoken within a pair had increasingly similar stress patterns. The implication was that the increase in shared postural activity with conversational partners was at least partially mediated by the similarity in speech patterns. In other words, because conversational partners tend to converge in speech patterns and speech has been shown to influence postural sway patterns, the shared postural activity observed by Shockley *et al.* (2003) may have been an indirect result of convergent speech patterns involved in cooperative conversation. Importantly, however, Shockley *et al.* (2007) found that the increase in shared postural activity could not be solely attributed to the biomechanical influences of convergent speech patterns. They performed a secondary analysis, this time pairing participants in the same experimental conditions, but who were not co-present. That is, they analyzed virtual pairs—pairs who were speaking the same word sequences in the same order, but pairs who did so in the presence of a different partner. Although members of a virtual pair did not perform the task simultaneously, their data were aligned with respect to the task (i.e. with respect

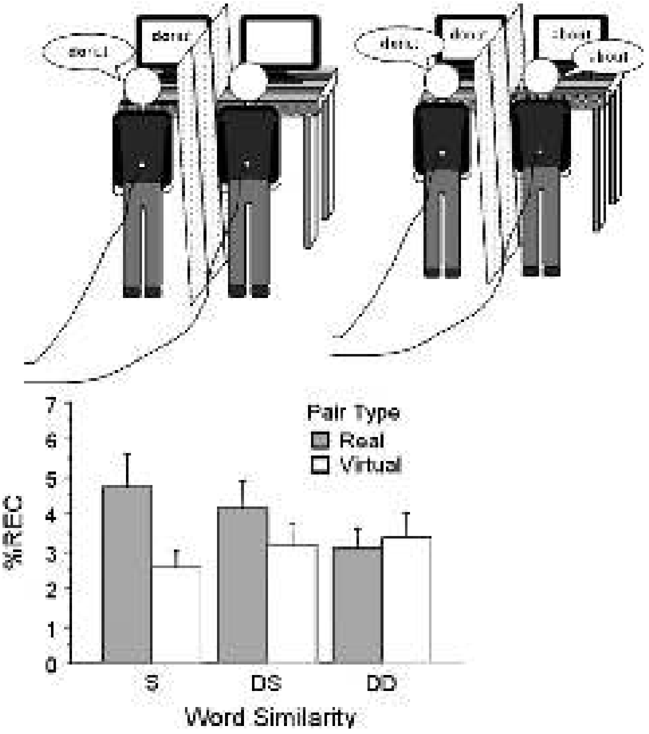

Figure 4.3 (Top) Method used by Shockley *et al.* (2007). (Bottom) Mean percent recurrence (%REC) for participant pairs and for "virtual pairs" (pairs who did not perform the task together) who uttered the same words (S), different words with the same stress patterning (SD), or different words with a different stress patterning (DD). From Articulatory constraints on interpersonal postural coordination, by K Shockley, A Baker, M J Richardson, and C A Fowler *Journal of Experimental Psychology: Human Perception and Performance*, 33, p. 203–205. Copyright 2007 by the American Psychological Association. Adapted with permission.

to word stimulus onset in their respective data collection sessions) via markers in the motion data that indicated onset of word stimuli. Thus, this secondary analysis permitted evaluation of how similarly the two postural time series unfolded during the course of the experimental task. As illustrated in Figure 4.3, they found that virtual pairs did not show increasing shared postural activity with increasing word similarity. They concluded that although speech similarity influences interpersonal postural coordination, the presence of another individual invites interpersonal coordination beyond that coordination resulting from utterance similarity (cf. Latané 1981; Zajonc 1965).

We have found that embodied conversations are spontaneously coordinated on multiple levels. These findings open up many intriguing questions. In particular conversational interactions, what factors predict the degree of postural entrainment? Attitudes of the conversants—whether they like each other or not—may influence levels of coordination. Many different goals play out in the course of a conversation, such as persuading, competing, and cooperating. Will these predict different types of coordination (cf. Giles 1973)?

Finally, does coordinated movement, likewise, foster effective communication (cf. Lakin and Chartrand 2003)? There is promise of interesting lessons to be learned from manipulating such factors. Until advances in virtual reality, it was impossible to exert precise experimental control over spontaneous, embodied interactions. Recently, Bailenson and colleagues have participants interact in a virtual world via digital version of themselves known as an "avatar". Bailenson and Yee (2005) introduced participants to another avatar who exactly mimicked their own head movements (with a delay introduced). Not only did participants rarely detect this mimicry, they found the mimic to be both persuasive and likable. Since coordination can be controlled, these methods could be extended to test a wide range of hypotheses investigating its behavioral consequences.

4.2.3 **Gaze coordination**

Consider an argument over a map, a debate over a proof written out on a black board, or a civilized conversation about a painting at a gallery. In all these cases, the stream of speech will be punctuated by hand waving and pointing to the shared visual scene, and perhaps even grabbing the map and turning it the right way up. Previously, we saw evidence that at one level, these physical movements will be coordinated during a conversation (Shockley *et al.* 2003). At another level, conversants also use such actions to influence each other's visual attention (Clark HH 1996). Here we use techniques from studying motor systems to quantify coordination in perceptual systems. We described research that tracks the gaze of two people while they look at an image and have a conversation.

The relationship between language use and visual attention has typically been studied by one of two approaches. One set of researchers have used eye-movement technology to explore the link between a speakers' eye movements and their language comprehension (e.g. Tanenhaus *et al.* 1995), and a listener's eye movements and their language production (e.g. Griffin and Bock 2000). The other set of researchers have studied interaction between participants and have focused on the actions they use to coordinate attention, such as gestures and pointing (Bangerter 2004; Clark HH and Krych 2004).

DC Richardson and Dale (2005) took a different tack in studying visual attention and language use. In contrast to the first approach, they did not track just one individual's eye movements, but recorded the eye movements of two participants while they discussed a shared visual scene. In contrast to the second set of researchers, they did not measure the actions participants make to coordinate attention, but measured the coordination of attention itself. Similar to the strategy used by Shockley *et al.* (2003, 2007) to quantify the temporal coupling between postural sway trajectories, they used cross recurrence analysis to quantify the temporal coupling between the conversants' eye movements.

This approach allowed a number of interesting questions. In this paradigm, the conversants cannot see each other, and hence cannot use pointing actions to coordinate their attention. Nevertheless, will their visual attention be coupled? Previous research has found reliable links between an individual's eye movements and their language comprehension and production in the case of short sentences (e.g. Griffin and Bock 2000; Tanenhaus *et al.* 1995). Will these results generalize to cases of extended, spontaneous

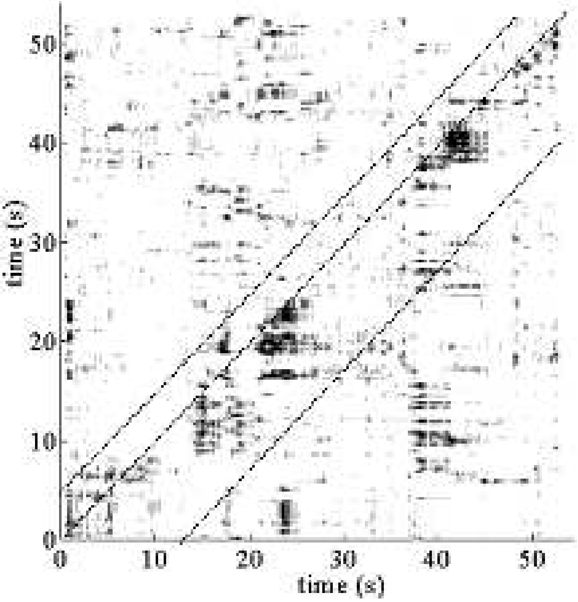

Figure 4.4 Composite cross recurrence plot of the eye movements of 49 speaker–listener dyads in Richardson and Dale (2005). Speakers' eye movements are along the y-axis, and listeners' along the x-axis. Dark dashed lines mark regions of analysis shown in Figure 4.5 (with the center dashed line representing the line of identity).

speech between two people? If so, what factors enable conversants to coordinate their visual attention by verbal means?

These questions were first addressed using a monologue version of the task. DC Richardson and Dale (2005) recorded the speech and eye movements of one set of participants as they looked at pictures of six cast members of a TV sitcom (either "Friends" or "The Simpsons"). They spoke spontaneously about their favorite episode and characters. One-minute segments were chosen and then played back unedited to a separate set of participants. The listeners looked at the same visual display of the cast members, and their eye movements were also recorded as they listened to the segments of speech. They then answered a series of comprehension questions. Recurrence analysis generated plots that quantified the degree to which speaker and listener eye positions overlapped at successive time lags.

Figure 4.4 shows a composite cross recurrence plot. The plots from all 49 of the speaker–listener pairs in our experiment were superimposed upon one another in grayscale. Recurrence at a particular time lag is shown by density along a particular x (speaker) = y (listener) + lag diagonal. This shows heaviest recurrence near to the line of incidence, representing the fact that speaker and listener eye movements were more coordinated when their time series were aligned within a few seconds of each other. There is little recurrence in the top left and bottom right regions of the plot. What the speaker was looking at during the start of the speech was not coordinated with what the listener

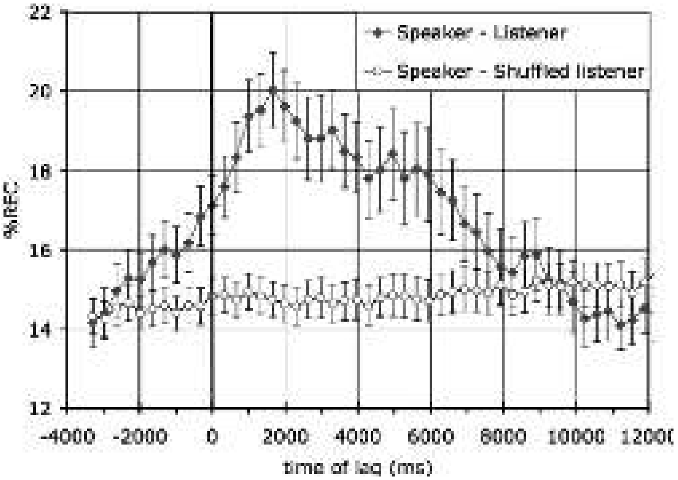

Figure 4.5 Speaker-listener gaze recurrence at different time lags, compared to a baseline of randomized listener eye movements (from Richardson and Dale 2005). See text for details.

was looking at towards the end, and vice versa. To examine the coordination more closely, we looked at the section of this graph where the speakers' gaze lagged the listeners by –4000 to 12000 ms. For this region, we computed the average recurrence at each time lag, in effect, calculating the density of recurrence along each diagonal in the cross recurrence plot (Figure 4.5). This speaker × listener distribution of fixations was compared to a speaker × randomized-listener distribution, produced by shuffling the temporal order of each listener's eye movement sequence and then calculating the cross recurrence with the speaker.

From the moment a speaker looks at a picture, and for the following 6 s, a listener was more likely than chance to be looking at that same picture (Figure 4.5). The overlap between speaker and listener eye movements peaked at about 2000 ms. In other words, 2 s after the speaker looked at a cast member, the listener was most likely to be looking at the same cast member. The timing of this peak roughly corresponds to results in the speech production and comprehension literatures. Speakers will fixate objects 800–1000 ms (Griffin and Bock 2000) before naming them, and listeners will typically take 500–1000 ms to fixate an object from the word onset (Allopenna *et al.* 1998). Planning diverse types of speech appears to systematically influence the speaker's eye movements, and, a few seconds later, hearing them will influence the listener's eye movements.

Importantly, this coupling of eye movements between speaker and listener was not merely an epiphenomenal by-product of conversation. The amount of recurrence between individual speaker–listener pairs reliably predicted how many of the comprehension questions the listener answered correctly. This correlation was supported by a follow-up study that experimentally manipulated the relationship between speaker and listener eye movements. We found that by flashing the pictures in time with the speakers' fixations (or a randomized version) we caused the listeners' eye movements to look more (or less) like the speakers', and influenced the listeners' performance on comprehension questions.

Though the language use in DC Richardson and Dale's (2005) study was spontaneous, it lacked a key element of everyday conversations—interaction. In a second set of studies, DC Richardson, Dale, and Kirkham (2007) tracked the gaze of two conversants simultaneously while they discussed TV shows, politics, and surrealist paintings. In the case of a live, interactive dialogue, conversants' eye movements continued to be coupled as they looked at a shared visual display. This coupling peaked at a lag of 0 ms. In other words, the conversants were most likely to be looking at the same thing at the same point in time. As in the monologue results, this coupling was at above chance levels for a period of around 6 s, suggesting that conversants may keep track of a subset of the depicted people who are relevant moment-by-moment (Brown-Schmidt *et al.* 2004).

Cross recurrence analysis has revealed a close temporal coupling between conversants' eye gaze during both monologues and dialogues. The strength of this coupling appears to determine comprehension, in part. Gaze coordination occurs as part of communication, therefore, and plays a functional role in communication. One could argue that it is still remarkable that it happens at all, however, given the high frequency of eye movements (three or four per second) and the low bandwidth of speech (approximately one per second). For example, in DC Richardson and Dale (2005), at above chance levels silent listeners were at times looking at pictures a full second before they were mentioned or even fixated by the speakers. How was this achieved?

Part of the answer is that conversants' shared more than a stream of words. According to HH Clark (1996), conversation is only understandable against a backdrop of "common ground". This knowledge is shared between conversants, and allows speech to be interpreted despite ambiguity and indefiniteness. In the case of our experiments, it allowed listeners to anticipate which pictures were the speaker's impending focus of attention. In other words, it was the mutual knowledge of the characters in Friends or The Simpsons that allowed conversants' gaze to be so tightly coupled (anticipatorily and reactively). In the experiments described above, common ground knowledge was high: participants were excluded if they had never seen an episode of either sitcom (resulting in a 0% attrition rate). In the second experiment of DC Richardson *et al.* (2007) the level of common ground knowledge was experimentally manipulated, and its effect on gaze coordination assessed.

Participants were asked to talk about a relatively obscure painting by Salvador Dali. Before their conversation, they heard either the same or different discussions of Dali's art. Accordingly, the participants then listened to 90-s passages that related either the history, content, and meaning of the specific painting (e.g. "the still life objects in the original canvas have separated from the table and float in the air, and even the particles of paint have broken loose from the canvas"), or Dali's personality and theory (e.g. "the paranoiac critical method entailed the creation of a visionary reality from elements of dreams, memories and psychological or pathological distortions. At times Dali would stand on his head to induce hallucinations."). They then saw the painting and discussed it while their gaze was tracked. DC Richardson *et al.* (2007) found that conversational partners who heard the same information had 33% more eye-movement coordination than those who heard different information. Interestingly, it did not seem to matter which of the passages the conversants heard—the one about the painting or the one about the artist. What was important was that they had the same information and knew this to be the case.

In further studies, DC Richardson and Dale (in preparation) are investigating how such common ground information might be created between conversants. Participants took part in three rounds of the tangram matching task (Clark HH and Brennan 1991). They saw the same six abstract, humanoid shapes in different orders. One participant was instructed to describe his shapes in turn so that the other could find them. In the first round, participants typically established descriptors of the ambiguous shapes (e.g. "the dancer", "the skier"). This process of grounding and confirming descriptors is reflected in the eye-movement recurrence. Typically, eye-movement couplings increased during a trial until the matcher was fixating the right shape. At that point, a descriptor would be proposed. For the rest of the trial, the eye-movement coupling decreased as both director and matcher looked around at other shapes to see if the descriptor was a good one. In later rounds, these established "conceptual pacts" (Clark HH and Brennan 1991) provided a quicker way to find the shapes, and eye-movement recurrence peaked more quickly.

There is an interesting reciprocity between gaze couplings and HH Clark's (1996) notion of common ground. If conversants begin a conversation with more knowledge in common, they will find it easier to coordinate their gaze. If conversants are looking at the same thing at the same time, then their shared perceptual experience will boost their common ground. Lastly, as conversants generate their own common ground knowledge anew, their gaze to abstract tangram shapes becomes more tightly linked.

4.3 **Conclusion**

Spontaneous speech is messy. Spontaneous verbal interactions between people are messier still. In the face of this complexity, some language researchers have (quite rightly) simplified things, studying spoken conversations instead of face-to-face conversations, speech comprehension instead of interactive conversation, text instead of speech, and single word presentation instead of reading. In contrast to that approach, the experiments described here have embraced the complexity in communicative behavior. Here we have taken a pair of fluctuating, dynamic, and noisy signals—posture and gaze position—and used recurrence analysis to reveal an intimate temporal coupling between conversants.

While the two groups of studies have marked differences, they reveal very similar patterns of underlying coordination. For example, Shockley *et al.* (2007) showed that co-presence (Clark HH 1996) seems to be part of speech-driven postural coordination, and not only the specific rhythm of speech itself. At the same time, Richardson *et al.* (2007) find that conversants who share (visual) co-presence along with common ground information about a painting more strongly coordinate visual attention during conversation. Both lines of work suggest future studies that can explore the source of this coordination. What is it about co-presence and common ground that generates rich coordination of these low-level signals? One avenue may be unleashing this analytic technique of recurrence analysis onto multichannel time series—revealing how patterns from word usage, to postural control and eye movements, are intricately tied together into the coordinative processes of conversation.

Another possible direction is to reanalyze pre-existing transcripts of dialogue with these new modes of inquiry. For example, Dale and Spivey (2005, 2006) used very large sets of transcripts of child–caregiver interaction to reveal similar patterns of coordination in syntax used by conversation partners. In addition, by using child–caregiver corpora, they demonstrated that this coupling has a developmental trajectory (the younger the child, the stronger the coupling), and provided evidence that there may be subtle individual differences underlying who leads or follows this coupling (subtle leading by the linguistically advanced child). If coordination in conversation can be likened to a dance, or a jazz quintet, then recurrence analysis may unveil other interesting underlying characteristics of these patterns, such as who is leading and following (see Dale and Spivey 2006, for a description how recurrence analysis can do this).

The behavioral couplings in embodied conversation discussed in this chapter reveal an intimate relationship between discourse processes, visual attention, and motor control. We argue that studying disembodied language is like studying music only as notes on a stave. Whilst one can learn a lot about form and structure, no one reads sheet music for pleasure. The function of music is in its performance, its embodiment. When music is played, multiple levels of behavioral coordination emerge. Likewise, when words are spoken between two people. We argue that the linguistic, postural and attentional coordination that ensues is not a byproduct of the interaction. When conversants are co-present, they synchronize their sway, regardless of whether they can see each other, and independently of the words that are said. The coupling between conversants' eye movements reflects both the process and the success of their communication. We claim that in a precise and profound way, embodied conversations resemble what HH Clark (1996) described as the joint activity of language use.

Acknowledgements

Preparation of the manuscript was supported by a grant from the National Science Foundation to the University of Cincinnati (CMS-0432992).

References

Abrams RA and Balota DA (1991). Mental chronometry: Beyond reaction time. *Psychological Science*, **2**, 153–7.

Allopenna PD, Magnuson JS, and Tanenhaus MK (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, **38**, 419–39.

Bailenson JN and Yee N (2005). Digital chameleons. Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, **16**, 814–9.

Ballard DH, Hayhoe MM, Pook PK, and Rao RP (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, **20**, 723–42; discussion 743–67.

Balota DA and Abrams RA (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1289–302.

Balota DA, Boland JE, and Shields LW (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, **28**, 14–36.

Bangerter A (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, **15**, 415–9.

Barsalou LW (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, **22**, 577–660.

Belen'kii VY, Gurfinkel VS, and Pal'tsev YI (1967). Elements of control of voluntary movement. *Biophysics*, **12**, 154–61.

Broer H, Krauskopf B, and Vegter G (2001). *Global Analysis of Dynamical Systems: festschrift dedicated to Floris Takens*. Bristol, UK: Institute of Physics.

Brown-Schmidt S, Campana E, and Tanenhaus MK (2004). Real-time reference resolution by naïve participants during a task-based unscripted conversation. In JC Trueswell and MK Tanenhaus, eds. *World-Situated Language Processing: Bridging the language as product and language as action traditions*. Cambridge: MIT Press.

Caminiti R, Ferraina S, and Mayer AB (1998). Visuomotor transformations: Early cortical mechanisms of reaching. *Current Opinion in Neurobiology*, **8**, 753–61.

Capella J and Planalp S (1981). Talk and silence sequences in informal conversations, III: Interspeaker influence. *Human Communications Research*, **7**, 117–32.

Carello C and Moreno M (2005). Why nonlinear methods? In MA Riley and GC Van Orden, eds. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, pp. 1–25. Retrieved March 1 2005, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp

Cisek P and Kalaska JF (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron*, **45**, 801–14.

Clark A (1997). *Being There: Putting brain, body, and world together again*. Cambridge, MA, USA: MIT Press.

Clark HH (1996). *Using Language*. New York: Cambridge University Press.

Clark HH and Brennan SE (1991). Grounding in communication. In LB Resnick, JM Levine and SD Teasley, eds. *Perspectives on Socially Shared Cognition*, pp. 127–49. Washington, DC: APA.

Clark HH and Krych MA (2004). Speaking while monitoring addressees for understanding. J*ournal of Memory and Language*, **50**, 62–81.

Condon W (1976). An analysis of behavioral organization. *Sign Language Studies*, **13**, 285–318.

Condon W and Ogston W (1966). Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease*, **143**, 338–47.

Condon W and Ogston W (1971). Speech and body motion synchrony of the speaker-hearer. In D Horton and J Jenkins, eds. *The Perception of Language*, pp. 150–84. Columbus, OH: Charles E. Merrill.

Conrad B and Schonle P (1979) Speech and respiration. *Archiv fur Psychiatrie und Nervenkrankheiten*, **226**, 251–68.

Dale R, Kehoe C, and Spivey M (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory and Cognition*, **35**, 15–28.

Dale R and Spivey MJ (2005). Categorical recurrence analysis of child language. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, pp. 530–535. Mahwah, NJ: Lawrence Erlbaum.

Dale R and Spivey MJ (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, **56**, 391–430.

Dault MC, Yardley L, and Frank JS (2003). Does articulation contribute to modifications of postural control during dual-task performance? *Cognitive Brain Research*, **16**, 434–40.

Dreyfus HL (1972). *What Computers Can't Do: A critique of artificial reason*. New York: Harper and Row.

Dreyfus HL (1992). *What Computers Still Can't Do: A critique of artificial reason*. Cambridge, MA, USA: MIT Press.

Eckmann J-P, Kamphorst SO, and Ruelle D (1987). Recurrence plots of dynamical systems. *Europhysics Letters*, **5**, 973–7.

Farmer T, Cargill S, Hindy N, Dale R, and Spivey M (2008). Tracking the continuity of language comprehension: Computer-mouse trajectories suggest parallel syntactic processing. *Cognitive Science*, **31**, 889–909.

Feldman AG (1966). Functional tuning of the nervous system during control of movement or maintenance of a steady posture: III. Mechanographic analysis of the execution by man of the simplest motor tasks. *Biophysics*, **11**, 766–75.

Gentilucci M, Benuzzi F, Bertolani L, Daprati E and Gangitano M (2000). Language and motor control. *Experimental Brain Research*, **133**, 468–490.

Giles H (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, **15**, 87–105. Giles H,

Coupland N, and Coupland J (1991). Accommodation theory: Communication, context, and consequence. In H Giles, J Coupland and N Coupland, eds. *Contexts of Accommodation: Developments in applied sociolinguistics*, pp. 1–68. New York: Cambridge University Press.

Glenberg AM and Robertson DA (2000). Symbol grounding and meaning: A comparison of highdimensional and embodied theories of meaning. *Journal of Memory and Language*, **43**, 379–401.

Gold JI and Shadlen MN (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, **404**, 390–4.

Goldinger S (1998). Echoes of echoes: An episodic theory of lexical access. *Psychological Review*, **105**, 251–79.

Goodale MA, Pelisson D, and Prablanc C (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, **320**, 748–50.

Griffin ZM and Bock K (2000). What the eyes say about speaking. *Psychological Science*, **11**, 274–9. Jeong

BY (1991). Respiration effect on standing balance. *Archives of Physical Medicine and Rehabilitation,* ***72****, 642–5.*

Kalaska JF, Scott S, Cisek P, and Sergio LE (1997). Cortical control of reaching movements. *Current Opinion in Neurobiology*, **7**, 849–59.

Kelso JAS (1995). *Dynamic Patterns: the self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Kendon A (1970). Movement coordination in social interaction: Some examples. *Acta Psychologica*, **32**, 1–25.

Kuhl P and Meltzoff A (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, **100**, 2425–38.

LaFrance M (1982). Posture mirroring and rapport. In M Davis, ed. *Interaction Rhythms: periodicity in communicative behavior*, pp. 279–98. New York: Human Sciences Press.

Lakin JL and Chartrand TL (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, **14**, 334–9.

Lakoff G and Johnson M (1999). *Philosophy in the Flesh: THE embodied mind and its challenge to western thought*. New York: Basic Books.

Latané B (1981). The psychology of social impact. *American Psychologist*, **36**, 343–56.

Marwan N, Romano MC, Thiel M and Kurths J (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, **438**, 237–329.

Mattes S, Ulrich R, and Miller J (2002). Response force in RT tasks: Isolating effects of stimulus probability and response probability. *Visual Cognition*, **9**, 477–501.

McHugo G, Lanzetta J, Sullivan D, Masters R, and Englis B (1985). Emotional reactions to a political leader's expressive displays. *Journal of Personality and Social Psychology*, **49**, 1513–29.

McKinstry C, Dale R and Spivey MJ (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, **19**, 22–24.

Meltzoff A and Moore M (1983). Newborn infants imitate adult facial gestures. *Child Development*, **54**, 702–9.

Natale M (1975). Convergence of mean vocal intensity in dyadic communications as a function of social desirability. *Journal of Personality and Social Psychology*, **32**, 790–804.

Newtson D (1994). The perception and coupling of behavior waves. In RR Vallacher and A Nowak, eds. *Dynamical Systems in Social Psychology*, pp. 139–67. San Diego, CA: Academic Press.

Newtson D, Enquist G, and Bois J (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, **35**, 847–62.

Newtson D, Hairfield J, Bloomingdale J, and Cutino S (1987). The structure of action and interaction. *Social Cognition*, **5**, 191–237.

Osman A, Kornblum S, and Meyer DE (1986). The point of no return in choice reaction time: Controlled and ballistic stages of response preparation. *Journal of Experimental Psychology: Human Perception and Performance*, **12**, 243–58.

Port R and van Gelder T (1995). *Mind as Motion: explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

Richardson DC and Dale R (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, **29**, 1045–60.

Richardson DC, Dale R, and Kirkham NZ (2007). The art of conversation is coordination: common ground and the coupling of eye movements during dialogue. *Psychological Science*, **18**, 407–13.

Rimmer KP, Ford GT, and Whitelaw WA (1995). Interaction between postural and respiratory control of human intercostals muscles. *Journal of Applied Physiology*, **79**, 1556–61.

Rizzolatti G, Riggio L, Dascola I, and Umiltá C (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia* (Special Issue: Selective visual attention), **25**, 31–40.

Sheliga BM, Craighero L, Riggio L and Rizzolatti G (1997). Effects of spatial attention on directional manual and ocular responses. *Experimental Brain Research*, **114**, 339–351.

Shockley K, Baker AA, Richardson MJ, and Fowler CA (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, **33**, 201–8.

Shockley K, Butwill M, Zbilut J, and Webber C (2002). Cross recurrence quantification of coupled oscillators, *Physics Letters A*, **305**, 59–69.

Shockley K, Sabadini L, and Fowler CA (2004). Imitation in shadowing words. *Perception and Psychophysics*, **66**, 422–9.

Shockley K, Santana MV, and Fowler CA (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, **29**, 326–32.

Spencer JP and Schoener G (2003). Bridging the representational gap in the dynamical systems approach to development. *Developmental Science*, **6**, 392–412.

Spivey M (2007). *Continuity of Mind*. New York, NY: Oxford University Press.

Spivey MJ, Grosjean M, and Knoblich G (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, **102**, 10393–8.

Spivey M, Richardson DC, and Dale R (in press). The movement of eye and hand as a window into language and cognition. In E Morsella, P Gollwitzer, and J Bargh, eds. *The Psychology of Action*. New York, NY: Oxford University Press.

Stoffregen TA, Pagaulayan RJ, Bardy BG, and Hettinger LJ (2000). Modulating postural control to facilitate visual performance. *Journal of Experimental Psychology: Human Perception and Performance*, **19**, 203–20.

Stoffregen TA, Smart LJ, Bardy BG, and Pagulayan RJ (1999). Postural stabilization of looking. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 1641–58.

Street RL (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, **11**, 149–69.

Tanenhaus MK, Spivey Knowlton MJ, Eberhard KM, and Sedivy JC (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, **268**, 1632–4.

Thelen E and Smith LB (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: The MIT Press.

Tipper SP, Howard LA, and Jackson SR (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition*, **4**, 1–38.

Ulrich R, Mattes S, and Miller J (1999). Donders's assumption of pure insertion: An evaluation on the basis of response dynamics. *Acta Psychologica*, **102**, 43–75.

Van Orden GC, Holden JG, and Turvey MT (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, **132**, 331–50.

Varela FJ, Thompson E, and Rosch E (1991). *The Embodied Mind: cognitive science and human experience*. Cambridge, MA, US: MIT Press.

Von Heijne G (1987). *Sequence Analysis in Molecular Biology: treasure trove or trivial pursuits?* San Diego, CA: Academic Press.

Webber CL, Jr and Zbilut JP (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, **76**, 965–73.

Webber CL, Jr and Zbilut JP (2005). Recurrence quantification analysis of nonlinear dynamical systems. In MA Riley and GC Van Orden, eds. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, pp. 26–94. Retrieved March 1 2005, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp

Yardley L, Gardner M, Leadbetter A, and Lavie N (1999). Effect of articulation and mental tasks on postural control. *Neuroreport*, **10**, 215–9.

Zajonc RB (1965). Social facilitation: A solution is suggested for an old social psychological problem. *Science*, **149**, 269–74.

Zbilut JP, Giuliani A, and Webber CL, Jr (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A*, **246**, 122–8.

Zbilut JP and Webber CL, Jr (1992). Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, **171**, 199–203.