Ecological Psychology, 20:32–64, 2008
Copyright © Taylor & Francis Group, LLC
ISSN: 1040-7413 print /1532-6969 online
DOI: 10.1080/10407410701766643

Routledge
Taylor & Francis Group

Gestures and Phases: The Dynamics of Speech-Hand Communication

Paul Treffner and Mira Peter
Metaffordance Consultants

Mark Kleidon
University of Southern Queensland, Australia

We investigated how a listener's perceived meaning of a spoken sentence is influenced by the relative timing between a speaker's speech and accompanying hand gestures. Participants viewed a computer-animated character who uttered the phrase, "Put the book there now." while executing a simple right-handed beat gesture whose location relative to the utterance was precisely controlled in a frame-by-frame fashion. The participant's task consisted of making a judgment about two related aspects of the actor's perceived speech: (a) Which word was emphasized? and (b) How clear was the emphasis? That is, did it make sense? The results revealed that the perceived emphasis was determined by the timing (phasing) of the speaker's hand gesture. Furthermore, the clarity of the perceived emphasis (i.e., meaningfulness) was influenced by the affordances in the immediate environment of the speaker. Discussion addresses the primacy of ostensive specification and gesture in communicative events, the dynamics of speech-hand coordination during both actual and virtual dialogue, and the role of environmental affordances in grounding informative communicative acts in the ecology of organism-environment dynamics.

The question of the extent to which hand gestures (and other movements sometimes referred to as "body language") might provide a basis for human communication continues to challenge theories of language and language development. The fact that most of us move our hands in spontaneous gesture even if we cannot be seen, for example when talking to a blind person or to someone by telephone (Iverson & Goldin-Meadow, 1997), suggests that a speaker's accompanying hand

Correspondence should be addressed to Paul Treffner. E-mail: L84sky@gmail.com

motion may play a role in helping the listener to *directly perceiv*e—or “grasp”—the meaning of an utterance. Why might hand and body movement play such a central role in language? While some researchers believe that manual gestures preceded speech and that a relatively sharp transition from a predominantly gestural to a predominantly vocal form of language occurred as a consequence of a speciation event (Corballis, 2003), others maintain that the process whereby human communication altered from monkey-like actions towards affordances (action-related properties of the environment relevant to action; Gibson, 1979/1986) to a predominantly vocal process was relatively slow and cumulative (Arbib, 2005). However, the observation that coherent, fluent speech is usually accompanied by similarly coherent gestures (Blake, Olshansky, Vitale, & Macdonald, 1997) has led to the proposal that speech and hand gestures may have necessarily co-evolved in a reciprocal manner, with speech and gesture manifesting as complementary aspects of a single production-perception mechanism rather than with gestures preceding speech or vice versa (e.g., McNeill, 2000, 2005).

The reason for the tight temporal coupling between speech and gestures in everyday communication may stem from the common source from which both gestures arose, that is, a general perceptual sensitivity to specific rhythmic patterns (McNeill, 1992). It has been hypothesized that the “ba-ba” babbling of infants (with onset around 7 months) may be a nonverbal motor activity related to the emergent control over mouth and jaw, or a linguistic activity reflecting early sensitivity to phonetic–syllabic patterns (Pettito, Holowka, Sergio, Levy, & Ostry, 2004). Drawing on the classic work by Kendon (1972), McNeill has long argued that there is an extremely close synchrony between gesture and speech such that the two operate as an inseparable, coherent unit that embodies the language production process itself rather than just reflecting different outputs from it (McNeill, 1992, 2000; see also Furuyama, 2002). Thus, words and gestures are not just expressions of thought but instead such acts *constitute the thinking process itself.* Evidence of this tight synchrony includes the fact that disrupting speech also disrupts gestures and vice versa and that stutterers modify their gestures to maintain synchrony with speech (Mayberry & Jaques, 2000; Mayberry & Shenker, 1997; see also van Lieshout, 2004, and van Lieshout, Hulstijn, & Peters, 1996). At the more macroscopic level of collectives or groups of gestures and their relation to explanatory speech, appropriate coordination is also crucial. The “mismatch” of information expressed in speech and that expressed via gesture can, surprisingly, facilitate learning. For example, children (especially those who were in a transitional phase of comprehending) learned more effectively when a teacher attempted to explain the mathematical concept of “equivalence” through simultaneously offering complementary but not identical information in speech and hand gesture (Goldin-Meadow, 1999; Singer & Goldin-Meadow, 2005). Likewise, research on language acquisition indicates that gesture plays a key role in a child’s learning, for example, in learning to count (Alibali & DiRusso, 1999; Mayberry & Nicholadis, 2000).

From the perspective of ecological psychology, in order to satisfy the most basic theoretical demands of logical consistency and empirical relevance, communicative behavior and its corresponding semantics must be “bound to” or “grounded within” an organism-relevant, action-oriented environmental context (Shaw, Turvey, & Mace, 1981; Turvey, Shaw, Reed, & Mace, 1981). Ostensive specification (e.g., pointing with the finger) is the most basic act of making another human aware of one’s intentions via a communicative technique that makes *direct epistemic contact* with the environment—encountering without “physically” touching—but resulting in knowledge of comparable reliability (Reed, 1996; Shaw, 2001, 2003; Shaw et al., 1981). Ostensive specification in the form of pointing typically has a standardized form within a given culture (McNeill, 2000). Although the Western standard form for pointing is with the index finger extended and the other fingers curled in, some cultures use two fingers or the entire hand; such forms of pointing are still recognized in Western cultures. As a form of ostensive specification, pointing (and its meaning) can be fully perceived without accompanying speech. Pointing can provide the somatic basis for a perceiver to apprehend (literally, “grasp”) a speaker’s intended meaning. That is, a listener may directly perceive (understand) what a speaker means largely because the listener has an inherent haptic awareness of his or her own limbs and extremities, specifically their position and orientation (Pagano & Turvey, 1995), as well as, it is important to note, their dynamics and phase relations (Wilson, Bingham, & Craig, 2003). At stake is an appreciation of the possible role that gesture plays in grounding communication by affirming the primacy of direct perception—not just perception of an object, environment, or situation—but perception of an object, environment, or situation *as indicated to a perceiver by a speaker* during a communicative speech act (Shaw et al., 1981). Ostensive specification can turn a communicative act from semantically opaque (i.e., meaningless) into a semantically transparent and clearly understandable utterance. In such specifying acts, an object or event, or more precisely, a “complex particular,” is specified by the speaker and consists of the coordination of agent, situation, and occasion as a *self-presenting* (i.e., represents itself) *fact* rather than as a symbolic representation (i.e., represents something other than itself). Said differently, it is the *force of existence* (knowledge from acquaintance) rather than the force of argument (knowledge from description) that underlies even the possibility (and certainly the felicity) of communicative acts between a speaker-actor and a perceiver-listener (Shaw et al., 1981; see especially pp.199–209). Indeed, the power of ostensive specification is such that even nonhuman species apparently understand (or can directly perceive) the meaning of a human’s gestural command, such as when a dog’s owner gestures to “go fetch”—not only to go fetch an object but even to go to a particular location if several fetchables are available.

However, pointing is by no means the only form of manual gesture that humans exhibit. Pointing is one component of a wider class of gestures known as gesticulation, which can be subdivided into iconic, metaphoric, deictic, and beat gestures (Kendon, 1974). Both iconic and metaphoric gestures are considered "pictorial" gestures in that the speaker uses the gesture to help convey awareness of the shape or style of the object or abstract concept being referred to (e.g., the shape of a spiral staircase or the idea of openness). Pointing gestures are also known as *deictic* gestures and often clearly specify a referent in the environment or context of the discourse. A beat gesture or simply, beat, is a short, rhythmic movement or series of movements of the hand or hands that can gauge and influence both the overall rate at which a discourse occurs, the transitions between sections and topics of the discourse, as well as more subtle aspects of emphasis and semantics. In this research we are primarily concerned with beats and the extent to which they may also play a dual "deictic role" that is dependent on both their temporal relation to speech and on the environmental context within which they occur.

Numerous findings support the idea that manual gesture and speech execution are tightly coupled, both in fluent and in stuttered speech (Mayberry & Jaques, 2000). Additionally, it has been observed that the majority of gesture strokes (the most meaningful and effortful part of the gesture) occur just before or during the speaker's articulation of the most contrastively stressed syllable (Kendon, 1974; McNeill, 1985, 1992), whereas the preparation phase often anticipates speech (McNeill, 1992). Similarly, Nobe (2000) showed that most representational gestures had their onset during speech articulation. However, it was observed that in 10% of cases speech and the accompanying hand gesture were actually asynchronous—the onset of the hand gesture preceded the co-expressive speech by at least 250 ms, which roughly corresponds to the duration of a syllable (Nobe, 2000). Similarly, it was shown that gesture onset and speech production were synchronized within a few milliseconds of one another although the hand gestures slightly preceded speech (Morrel-Samuels & Krauss, 1992).

It has been suggested that beat gestures are used to emphasize a particular word or phrase and usually correspond temporally to a single syllable. It has also been proposed that they are used as a rhythmic marker that can guide organization of prosodic phrases (McNeill, 1992). However, variability has been observed in the timing between stressed syllables and accompanying beat gesture strokes. While the synchronization between deictic gestures and speech has been shown to be clear, it has been argued that gestural beats and verbal stress are not synchronized in such a strict rhythmic manner (McClave, 1994). As beat and deictic gestures may be considered the most primitive type of communicative movements, and because beats are relatively conspicuous and easy to analyze (e.g., a simple up-down or left-right movement of the hand or forearm), it is instructive to focus on this form of linguistic action before addressing more complex phenomena such as the iconic or metaphoric gestures found in everyday conversation.

Although beat gestures usually occur naturally during speech, on some occasions there may be little or no perceptible movement of the hands during conversation, such as when the hands are physically restricted when employed for another activity or when the hands of the speaker are not visible such as during a phone conversation. How then does this affect clarity of linguistic perception during communication? In driving a car, it has been shown that when the driver's conversation is directed outside of the immediate environment of the vehicle—as when using a hands-free mobile phone—coordination and control of driving seriously deteriorates (Treffner & Barrett, 2004). Data on the stability and skill of driving suggest that under such conditions a driver's perception and attention are compromised (Treffner, Barrett, & Petersen, 2002). This addresses the oft-noted observation that it is easier (less demanding of linguistic attention) for a driver to converse with a front-seat passenger than for a driver to speak to someone at a distance using a mobile phone. This may be because the driver can (peripherally) view various informative gestures of the adjacent passenger such as hand motion, head nodding, and facial expressions while simultaneously maintaining awareness of the immediate driving environment. The driver who uses a mobile phone has no such nonverbal information available and so deciphering the other person's speech is made all the more demanding of attention. Given the key role of attention in maintaining dynamic stability during critical acts of motor coordination (e.g., Treffner & Kelso, 1999), having nonverbal information available facilitates a more immediate and direct apprehension of linguistic content, which in turn can release attentional resources for the critical task of safely controlling the vehicle. Examples such as these demonstrate the prevalence and importance of nonverbal gestures in everyday communication.

Regarding hand preference during gestural coordination, Kimura (1973) noted that the right hand of right-handers was chosen for free movements that accompanied speech. However, no preference was found when displaying iconic gestures while speaking (Lausberg & Kita, 2003) and that the left and right hands were used equally often for beat gestures (Sousa-Poza, Rohrberg, & Mercure, 1979). In the experiment reported herein, hand gestures of the animated character were performed unimanually and exclusively using the right hand. Although not essential to the design of our experiment, this is consistent with our previous data on both left- and right-handers suggesting that speech-hand coordination dynamics are supported by a common timing system primarily involving the left cerebral hemisphere, regardless of handedness (Treffner & Peter, 2002; Treffner & Turvey, 1995, 1996).

In order to investigate the effects of speech-hand timing on sentence perception, it is necessary to strictly control the relative temporal positions of the manual gesture and its lexical affiliate. This is impossible in a normal biological system as the constraints, both neuromuscular and dynamical, are adamant of the requirement that normal functioning (goal-directed coordination for conveying meaning) should not be violated. For this reason we must create a situation that is not so artificial that it cannot be treated as potentially realistic but not so natural that it cannot be parametrically controlled. We therefore turned to computer animation and the ability to create a reasonably lifelike character or "avatar" that exhibits gestures during accompanying speech (Stanney, 2002).

Our prior research into speech-hand coordination focused on the *production* side of the phenomenon and how the synchrony or asynchrony (i.e., phasing) of speech-hand coordination evolves as a function of handedness, hand, and rate of production (Treffner & Peter, 2002). In the current research we focus on the *perceptual consequences* of viewing differential phasing phenomena. More precisely, we investigate how a listener's perception of the meaning of a speaker's utterance is influenced by visual information that specifies the relative timing of beat gestures and associated speech. In particular, regarding a sentence's semantics, we investigated whether the listener's *perceived focus* of a sentence depends on the temporal *coordination* of a speaker's gesture and speech and also whether the *perceived clarity* of a sentence depends on *environmental context*.

METHODS

Participants

A total of 14 individuals (5 females and 9 males) ranging from 15 to 48 years of age, volunteered to participate. Participants did not know that gestures were the topic of interest. Advertisements and instructions simply described the experiment as "about communication."

Materials

A full-color 3D computer-generated animated character was created as a means to synchronize a hand gesture to precise locations along the acoustic speech signal. Animations were created on a computer using key framing techniques using a combination of the animation editing programs Poser, 3D Studio Max, and Premiere. The phrase "Put the book there now." was chosen as an utterance and the speaker's mouth, cheeks, and chin were rendered to reproduce, as faithfully as possible (given available desktop PC resources) the movements that would accompany such a phrase. An inter-frame resolution of 33.33 ms yielded an animation of 58 frames in length and a total animation sequence of 1.91 s in duration (Figure 1).[1]

The times between the end of one word and the beginning of the following were adjusted such that the gaps between words sounded equal. Almost all intonation

[1] Example animations from the experiment can be viewed and downloaded on the Internet at http://gesturesandphases.tripod.com and http://www.trincoll.edu/depts/ecopsyc/gestures.htm. They can also be requested from Paul Treffner at L84sky@gmail.com.

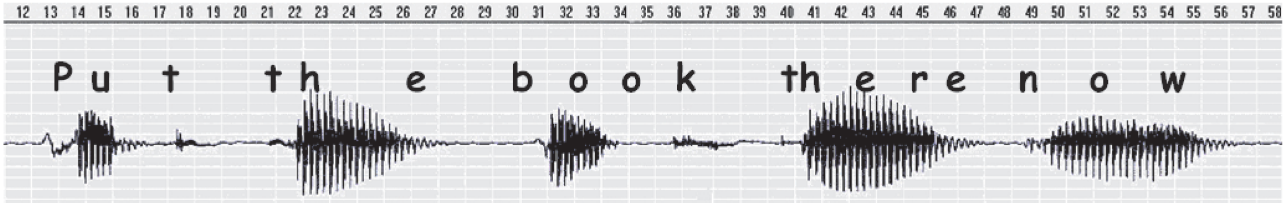

FIGURE 1 Voice sonogram of full phrase, "Put the book there now." with lexical correlates approximated. Each animation involved centering a beat gesture on a single frame of the speech signal. Animations were created with gestures centered on frames 27 through 49 (end of "the" through beginning of "now").

(e.g., relative pitch and tone differences) was removed from the speech signal in an attempt to eliminate acoustic information about the “focus” of the utterance and, also important, removed a possible gaze bias toward hand gesture stemming from prosody (McClave, 1997). Although somewhat monotonic, the utterance was a processed version of actual continuous human utterances and was still reasonably natural; it certainly sounded more natural than synthesized “computer-speech.”

Because we are interested in the relation between relative timing of gestural movement and meaning, we chose as simple a hand movement as possible while still retaining its “gestural” properties. We chose a beat gesture that involved a simple “out-hold-in” hand motion beginning with the right hand with closed fist postured in front of the chest, extension at the wrist with extension of fingers (pre-stroke, 3 frames in duration), a brief hold of this posture (mid-stroke, 4 frames), then flexion of wrist and fingers (post-stroke, 3 frames) to return to a posture with closed fist positioned in front of the chest (Figure 2). The complete gesture lasted 333 ms or a “window” of 10 frames in size. In order to create animations for the experimental conditions, the gesture was incorporated into the animation such that the midpoint of the 10-frame gesture window was subsequently centered on each of 23 positions corresponding to frames 27 through 49 (inclusive) within the subsegment of the utterance corresponding to “… book … there …” (Figure 1). Because in each animation the middle of the 10-frame gesture window was centered on one of 23 individual frames from the “book-there” segment, the initiation of the gesture would begin earlier. For example, for a gesture 10 frames in length centered on frame 32 (beginning of “book”), the initiation of the gesture would occur in frame 27 (i.e., 32 – 5 = 27), which corresponds to the end of “the” (Figure 1). In sum, the animated speaking articulatory movements and audio speech signal remained constant for all animations. The main difference between the animations involved where the hand gesture was synchronized (i.e., which section of the

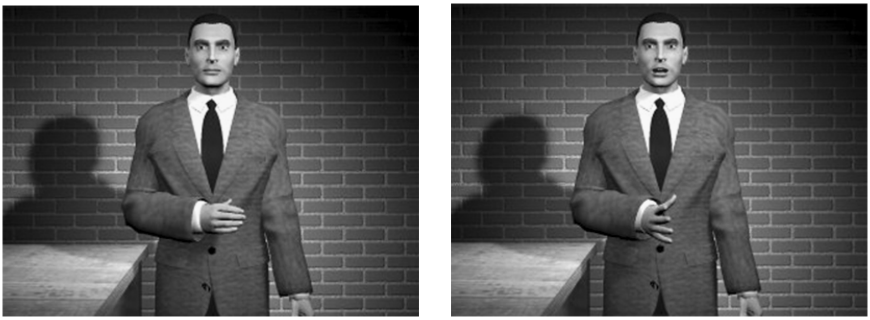

FIGURE 2 Beat gesture and stroke. The hand is held in a pre-stroke posture suspended in mid-air (left panel). At the mid-stroke position the hand moved forward with fingers fully extended (right panel). Finally, the hand and fingers return to a post-stroke position (which is identical to the pre-stroke posture shown). (Still images rendered from the 10-frame animated gestural sequence.)

speech signal). Since a “gesture” is not a momentary occurrence but rather a spatiotemporal event lasting 10 frames, synchronizing the gesture with different portions of the speech event involved sliding a gestural window along the speech signal; gestural location is then defined as the frame of the speech signal on which the gestural window is *centered*.

As we wished to investigate the role of environmental context in the perception of speech-hand gestural communication, half of the animations included a table in the background of the animated speaker while in the other half the table was absent. For control condition purposes, two animations were created whereby the speaker made no gesture (with and without a table). In these cases, the speaker’s arms and hands remained motionless by his sides (Figure 3).

Either a large back-projection video screen (2.5 m x 2 m) or a computer monitor was used to present stimuli to participants. Participants were positioned at a distance from the large screen such that the visual angle subtended was similar to that when using the smaller computer monitor. Audio volume presented via loudspeak-

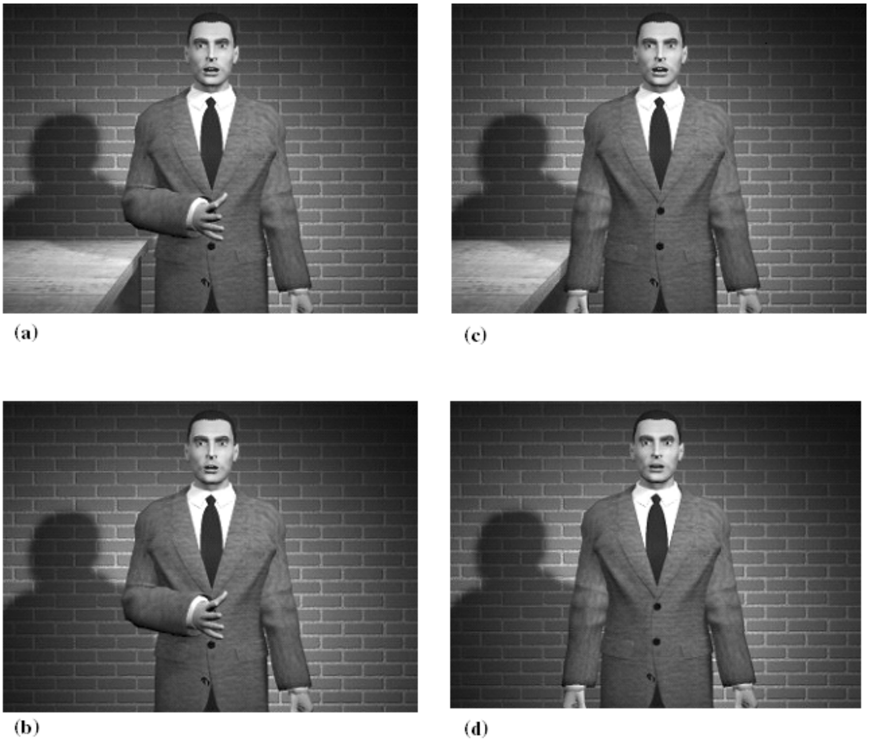

FIGURE 3 The four situations used in the experiment: (a) Gesture, table; (b) Gesture, no table; (c) No gesture, table (control); (d) No gesture, no table (control). (Still images from the animated sequence.)

ers behind either the large screen or the smaller monitor was adjusted to provide similar volumes. Video segments were played to participants using Media Player.

Procedure

Each participant was seated in front of the display and a series of animations was shown whereby an animated male character uttered the phrase "Put the book there now." The participant was required to view each animation and record responses on paper at the end of each individual 1.5-s animation. Participants were asked to address two specific issues:

1. *"Which word is the intended focus of the sentence, that is, what is being emphasized?"*
2. *"The focus of the sentence was emphasized clearly."*

Response sheets consisted of three columns headed "Trial number," "Focus," and "Rating." In the "Focus" column, a participant was asked to circle one of the five words printed to indicate which word he or she thought was the speaker's *intended focus* of the sentence (i.e., "categorization"). Similarly, for each trial a participant had to indicate on a 5-point scale his or her *level of agreement* with the statement "The focus of the sentence was emphasized clearly" (i.e., "clarity"). The participant was asked to record his or her level of agreement by circling one of a list of five numbers (1 = *strongly agree,* 2 = *agree,* 3 = *neither agree nor disagree,* 4 = *disagree,* 5 = *strongly disagree*).

Several reasons pertain to designing a dependent measure of "clarity" for perceptual experiments. It provides an additional insight into the participant's perception. It urges the participant to attend carefully to what is perceived by requiring a "qualitative" more reflective second-order evaluation (*how clear* was the perception?) as well as the first-order "quantitative" judgement (*what* was perceived?). This avoids potentially less conscientious pencil-circling indications of one of the five words heard. It provides a way to evaluate the degree of "meaningfulness" of perception—to what extent the perceived word "makes sense." It is one thing to "hear" a certain word uttered (e.g., consider a computerized parser or "speech perception" program); it is quite another to perceive that the word perceived *makes sense* given the accompanying spatiotemporal context of body movements and environment. The latter reason is perhaps the most forceful for incorporation of a measure of clarity of perception. The main thrust of our experimental design was to disassociate the normal tight synchrony between body and vocal languages in order to reveal the consequent effect on meaningfulness (or "clarity" or "certainty" or "sensibleness") of judgment.

Participants were not constrained as to where in the animation they looked. Thus, participants were, as is normal, free to focus on the head, torso, face, mouth, hand, or any other combination of body or environmental features. However, in order to complete the experimental requirements of circling which word was perceived and how clearly it was perceived to be emphasized, it must be assumed that participants were aware of both hand motion and words uttered.

In the experiment there were 144 trials consisting of three repetitions of each of the 48 conditions The trials were completely randomized and each participant was presented with a different random order of 144 trials. Conditions included (a) 23 animations with gesture and a table, (b) 23 animations with gesture but no table, (c) one animation with no gesture and a table, (d) one animation with no gesture and no table. Conditions "c" and "d" were used as control conditions (Figure 3).

Data Reduction

Mean percentages were calculated for the categorization and ratings data for each frame (on which the gesture was centered) by averaging across the three repetitions of each condition and across participants. The paired sample *t* test ($p < .05$) was used for both categorization and ratings data to compare table-present and no-table conditions as well as pairs of frames within a table-present or no-table condition. In accord with standard psychophysics conventions, the response threshold was set at 50% to determine the region in which speech-hand synchrony could be considered to influence the perceived focus of the sentence (i.e., categorization). To determine the region where the relative timing of gesture and speech appears to yield important information for perception and comprehension, we introduce the *"perceived synchrony region (PSR)"* for each of the five words in the utterance "Put the book there now." The PSR is defined, using the results data, as the region bounded by the earliest and latest frames in which a particular word (e.g., "book") is perceived (chosen) by the participant as the focus of the sentence 50% or more of the time when the gesture in the animation was synchronized with (centered on) that frame.

RESULTS AND DISCUSSION

Categorization

Figure 4 shows the results for perceived focus and word categorization when the gesture was centered on (synchronized to) a particular frame of the animation and utterance. As there was no intonation information available in the speech of the animation, any perceived focus exhibited by participants must be due to factors other than tonal emphasis (e.g., gestural position, context, etc.). For the word *book* the PSR extended from Frame 28 to Frame 37 (69.05% and 64.28%, respectively) when the table was present and from Frame 29 to Frame 37 (64.28% and 73.81%, respectively) when the table was absent. For the word *there* the PSR extended from Frame 38 to Frame 48 (52.38% and 52.38%, respectively) for both table-present

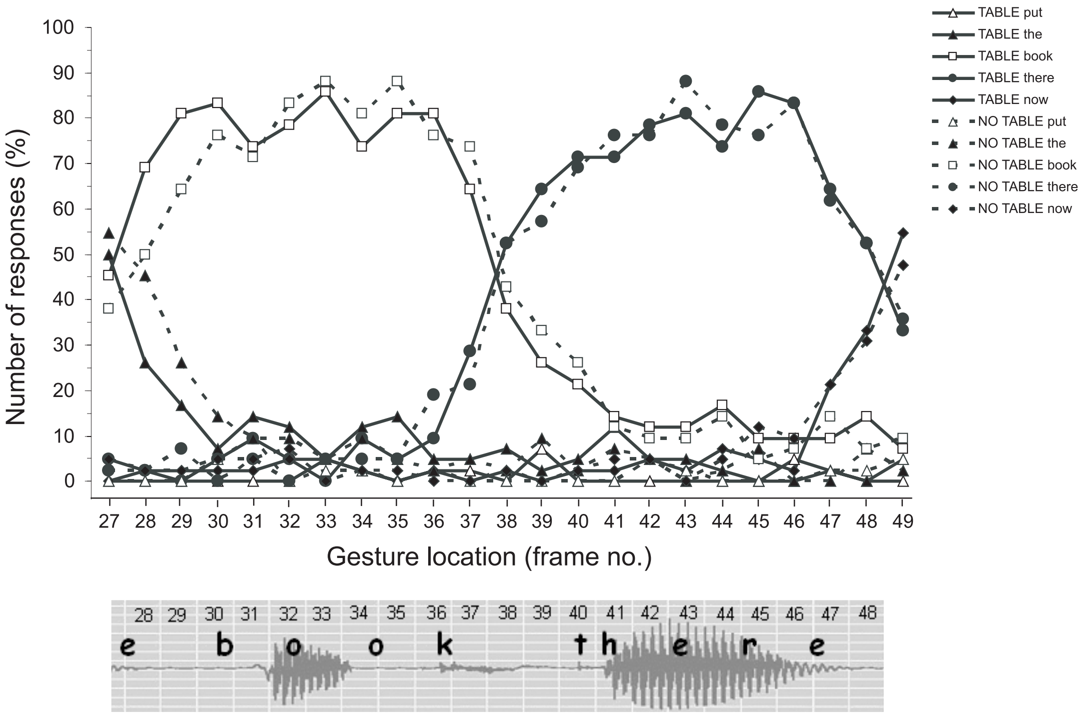

FIGURE 4 Word categorization results. The x-axis indicates frame position whereby the gesture was centered on or synchronized to each of Frames 27 through 49. The y-axis indicates the extent to which listeners considered a particular word as the *focus* of the sentence when the gesture was correspondingly synchronized. Solid lines represent table-present condition; dotted lines represent the no-table condition. Also shown is a segment of the raw speech signal and its relation to the words uttered.

and table-absent conditions. Clearly, when a gesture is coordinated with a speaker's utterance, the gesture can dramatically influence a listener's perception of which word is emphasized. Further, perceived emphasis can be influenced by a gesture that *precedes* the start of the acoustic signal of the perceived word. Thus, in Figure 4 the "book" PSR (table present) begins at Frame 28 but the corresponding acoustic signal does not begin until Frame 31. At Frame 28, 69% of responses indicated "book" as the focus even though at this point the acoustic speech for "book" has not yet commenced. The middle of the gesture precedes the start of the speech signal for "book" by about 133 ms (4 frames). It should be noted the start of the gesture is a further 5 frames before the midpoint so in this example the onset of the associated gesture actually began at Frame 23 (= 28 – 5), which is in the middle of the speech for "the" and approximately 300 ms before speech onset for "book." The "priming" effect of gesture—either fledgling pre-stroke or fully-fledged mid-stroke—on speech, may surreptitiously provide "heads-up" information to the listener that allows the listener's attention to be directed to currently unfolding (nonacoustic) information about the current communicative event, which can help a listener attune to and anticipate what the speaker's intended emphasis should be perceived to be.

Categorization of "book," table. Within the "book" PSR (Frame 28 to Frame 37), "book" was selected significantly more often when the gesture was centered on Frame 33 than on Frame 34 (85.71% vs. 73.81%, respectively; $t(13) = 2.69, p < .05$) (Figure 4). No other significant differences were present within these frames.

A significant difference was noted between Frame 27 at the end of the region for "the" and Frame 28 at the beginning of the "book" PSR (45.24% vs. 69.05%, respectively; $t(13) = 2.22, p = .05$). Also notable is the difference between the last Frame 37 in the "book" PSR and Frame 38, the first frame of the "there" PSR (64.28% vs. 38.09%, respectively; $t(13) = 2.35, p < .05$).

Categorization of "book," no table. Within the PSR for "book" (Frame 28 to Frame 37), the word *book* was selected significantly more often with the gesture centered on Frame 33 than on Frame 31 (88.09% vs. 71.43%, respectively; $t(13) =2.18, p = .05$). No other significant differences were present within the PSR for "book."

No significant differences were noted between Frames 27 and 28 (38.09% vs. 50%, respectively). However, the difference noted between Frame 37 at the end of the "book" PSR and Frame 38 at the beginning of the "there" PSR was significant (73.81% vs. 42.86%, respectively; $t(13) = 2.88, p < .05$).

Categorization of "there," table. Within the "there" PSR (Frames 38 to 48), no significant differences between frames were found. A difference was found between Frame 37 at the end of the "book" PSR and Frame 38 at the beginning of the "there" PSR (28.57% vs. 52.38%, respectively; $t(13) = 2.22$, $p = .05$). Also notable is the difference between the last frame (48) in the "there" PSR and the first frame (49) in the "now" PSR (52.38% vs. 33.33%, respectively; $t(13) = 2.51$, $p < .05$).

Categorization of "there," no table. Within the "there" PSR (Frames 38 to 48), "there" was selected significantly more often in Frame 43 than in Frame 41 (88.09% vs. 76.19%, respectively; $t(13) =2.18$, $p = .05$). Also, "there" was selected significantly more often in Frame 43 than in Frame 44 (88.09% vs. 78.57%, respectively; $t(13) =2.18$, $p < .05$) and more often in Frame 46 than in Frame 47 (83.33% vs. 61.90%, respectively; $t(13) = 2.59$, $p < .05$). No other significant differences were found within this PSR.

A significant difference was noted between Frame 37 at the end of the "book" PSR and Frame 38 at the beginning of the "there" PSR (21.43% vs. 52.38%, respectively; $t(13) = 3.04$, $p < .01$). However, the difference between the last frame (48) in the "there" PSR and the first frame (49) in the "now" PSR was nearly significant (52.38% vs. 35.71%, respectively; $t(13) = 1.99$, $p = .07$).

No-gesture, table versus no table. Regarding the two control conditions (i.e., utterance alone, no synchronous gesture, arms by sides), with the table present, "there" was chosen most often as the focus of the sentence (19.05%), although the magnitude was not significantly greater than for any other word categorized. It is important to note that with the table absent, "there" was not selected at all and "book" was perceived significantly more often than "there" to be the focus of the sentence (26.19% vs. 0%, respectively; $t(13) = 2.78$, $p < .05$). However, "book" was not selected more often than "put" (7.14%), "the" (16.67%), or "now" (11.90%), and "now" was chosen significantly more often than "there" ($t(13) = 2.11$, $p = .05$).

Discussion of categorization results. The PSRs spanned between 9 and 11 frames (e.g., the PSR for "there" spanned Frames 38 through 48, inclusive). Within the PSR of a given word, few frames exhibited a selection rate that was significantly different from any other frame in the region. This result indicates that a synchronous hand gesture has a relatively uniform influence on the perceived focus of a spoken sentence regardless of position or timing of the gesture.

The effect of a beat gesture was evaluated by synchronizing the mid-stroke of the gesture to different positions along the speech signal of the animation. The participant's response to (perception of) this synchrony showed that the perceived focus of the speaker's utterance was based on a temporal region (the PSR) that begins well *before* the onset of the perceived word's speech signal and a region that ends fairly abruptly during or after the speech signal's offset. The PSR for "book" begins at

Frame 28 before the speech signal that begins at about Frame 31, whereas both the PSR and the speech signal end at about Frame 37 (Figure 4). Likewise, the PSR for “there” begins at Frame 38 but the speech begins later, at around Frame 40; the PSR ends at Frame 48, as does the speech signal. This finding shows that a gesture that slightly precedes or is *phase-advanced* relative to its lexical affiliate will have greater influence on perceived emphasis than a gesture synchronized later or at the end of its associated speech-generated acoustic event. A gesture “perfectly synchronized” or *in-phase* relative to the accompanying acoustic event has the greatest chance of influencing a listener’s perception of intended emphasis. However, given the phenomenon of coarticulation, it may be that phase-advanced gestural events help provide the important *anticipatory information* necessary to specify the intended content of speech acts that might otherwise become lost in and muddled by the astounding unsegmented continuity of the human acoustic speech signal.

The paradigm also allowed us to investigate the influence of environmental context and whether perception of the focus of a sentence was dependent on the presence of a relevant object (a table) in the environment that was potentially related to the topic of the sentence. Results indicated that there was no real difference in categorization likelihood between the table-present and table-absent conditions *when gesture was involved*. This held except for “book” near the start of its PSR (Frame 29), where “book” was perceived as the focus more often with the table than without the table (80.95% vs. 64.28%, respectively; $t(13) = 2.46$, $p < .05$). However, recall that for the no-gesture condition there was a definite effect of the table such that “there” was perceived as the focus significantly more often when the table was present than when absent.

Thus, when a gesture accompanies (monotonic) speech, the influence of environmental context appears diminished in comparison with the salient visual information provided by a synchronous hand gesture. However, the results suggest that environmental context can influence the perceived focus of a sentence in the absence of disambiguating information from gestures. An explanation for this finding would be that the table is perceived to afford a surface upon which to put a book and, moreover, to put it in a definite location (i.e., *there*). We take this as evidence that when apprehending the meaning of communicative utterances, perceivers take account of the *affordances* of the immediate environment and that the latter can contribute to the (direct) perception of the meaning of a speaker’s utterance (Fowler, 1986; Gibson 1979/1986).

Transition region: “book” versus “there.” In Figure 4 a clear transition can be seen between the PSRs for “book” and “there.” Response data on the two PSRs were compared around the crossover region (Frames 36, 37, 38, 39, and 40). In the table-present condition, for Frame 36, “book” was chosen significantly more often than “there” (80.95% vs. 9.52%, respectively; $t(13) = 6.20$, p < .001). For Frames 37, 38, and 39, no significant differences were found between rates for choosing "book" and "there." For Frame 40, "there" was chosen in preference to "book" (71.43% vs. 21.43%, respectively; $t(13) = 3$, p = .01). Thus, a 5-frame book-there crossover region exists for Frames 36 through 40, where uncertainty between perceiving "book" and "there" exists for the central 3 frames (37, 38, and 39). Preceding this uncertainty region with a gesture synchronized on Frame 36, participants overwhelmingly chose "book" over "there." Likewise, following this region with a gesture synchronized on Frame 40, participants overwhelmingly chose "there."

In the no-table condition, for Frame 36, "book" was chosen more often than "there" (76.19% vs. 19.05%, respectively; $t(13) = 3.92, p < .005$). Interestingly, for Frame 37, in contrast to the table-condition, "book" was chosen more often than "there" (73.81% vs. 21.43%, respectively; $t(13) = 3.14$, $p < .01$). For Frames 38 and 39, no differences were found between responses of "book" and "there". For Frame 40, "there" was chosen more often than "book" (69.05% vs. 26.19%, respectively; $t(13) = 2.59, p < .05$). Hence, in the absence of a table, the crossover region and region of uncertainty was compressed compared with when a table was present and encompassed only 4 frames instead of 5 (37, 38, 39, and 40) with the central frames 38 and 39 providing a region of uncertainty between choosing "book" and "there." Therefore, *with a table present*, a gesture centered on Frame 37 yielded uncertainty as to whether "book" or "there" had been emphasized (i.e., perhaps the observer thinks, *"Maybe the word 'there' instead of 'book' was emphasized—that would make sense because there's a table in the background!"*). This is all the more telling because at Frame 37 the acoustics speech signal for "book" has not finished and the speech signal for "there" has yet to commence.

In sum, the preceding results show that the absence of a table increases the disposition for a participant to perceive "book" as the focus when a synchronous gesture is performed, whereas the presence of an environmental object (a table) increases the disposition to perceive the focus of the sentence as a word that can also meaningfully relate to the contextual environment (i.e., "there"). This subtle but definite effect of the location of a synchronous gesture demonstrates that the environment of a speaker, if relevant to the speaker's utterance (i.e., it has action-related affordance properties), does influence the semantic content of perceived speech.

Clarity Ratings

Clarity of "book," table. In the following sections, we report the extent to which participants agreed or disagreed that their perceived focus was *emphasized clearly*. Considering the table-present condition and those participants who perceived that the focus of the sentence was "book" (cf. the PSR for "book," Frames 28–37), more participants "strongly agreed" that the emphasis was clear when the gesture was centered on Frame 29 (nearer the beginning of the speech signal for "book") than when the gesture was synchronized with Frame 28 (21.43% vs.

9.52%, respectively; $t(13) = 2.11, p = .05$) (Figure 5). Also, more “strongly agreed” that the emphasis on “book” was clear for a gesture synchronized with Frame 34 (farther from the end of the speech signal for “book”) than when synchronized with Frame 35 (21.43% vs. 9.52%, respectively; $t(13) = 2.11, p = .05$).

For the “agree” rating, participants judged the emphasis on “book” to be clear more often when the gesture was centered on Frame 30 than on Frame 31 (52.38% vs. 30.95%, respectively; $t(13) = 2.59, p < .05$). Also, participants “agreed” the emphasis was clear more often for Frame 33 than for Frame 31 (45.24% vs. 30.95%, respectively; $t(13) = 2.12, p = .05$). More participants “agreed” the emphasis was clear for Frame 36 than for Frame 37 (47.62% vs. 23.81%, respectively; $t(13) = 2.11, p = .05$). No other significant differences were found for any of the other clarity ratings within the PSR for “book.” Also, no other differences were found for any of the ratings between Frame 27 at the end of the PSR for “the” and Frame 28 at the beginning of the “book” PSR. There was also no significant difference found between Frame 37 at the end of the “book” PSR and Frame 38 at the beginning of the “there” PSR.

Clarity of “book,” no table. Without the table present, within the “book” PSR, more participants “strongly agreed” the emphasis on “book” was clear when the gesture was centered on Frame 33 than on Frame 31 (38.09% vs. 19.05%, respectively; $t(13) = 2.51, p < .05$) (Figure 5). Participants also exhibited ambivalence regarding clarity of their perceived focus such that more selected the rating “neither agree nor disagree” for gestures centered on Frame 35 (the relatively silent closure region toward the end of the speech for “book”) than for Frame 36 (corresponding to the voiceless stop, “k,” at the end of “book”) (21.43% vs. 4.76%, respectively; $t(13) = 2.46, p < .05$). No other significant differences were noted for adjacent frames of this PSR.

More participants “disagreed” that the emphasis on “book” was clear when the gesture was centered on Frame 27 (end of “the” PSR) than on Frame 28 (beginning of “book” PSR) (2.38% vs. 21.43%, respectively; $t(13) = 2.51, p < .05$). Correspondingly, more participants “agreed” that the emphasis on “book” was clear for Frame 37 (end of “book” PSR) than for Frame 38 (beginning of “there” PSR) (40.48% vs. 11.9%, respectively; $t(13) = 2.28, p < .05$).

Clarity of “there,” table present. When the table was present, within the PSR for “there” (Frames 38–48) more participants “strongly agreed” that the emphasis on “there” was clear when the gesture was centered on Frame 39 closer to the start of the speech for “there” than farther from it on Frame 38 (19.05% vs. 4.76%, respectively; $t(13) = 3.12, p < .01$) (Figure 6). Similarly, more participants “agreed” the emphasis was clear for Frame 40 than for Frame 39 (42.86% vs. 26.19%, respectively; $t(13) = 2.19, p = .05$). Surprisingly, more participants “agreed” the emphasis was clear when the gesture was centered on Frame 43 than

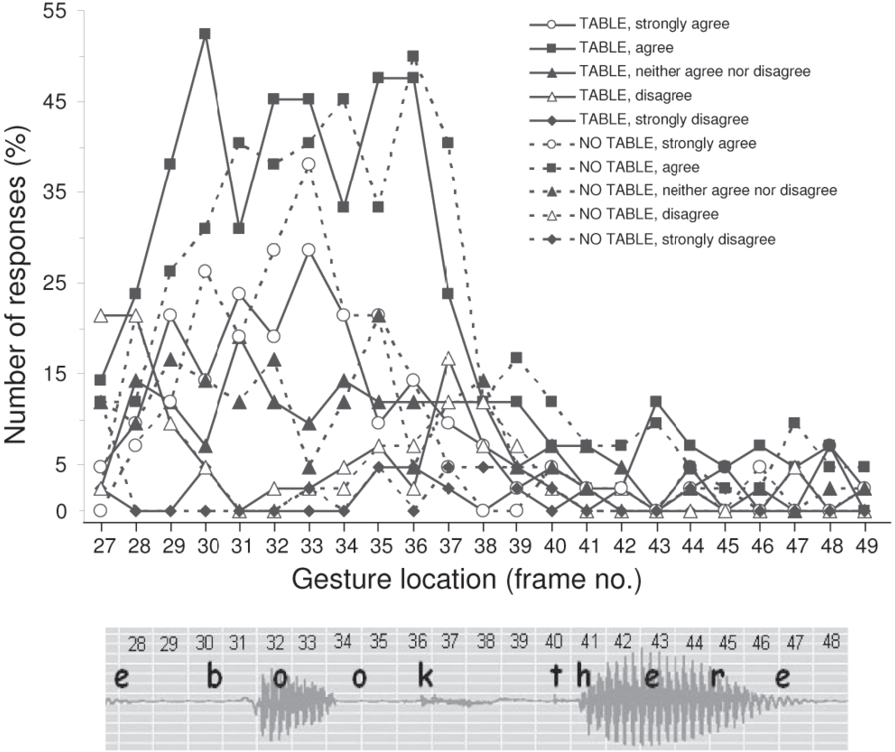

FIGURE 5 Clarity ratings for the word "book." The x-axis indicates frame position whereby the gesture was centered on each of Frames 27 through 49. The y-axis indicates the extent to which participants agreed that the focus on "book" *was clear*, given that "book" was the perceived focus. Solid lines indicate responses for table-present conditions; dotted lines indicate responses for the no-table conditions. Also shown is a segment of the raw speech signal and its relation to the words uttered.

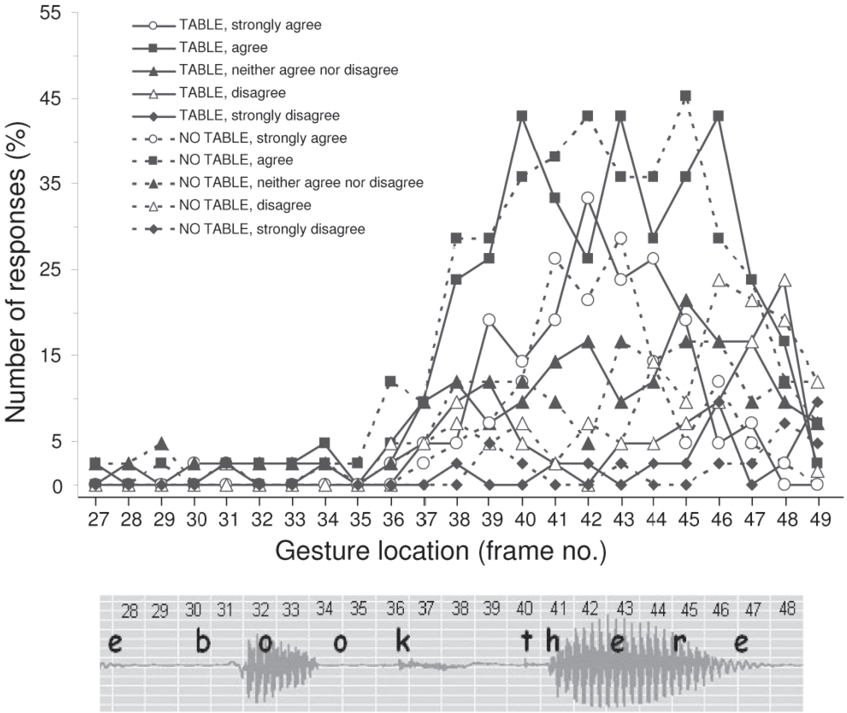

FIGURE 6 Clarity ratings for the word "there." The x-axis indicates frame position whereby the gesture was centered on each of Frames 27 through 49. The y-axis indicates the extent to which participants agreed that the focus on "there" *was clear*, given that "there" was the perceived focus. Solid lines indicate responses for table-present conditions; dotted lines indicate responses for the no-table conditions. Also shown is a segment of the raw speech signal and its relation to the words uttered.

on Frame 42 (42.86% vs. 26.19%, respectively; *t*(13) = 2.46, *p* < .05) and, as might be expected, for Frame 46 compared with 47 (toward the end of the speech) (42.86% vs. 23.81%, respectively; *t*(13) = 2.28, *p* < .05). No other significant differences were noted between frames within the "there" PSR for the other clarity ratings.

Although no significant differences were found for any ratings between Frame 37 at end of "book" PSR and Frame 38 at the beginning of "there" PSR, more participants "agreed" the emphasis on "there" was clear for gestures centered on Frame 48 at the end of "there" PSR than on Frame 49 at the beginning of "now" PSR (16.67% vs. 2.38%, respectively; *t*(13) = 2.12, *p* = .05).

Clarity of "there," table absent. Without the presence of a table, and for "there" PSR, more participants "strongly agreed" the emphasis on "there" was clear for Frame 41 (which is close to the center of the speech signal) compared with Frame 40 (which is close to speech onset) (26.19% vs.11.9% respectively; *t*(13) = 3.12, *p* < .01) (Figure 6).

For the frames flanking "there" PSR, the only significant difference noted was for the rating "agree" for gestures centered on Frame 37 at the end of "book" and on Frame 38 at the beginning of "there" (9.52% vs. 28.57%, respectively; *t*(13) = 2.51, *p* < .05). Also, more participants were ambivalent and "neither agreed nor disagreed" that the emphasis was clear for Frame 43 than for Frame 42 (16.67% vs.4.76%, respectively, *t*(13) = 2.11, *p* = .05).

Clarity of "the" and "now". As reported earlier, significant differences in clarity ratings were found most commonly within the PSRs for "book" and "there." In contrast, no differences were found for "put," and few differences were revealed for the words "the" and "now." The only significant difference for clarity ratings for "the" involved the rating "agree" and was found in the no-table condition between Frame 27 at the end of the speech for "the" and Frame 28 at the start of "book" PSR (28.57% vs. 11.9%, respectively; *t*(13) = 2.88, *p* = .01).

For categorization responses of "now," the only differences involved the ambivalent response, "neither agree nor disagree." With table present, the ambivalent response was chosen less for Frame 48 at the end of "there" PSR than for Frame 49 at the beginning of "now" (2.38% vs. 14.28%, respectively; *t*(13) = 2.11, *p* = .05). Similarly, with table absent, the ambivalent response was selected less for Frame 47 near the end of "there" PSR than for Frame 48 near the beginning of "now" (0% vs. 9.52%, respectively; *t*(13) = 2.28, *p* < .05). Ambivalence may reflect indefinite perception. Since no conditions involved instances where a gesture was synchronized with the words "put," "the," or "now," and PSRs were only found for those words where a synchronous gesture existed ("book" and "there"), the paucity of significant differences in clarity ratings for the other words reflects the lack of perceiving these words as the focus.

Clarity ratings, gesture: Table versus no-table. With gesture present, for the clarity rating "strongly agree," no significant differences were found between table-present and table-absent conditions for any frame for any of the words. For the rating "agree," a significant effect of the table was noted for the words "book," "there," and "the." The number of participants who "agreed" that the emphasis on "book" was clear when the gesture was centered on Frame 28 (at the start of "book" PSR) was significantly greater when the table was present compared with when it was absent (23.81% vs. 11.9%, respectively; $t(13) = 2.11$, $p = .05$) (Figure 5). Oddly, for Frame 31, significantly more participants "agreed" that the emphasis on "book" was clear when the table was absent compared with when it was present (40.48% vs. 30.95%, $t(13) = 2.28$, $p < .05$). However, for Frame 35 (as with Frame 28), more participants "agreed" the emphasis on "book" was clear when the table was present compared with when it was absent (47.62% vs. 33.33%, respectively; $t(13) = 2.12$, $p = .05$.). Overall, it seemed the presence of a table increased clarity of perception that the focus was on "book."

For perceptions of "there" as the focus of the sentence, when the gesture was centered on Frame 46, significantly more participants "agreed" that the emphasis on "there" was clear when the table was present compared with when it was absent (42.86% vs. 28.57%, respectively; $t(13) = 2.48$, $p < .05$) (Figure 6). Because of the close relation between the affordance properties of a table and the semantics of "there," the preceding result can be well motivated.

For perceptions of "the" as the focus of the sentence, when the gesture was centered on Frame 27 more participants "agreed" that a focus on "the" was clear when the table was absent compared with when it was present (28.57% vs. 14.28%, respectively; $t(13) = 2.12$, $p = .05$). This can be understood by realizing that without the table, there is less bias from "book" and "there" (table-related concepts) that could reduce the participant's certainty and clarity that "the" was the focus of the sentence. For the ambivalent clarity rating of "neither agree nor disagree," no significant differences were noted between table-present and no-table conditions in any of the frames for any word.

For the clarity rating "disagree," significant differences were found for the words "book," "there," and "the." In Fame 27 (immediately preceding "book" PSR), more participants "disagreed" that the perceived emphasis on "book" was clear when there was a table compared with when it was absent (21.43% vs. 2.38%, respectively; $t(13) = 2.51$, $p < .05$) (Figure 5). Thus, although the gesture occurred immediately prior to the PSR for "book," some still perceived "book" as the emphasis of the sentence. In such cases, however, a participant thought his or her perception of "book" was more unclear when there was a table in the background than when there was no table.

It is important that in Frame 46 near the end of "there" PSR, more participants "disagreed" that the perceived emphasis on "there" was clear when there was no table compared with when it was present (23.81% vs. 9.52%, respectively; *t*(13) = 2.12, *p* = .05) (Figure 6). That is, without a table to help justify the speaker's emphasis on "there," the listener thought his or her perception of "there" was unclear. This finding provides the complement for the previous result from the "agree" ratings for "there" at Frame 46, where it was shown that the presence of a table increased the clarity of perceiving "there" compared with when there was no table.

For those who perceived "the" as the emphasis, when the gesture was positioned at Frame 28 (after "the" PSR and beginning of "book" PSR), more "disagreed" that the perception of "the" was clear when there was no table compared with when it was present (11.9% vs. 0%, respectively; *t*(13) = 2.11, *p* = .05). That is, with the mid-stroke of the gesture located to follow "the" PSR, the participant was more *unclear* regarding his or her perception of "the" when there was no table than when there was a table. Indeed, when there *was* a table, no participant who perceived "the" thought that it was unclear. This result reflects the semantics of the verb "put" and the definite article "the." The action word, "put," entails a location in which to place something. A table in the speaker's background would be consistent with the first word spoken, "put." For those participants who perceived and chose "the" as the emphasis, and although "the" was perceived unclearly (not surprising given the gesture's location), it was perceived more unclearly when there was no appropriate affordance (e.g., a table) onto which an article could be put (e.g., the definite article referred to by "the" in the sentence). Finally, for the clarity rating "strongly disagree," no significant differences were found between table and no-table conditions for any of the words.

Clarity ratings, no gesture: Table versus no-table. When there was no co-occurrent gesture (i.e., the control condition), the only significant effect of the table was for the clarity rating "neither agree nor disagree" for the word "there." More participants were uncertain whether the emphasis on "there" was clear when the table was present compared with when it was absent (11.0% vs. 0%, respectively; *t*(13) = 2.11, *p* = .05).

From the preceding sections on clarity of perception we can conclude that environmental context plays a subtle but definite role in affecting a listener's perceived meaning. Participants who perceived (categorized) the focus of the sentence to be "there" (due to gestural location) thought that such an emphasis was clearer (i.e., "made more sense") when there was a table present than when there was no suitable referent for the word "there." Furthermore, for those who perceived (categorized) "there" as the emphasis, more disagreed it was clear when there was no table compared with when there was a table. These results support our hypothesis that gestural emphasis in concert with relevant environmental context contributes significantly to the *meaningfulness* of perceived (categorized) speech.

The clarity ratings indicated that the perceived focus of the sentence was emphasized more clearly when the gesture was centered well within the PSR. If clarity ratings for “there” are placed in order of preference, most participants “agreed” that the focus of the sentence was clear, followed by “strongly agree” and then “neither agree nor disagree.” Provided the gesture was within the PSR, very few participants “disagreed” or “strongly disagreed” that the focus of the sentence was clear, especially near the center point of “there” PSR (Frame 43) (Figure 6). Subsequently there was a steady increase in responses that “disagreed” that the emphasis on “there” was clear as the gesture was moved toward the end of “there” PSR and was most dramatic for Frame 48 before falling on Frame 49 as perception (categorization) of “now” commenced. Overall, the ratings for clarity generally followed the profile of the categorization results. Participants thought that the emphasis perceived was also clear well before the onset of the speech signal and clarity of the word perceived decreased dramatically toward the end of the word’s PSR.

GENERAL DISCUSSION

The results lend support to the first hypothesis that the perceived focus of a sentence does depend on the timely coordination of a speaker’s gestures and speech. The categorization results clearly show that the perceived semantic focus of a sentence alters as the gestural location shifts along the speech signal. Concomitantly, the clarity ratings reinforce the conclusion that the certainty of what speech is perceived increases with appropriate synchronization between body movement and vocalization. The second hypothesis, that the perceived semantic clarity of a sentence depends on environmental context, also found considerable support. Throughout the clarity ratings data, the presence of a table led to significantly different ratings of clarity compared with when there was no table. For example, more agreed that the emphasis on “book” was clearer when the table was *absent* compared with when it was present. Symmetrically, more agreed that the emphasis on “there” was clearer when a table was *present* than when absent. Further, in the absence of any biasing gesture, participants still selected “there” as the focus of the sentence significantly more when there was a table than without a table. Concomitantly, more participants who perceived “there” as the focus of the sentence *disagreed* that the emphasis was clear when there was no table compared with when there was a table. In sum, our data on the perception of gestured speech indicates that the clarity of perception, that is, how much of that which is perceived “makes sense,” significantly depends on relevant environmental context.

A variety of implications follow from these results in terms of understanding how humans naturally communicate, how best to conceive of the dynamical and neural bases for communication, and how technology such as computer animation might facilitate communication. Regarding the latter, our results emphasize that if animators wish to add emphasis, then the mid-stroke of a gesture should be synchronous or even precede the acoustic body of the word uttered. The addition of appropriate speech-hand coordination could add realism and accuracy and hence increase the effectiveness of computer-animated characters and avatars for human computer interfaces and virtual environments (Gullberg & Holmquist, 1999, 2002; Rogers, 1978; Stanney, 2002).

Our results add to (and suggest gestural versions of) recent demonstrations of the surprising flexibility of synchrony of speech articulators and the associated acoustics as seen in classic McGurk effect experiments, where the acoustics can lag visual stimuli by as much as 180 ms (Munhall, Gribble, Sacco, & Ward, 1996; Munhall &Tohkura, 1998), and investigations of asynchrony between facial motion and acoustics (Abry, Lallouache, & Cathiard, 1996; Santi, Servos, Vatikiotis-Bateson, Kuratate, & Munhall, 2003; Yehia, Kuratate, & Vatikiotis-Bateson, 2002; Yehia, Rubin, & Vatikiotis-Bateson, 1998). The studies of orofacial motion in particular have emphatically demonstrated that the speech acoustics can be better estimated by the 3D dynamics of the face than by the midsaggital motion of the anterior vocal tract (i.e., lips, tongue, and jaw) (Yehia, Rubin, & Vatikiotis-Bateson, 1998).

Our results expand the phenomenon of phase flexibility to include speech-hand and speech-gesture coordination. A recent study has also shown that nonverbal gestural phenomena such as head movement (visual prosody) have a marked effect on perceived speech, lexical segmentation, and comprehension. It was found that speech amplitude and pitch correlated strongly with natural head movement (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004). Further, when using an animated speaker such that the normal synchrony of head motion was disassociated from speech, perceivers had greater difficulty in perceiving relevant aspects of speech. Such findings attest to the importance of nonverbal gestures for understanding face-to-face communication.

With regard to understanding specifically interpersonal speech-hand communication, previous research has shown that in the *production* of speech, co-occurring hand gestures tended to be initiated either prior to or simultaneously with the initiation of the relevant word spoken (Krauss, 1988). This was thought to reflect a process whereby the gesture enhances retrieval of the word to be spoken (i.e., the “lexical affiliate”) from memory. This fits in with the current results that show, symmetrically, that the *perception* of the intended focus of a sentence is strongly influenced by a gesture provided that the gesture is produced prior to or simultaneous with the utterance. Further research could explore possibilities of helping individuals with speech impediments such as stuttering by focusing more on gestural techniques to increase communicative skills (e.g., Mayberry & Nicholadis, 2000; Mayberry & Shenker, 1997; van Lieshout, 2004).

What mechanisms might provide a basis for speech-hand coordination? It has been argued that language evolved from simple primate gestures rather than from vocal origins (Corballis, 2002, 2003). The articulatory phonology approach pioneered by Browman and Goldstein (e.g., 1992), as well as the dynamical systems-based modeling approach known as task dynamics (Saltzman & Byrd, 2000), proposes that phonological units, “articulatory gestures,” serve a dual role as units of language production and language perception and correspond to the constriction actions of distinct vocal organs. The constrictions can be modeled as states of a dynamical system such that the motion event of articulator movement constitutes an articulatory gestural unit. This view of articulatory gestures as dynamical systems provides a foundation for conceiving how produced and perceived language forms might be complementary for a language user.

The notion of gestural articulatory dynamics as a basis for speech production and perception has been extended to include co-occurrent simple manual activity (Treffner & Peter, 2002). It was proposed that the coordination dynamics governing the timing patterns of the hands and vocal tract also provides a basis for the direct perception of a speaker’s linguistic intentions during communication. Experiments were conducted using a phase transition paradigm to examine the coordination of speech-hand gestures in both left-handed and right-handed individuals. In those experiments it was shown that patterns of speech-hand synchrony are determined by definite coordination dynamics that have been extensively investigated in a variety of biological coordination tasks over the last 2 decades (i.e., the Haken-Kelso-Bunz (HKB) model; e.g., Kelso, 1995). Results showed that in a simple monosyllabic speech-finger tap synchronization task that required either in-phase (e.g., /ba/ + tap…/ba/ + tap…, etc.) or anti-phase (e.g., /ba/…tap…/ba/ …tap…, etc.) coordination with a pacing metrononome, the asynchrony of the manual tap over the speech utterance (i.e., tap *preceded* speech) decreased as rate of production increased. That is, the kinematics of finger and jaw became increasingly coincident as performance rate increased. This behavior can be understood as a lawful consequence of the generic dynamics that account for such coordination behavior, in this case, a particular parameterization of the extended asymmetric HKB model (Treffner & Peter, 2002; Treffner & Turvey, 1995, 1996). Specifically, when a constant phase offset of 50º between finger and jaw was introduced into the model (together with inclusion of “attentional” factors, the *c*-term and *d*-term of the extended HKB equation), then the empirical data was reproduced—both its quantitative form and qualitative evolution under parametric change. We considered the constant “perceptual offset” of 50º to be consistent with evidence from speech perception experiments whereby individuals synchronized an external acoustic or motor event with the perceived centre of a syllable (i.e., the word’s “p-center” or perceptual center). The perceptual offset of 50º helps explain the large corpus of data on perceptual synchrony since, in our experiment, subjective, *perceived* synchrony between tap and speech (i.e., the in-phase task requirement) was intentionally achieved even though there was approximately a 50º *asynchrony* between tap and speech kinematics (Treffner & Peter, 2002). The gesture (finger tap) tended to *precede* speech (jaw opening), and the degree of anticipation of gesture over speech *decreased* as the rate of production *increased*. That is, synchrony increased as rate of coordination increased.

What then is the relation of an intrinsic perceptual phase offset in the nonlinear dynamics underlying the production of speech-hand coordination (Treffner & Peter, 2002) to the current results on relative synchrony during *perception* of speech-hand coordination? We believe the relation lies in how a varying rate of production alters speech-hand synchrony and the consequences of such synchrony on the accuracy of perception, that is, in appropriately comprehending a speaker's utterances. If speech-hand coordination patterns follow lawful regularities determined and predicted by the dynamics of rhythmic oscillatory systems (e.g., Treffner & Peter, 2002), and social coordination relies primarily upon physical dynamics rather than logical inferential mechanisms (Marsh, Richardson, Barron, & Schmidt, 2006), then it is reasonable to propose that speech-gesture coordination co-evolved as a contemporaneous mechanism to speech alone and that it could be harnessed by communicative dyads for purposes of emphasis and clarity. Reciprocally, the fact that individuals in face-to-face conversation often misunderstand one another follows from the negative effect that speech-hand *asynchrony* has on clarity of meaning. However, mechanisms alone cannot explain phenomenal meaning. But a coordination mechanism based on dynamics can entail semantic properties precisely because it is situated within an ecological context, the biologically relevant environment in which the perception-action cycles supporting communicative intentions evolved.

Communication would be more accurately approached as a biological or, more generally, an ecological phenomenon (Millikan, 1984; Treffner, 1999b) that is based on mechanisms not unlike the phase transition or resonance dynamics of between-person entrainment (Schmidt & Turvey, 1994; Treffner, 1999a; Treffner & Turvey, 1993). Taking a step further, social psychologists could emphasize more subtle qualities and constraints of the ecologically real social environment, so-called *values* such as "caring"; these must be satisfied by those engaged in successful communication and should be recognized and incorporated into a full-fledged pragmatic theory of language (Hodges, 2007). Communication thus understood *instantiates a mechanism* (of resonance dynamics—the "syntax") that *entails awareness* (of affordances—the "semantics"). This model replaces more conventional assumptions of the transmission and reception of semantics-free symbols that require (somehow) interpretation. Such semantic embellishment and integration into meaningful "mental concepts" is, however, theoretically untenable. Such a rebuff of information-processing approaches that purport to explain language perception is not new (Reed, 1996; Shaw et al., 1981; Turvey et al., 1981), but it remains a daunting challenge for ecological scientists to provide sufficiently convincing empirical evidence to dislodge the currently accepted dogmas of seductive representationalism and symbolic computational wizardry.

But why did we then choose to use computer animation to address issues of natural communication? By definition, every experimental manipulation results in a reduction in richness and, unfortunately, sometimes a distortion of the ecological information available in natural situations. The goal is always to minimize the impact of such intervention, minimize (or avoid) the distortion, and yet distill and describe the natural dynamics at play (Kelso, 1995). From a Gibsonian standpoint, we support the emphasis on "ecological validity" in experimental design that has, in part, emerged from the ecological psychology of James Gibson and his associates (e.g., Shaw et al., 1981). We have attempted to create an experimental design in this vein (i.e., keep and manipulate the essential information and dispense with the nonessential information). The current work is an attempt to investigate natural language communication by altering the stimulus display in relevant ways rather than impoverishing it. The phenomenon of natural language involves more than vocal, body, and environmental components. It is somehow the spatiotemporal mixture of all these facets that permits successful and meaningful exchange between organisms of their perceptions, viewpoints, thoughts, and intentions.

What relation, if any, might exist between the concept of speech-hand coordination as based in dynamics, and contemporary research into the neural correlates of linguistic ability? Recently, there has been growing evidence that neural systems involved in speech perception are tightly coupled with those involved in speech production, or more generally, that the observation and execution of action share a common neural substrate. It has been reported that both listening to speech and observing speech-related lip movements increased the excitability of motor units underlying speech production (Watkins, Strafella, & Paus, 2003), with increased excitability pronounced in the left hemisphere. Similarly, a significant increase in motor evoked potentials was recorded from the tongue in response to transcranial magnetic stimulation of the left primary motor cortex when listening to speech containing consonants for whose production tongue movements were required (Fadiga, Craighero, Buccino, & Rizzolatti, 2002). These findings are consistent with those of Haueisen and Knosche (2001), who demonstrated involuntary neural activity above the primary motor hand area in pianists listening to music. Observations such as those of Floel, Ellger, Breitenstein, and Knecht (2003), which show that linguistic tasks such as speech production and perception activate hand motor cortex, and those of Meister et al. (2003), who showed that during reading aloud there is an increase in excitability of hand motor cortex of the language-dominant hemisphere, add to a body of evidence that indicates a functional connection between the motor area for manual activity of the language-dominant hemisphere and regions subserving language comprehension. Indeed, it was observed that hand gestures influenced acoustical encoding of co-produced speech. This suggests that hand gestures are integrated with speech (at least in a speaker) at an early stage of language processing (Kelly, Kravitz, & Hopkins, 2004).

All these observations support theories that favor a strong evolutionary and functional link between manual gestures and the development of speech and language. According to such theories, the evolution of language is based on a neural motor system that specializes in action recognition. Manual gestures, as a form of goal-directed action, are well disposed for conveying meaning to such an action-attuned perceptual system. The discoveries of Ferrari, Gallese, Rizzolatti, and Fogassi (2003) of new categories of "mirror" neurons in F5 in monkey cortex included mouth mirror neurons and a population of hand mirror neurons named audiovisual mirror neurons that became active when the monkey performed a hand action or when it only heard the action-related sound. Indeed, Taira et al. (1990) argued that cells of parietal cortex are responsive to the affordances for grasping (their term) available in the visual system and in turn transmit this activity to F5. A laudable attempt to create a neural network framework incorporating action, mirror neurons, and perception (of affordances) is provided by Arbib and colleagues (Arbib, 2005; see also commentaries). Further understanding of how a transition from a system for basic action recognition might have become equipped for language is provided by Kohler et al. (2002). The fact that audiovisual mirror neurons are a subcategory of hand mirror neurons offers an insight into language evolution and the strong coupling between manual gestures, sound, and ultimately, articulatory gestures. However, the evidence for an action-based perceptual system for language that is provided by data on mirror neurons is not unique to monkey. In an fMRI study in humans, it was shown that a participant's observation of actions performed with the hand, mouth, or foot led to the activation of distinctly different parts of Broca's area and offers further support for the existence of neural correlates of a speech-gesture coordination system in humans (Buccino, Binkfski, & Riggio, 2004).

Although we take such results on the neural mechanisms of speech-hand coordination as confirmation of a deeply seated relation between perception and action, such neural mechanisms are not by themselves explanatory. We hope that the present results help foster mutual progress in both relatively macroscopic (ecological/information-based) and relatively microscopic (neural/material-based) approaches to an emerging multidisciplinary science of information-based behavior. Furthermore, communication is necessarily a social phenomenon that occurs between persons and thus involves interpersonal information; neural measurements cannot delineate the specificational information about affordances at the ecological scale to which communicative acts often refer. Similarly, symbolic descriptions at the cognitive (representational) scale are inadequate for capturing the subtleties of perception-action behavior (Borghi, 2004; Shaw, 2001; Shaw et al., 1981; Turvey et al., 1981).

We hope that our work will further understanding of the complex temporal relations between gesture and speech. As one of the pioneers of gesture research, Adam Kendon, has implored, “Further studies of how gesture contributes to understanding in interaction are very much needed” (Kendon, 2004, p.94). However, in order to understand how gestures facilitate meaningful communicative events, a spotlight should also be shone on the affordance structure of the environment and how the coordination dynamics and associated rhythms of perceiver-actors interface with such environments. Anything less runs the risk of assumed mechanisms with little regard for context and binding relations. Nonverbal communication has the advantage of extending beyond the ephemeral temporal domain and into the domain of contextually situated, spatiotemporal, visible *events*. “Events are perceivable but time is not” (Gibson, 1975). Given the temporally extended dynamics of behavioral interaction (Schmidt, 2007; Treffner & Kelso, 1999), the challenge, it seems, is to harness technology in such a way that it does not obscure or distort the fundamental event-based nature of ecological social phenomena. If the mechanism of spatiotemporal event perception is dynamics-based and not memory-based (Treffner & Kelso, 1999), then of particular interest is how the precise timing relations we have shown might be harnessed for the benefit of emerging human interface technologies that exploit computer animation, social interaction in virtual environments, and animation-based provision of information such as via animated characters (avatars). The current research shows that clarity of exposition can be enhanced by the appropriate *phasing* of body gestures and speech, either with the accompanying gesture phase-advanced or approximately synchronous with speech. Instructional learning via displays of avatars or virtual instructors would benefit by increasing the prominence of gesture in actual or virtual instructors since it has been shown that accompanying gesture can significantly improve the acquisition of concepts in children (Singer & Goldin-Meadow, 2005). Conversely, gesturing with the requisite phase lag between speech and associated gesture is likely to result in lesser clarity. Interestingly, if intentionally created, such phase-lagged gestures could be exploited to a speaker’s advantage by increasing the ambiguity or “open-endedness” of an utterance, or even to intentionally obfuscate, confuse, and confound a listener, or to conceal a speaker’s true intentions. Whether avatars will remain as explicitly programmed animations produced by the rather old-fashioned explicit key framing method (as were the present simulations), or whether they will evolve (self-organize) into surprisingly naturalistic renditions of biological motion, perhaps using the wealth of research on biological coordination dynamics that has been produced by dynamics-inspired ecological psychologists, remains to be seen. The latter is surely inevitable.

The current results and those of our previous work on simultaneous tapping and babbling (Treffner & Peter, 2002) show that the inherent phase-lag in speech-hand dynamics can be entrained by an intentional communicative system to the mutual advantage of those involved in a dialogue (see also Pettito et al., 2004). We believe the perceived synchrony of gestures and words is based on a flexible spatiotemporal dynamics around which the binding of speech and movement, and its grounding in the environment, can occur in a meaningful or ecologically relevant way. This binding and grounding constitutes the speaker's intentions such that they can be presented to (not represented in) an appropriately attuned listener who can then effectively reach out, "grasp," and understand the intended message. Under such ecological conditions, a listener can experience the direct perception of meaning.

ACKNOWLEDGMENTS

Results were first presented at *Symmetry and Duality: Principles for an Ecological Psychology,* a Festschrift in honor of Bob Shaw and his contributions to experimental psychology, which occurred at the University of Connecticut, June 2004. Thanks to Eric Vatikiotis-Bateson, Pascal van Lieshout, and an anonymous reviewer for their helpful comments on the manuscript.

REFERENCES

Abry, C., Lallouache, M.-T., & Cathiard, M.-A. (1996). How can coarticulation models account for speech sensitivity to audio-visual desynchronization? In D. Stork & M. Hennecke (Eds.), *Speechreading by humans and machines: Vol. 150* (pp. 247–256). Berlin: Springer-Verlag.

Alibali, M. W., & DiRusso, A. A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development, 14*, 37–56.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, 28,* 105–124.

Blake, J., Olshansky, E., Vitale, G., & Macdonald, S. (1997). Are communicative gestures the substrate of language? *Evolution of Communication*, *1,* 261–282.

Borghi, A. M. (2004). Object concepts and action: Extracting affordances from objects parts. *Acta Psychologica, 115,* 69–96.

Browman, C.P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49,* 155–180.

Buccino, G., Binkfski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language*, *89,* 370–376.

Corballis, M. (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.

Corballis, M. (2003). Laterality and human speciation. In T. J. Crow (Ed.), *Speciation of modern Homo sapiens. Proceedings of the British Academy*, *106,* 137–152.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15,* 399–402.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions. *European Journal of Neuroscience*, *17,* 1703–1714.

Floel, A., Ellger, T., Breitenstein C., & Knecht, S. (2003). Language perception activates the hand motor cortex: Implications for motor theories of speech perception. *European Journal of Neuroscience*, *18,* 704–708.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3–28.

Furuyama, N. (2002). Prolegomena of a theory of between-person coordination of speech and gesture. *International Journal of Human-Computer Studies, 57,* 347–374.

Gibson, J. J. (1975). Events are perceivable but time is not. In J. T. Fraser & N. Lawrence (Eds.), *The study of time II* (pp. 295–301). Berlin: Springer-Verlag.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. (Original work published 1979)

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences, 3,* 419–429.

Gullberg, M., & Holmquist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics and Cognition, 7,* 35–63.

Gullberg, M., & Holmquist, K. (2002). Visual attention towards gestures in face-to-face interaction vs. on screen. In I. Wachsmuth & T. Sowa (Eds.), *Gesture and sign language based human-computer interaction* (pp. 206–214). Berlin: Springer.

Haueisen, J., & Knosche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience, 13,* 786–792.

Hodges, B. H. (2007). Values define fields: The intentional dynamics of driving, carrying, leading, negotiating, and conversing. *Ecological Psychology, 19,* 153–178.

Iverson, J. M. & Goldin-Meadow, S. (1997). What's communication got to do with it? Gesture in congenitally blind children. *Developmental Psychology, 33*, 453–467.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89,* 253–260.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behaviour.* Boston: MIT Press.

Kendon, A. (1972). Some relationships between body motion and speech. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication*. New York: Pergamon Press.

Kendon, A. (1974). Movement coordination in social interaction: Some examples described. In S. Weitz (Ed.), *Nonverbal communication*. New York: Oxford University Press.

Kendon, A. (2004). Review of Susan Goldin-Meadow (2003), Hearing gesture: How our hands help us think. *Gesture, 4,* 91–107.

Kimura, D. (1973). Manual activity during speaking–I. Right handers. *Neuropsychologia, 11,* 45–50.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297,* 846–848.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7,* 54–59.

Lausberg, H., & Kita, S. (2003). The content of the message influences the hand choice in co-speech gestures and in gesturing without speaking. *Brain and Language, 86,* 57–69.

Marsh, K. L., Richardson, M. J., Barron, R. M., & Schmidt, R. C. (2006). Contrasting approaches to perceiving and acting with others. *Ecological Psychology, 18,* 1–38.

Mayberry, R., & Jaques, J. (2000). Gesture production during stuttered speech: Insights into the nature of gesture-speech integration. In D. McNeill (Ed.), *Gesture and language* (pp. 199–213). Cambridge, UK: Cambridge University Press.

Mayberry, R. I., & Nicholadis, E. (2000). Gesture reflects language development: Evidence from bilingual children. *Current Directions in Psychological Science, 9,* 192–196.

Mayberry, R. I., & Shenker, R. C. (1997). Gesture mirrors speech motor control in stutterers. In W. Hulstijn, H. Peters, & P. van Lieshout (Eds.), *Speech motor production and fluency disorders* (pp. 183–190). Amsterdam: Elsevier Science.

McClave E. (1994). Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research, 23,* 45–66.

McClave E. (1997). Pitch and manual gestures. *Journal of Psycholinguistic Research, 27,* 69–89.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92,* 350–371.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

McNeill, D. (Ed.). (2000). *Language and gesture*. Cambridge, UK: Cambridge University Press.

McNeill, D. (2005). Gesture-first, but no gestures? *Behavioral and Brain Sciences, 28*, 138–139.

Meister, I. G., Boroojerdi, B., Foltys, H., Sparing, R., Huber, W., & Töpper, R. (2003). Motor cortext hand area and speech: Implications for the development of language. *Neuropsychologia, 41,* 401–406.

Millikan, R. G. (1984). *Language, thought and other biological categories.* Cambridge, MA: MIT Press.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts the temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 615–623.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics, 58,* 351–362.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science, 15,* 133–137.

Munhall, K. G., &Tohkura, Y. (1998). Audiovisual gating and the time course of speech perception. *Journal of the Acoustical Society of America, 104,* 530–539.

Nobe, S. (2000). Where do *most* spontaneous representational gestures actually occur with respect to speech? In D. McNeill (Ed.), *Language and gesture* (pp. 186–198). Cambridge, UK: Cambridge University Press.

Pagano, C. C., & Turvey, M. T. (1995). The inertia tensor as a basis for the perception of limb orientation. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1070–1087.

Pettito, L. A., Holowka, S., Sergio, L. E., Levy, B., & Ostry, D. J. (2004). Baby hands that move to the rhythm of language: Hearing babies acquiring sign languages babble silently on the hands. *Cognition, 93,* 43–73.

Reed, E. S. (1996). *Encountering the world: Towards an ecological psychology.* New York: Oxford University Press.

Rogers, W. (1978). The contribution of kinesic illustrators towards the comprehension of verbal behaviour within utterances. *Human Communication Research, 5,* 54–62.

Saltzman, E. L., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science, 19,* 499–526.

Santi, A., Servos, P., Vatikiotis-Bateson, E., Kuratate, T., & Munhall, K. (2003). Perceiving biological motion: Dissociating visible speech from walking. *Journal of Cognitive Neuroscience, 15,* 800–809.

Schmidt, R. C. (2007). Scaffolds for social meaning. *Ecological Psychology, 19,* 137–151.

Schmidt, R. C., & Turvey, M. T. (1994). Phase-entrainment dynamics of visually coupled rhythmic movements. *Biological Cybernetics, 70,* 369–376.

Shaw, R. E. (2001). Processes, acts, and experiences: Three stances on the problem of intentionality. *Ecological Psychology, 13,* 275–314.

Shaw, R. E. (2003). The agent-environment interface: Simon's indirect or Gibson's direct coupling? *Ecological Psychology, 15,* 37–106.

Shaw, R. E., Turvey, M. T., & Mace, W. M. (1981). Ecological psychology: The consequences of a commitment to realism. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes: Vol. II* (pp. 159–226). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16,* 85–89.

Sousa-Poza, J. F., Rohrberg, R., & Mercure, A. (1979). Effects of type of information (abstract-concrete) and field dependence on asymmetry of hand movements during speech. *Perceptual and Motor Skills, 48,* 1323–1330.

Stanney, K. (2002). *Handbook of virtual environments*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Taira, M., Mine S., Georgopoulos, A. P., Murata, A., & Sakata, H. (1990). Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Experimental Brain Research, 83*, 29–36.

Treffner, P. J. (1999a). Resonance constraints on between-person polyrhythms. In M. A. Grealy & J. A. Thomson (Eds.), *Studies in perception and action V* (pp. 165–169). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Treffner, P. J. (1999b). The common structure of concepts is the affordance in the ecology. Commentary on Ruth Millikan's "A common structure of individuals, stuffs, and real kinds." *Behavioral and Brain Sciences, 22,* 729–733.

Treffner, P. J., & Barrett, R. (2004). Hands-free mobile phone speech while driving degrades coordination and control. *Transportation Research, Part F: Traffic Psychology and Behaviour, 7,* 229–246.

Treffner, P. J., Barrett, R., & Petersen, A. J. (2002). Stability and skill in driving. *Human Movement Science, 21,* 749–784.

Treffner, P. J., & Kelso, J. A. S. (1999). Dynamic encounters: Long-memory during functional stabilization. *Ecological Psychology, 11,* 103–137.

Treffner, P. J., & Peter, M. (2002). Intentional and attentional dynamics of speech-hand coordination. *Human Movement Science, 21,* 641–697.

Treffner, P. J., & Turvey, M. T. (1993). Resonance constraints on rhythmic movement. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1221–1237.

Treffner, P. J., & Turvey, M. T. (1995). Handedness and the asymmetric dynamics of bimanual rhythmic coordination. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 318–333.

Treffner, P. J., & Turvey, M. T. (1996). Symmetry, broken symmetry, and the dynamics of bimanual coordination. *Experimental Brain Research, 107,* 463–478.

Turvey, M. T., Shaw, R. E., Reed, E., & Mace, W. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). *Cognition, 9,* 237–304.

van Lieshout, P. H. H. M. (2004). Dynamical systems theory and its application in speech. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 313–356). Oxford, UK: Oxford University Press.

van Lieshout, P. H. H. M., Hulstijn, W., & Peters, F. M. (1996). From planning to articulation in speech production: What differentiates a person who stutters from a person who does not stutter? *Journal of Speech and Hearing Research, 39,* 546–564.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*, 989–994.

Wilson, A. D., Bingham, G. P., & Craig, J. C. (2003). Proprioceptive perception of phase variability. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 1179–1190.

Yehia, H. C., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion, and speech acoustics. *Journal of Phonetics, 30,* 555–568.

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behaviour. *Speech Communication, 26,* 23–43.