Psychonomic Bulletin & Review
2006, 13 (3), 361-377

THEORETICAL AND REVIEW ARTICLES

The motor theory of speech perception reviewed

BRUNO GALANTUCCI
Haskins Laboratories, New Haven, Connecticut
and University of Connecticut, Storrs, Connecticut

CAROL A. FOWLER
Haskins Laboratories, New Haven, Connecticut,
Yale University, New Haven, Connecticut,
and University of Connecticut, Storrs, Connecticut

and

M. T. TURVEY
Haskins Laboratories, New Haven, Connecticut
and University of Connecticut, Storrs, Connecticut

More than 50 years after the appearance of the motor theory of speech perception, it is timely to evaluate its three main claims that (1) speech processing is special, (2) perceiving speech is perceiving gestures, and (3) the motor system is recruited for perceiving speech. We argue that to the extent that it can be evaluated, the first claim is likely false. As for the second claim, we review findings that support it and argue that although each of these findings may be explained by alternative accounts, the claim provides a single coherent account. As for the third claim, we review findings in the literature that support it at different levels of generality and argue that the claim anticipated a theme that has become widespread in cognitive science.

The motor theory of speech perception (see, e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985) is among the most cited theories in cognitive psychology.[1] However, the theory has had a mixed scientific reception. On the one hand, it has few proponents within the field of speech perception, and many authors cite it primarily to offer critical commentary (e.g., Sussman, 1989). On the other hand, it is perhaps the only theory of speech perception recognized outside the field of speech, and there, its reception has been considerably more positive (e.g., Rizzolatti & Arbib, 1998; Williams & Nottebohm, 1985).

With the deaths of Alvin Liberman and Ignatius Mattingly, the two investigators who contributed most to the development of the theory, it is timely to review its main claims in order to determine which of them should be set aside and which deserve further consideration. The three main claims of the theory are the following: (1) Speech processing is special (Liberman & Mattingly, 1989; Mattingly & Liberman, 1988); (2) perceiving speech is perceiving vocal tract gestures[2] (e.g., Liberman & Mattingly, 1985); (3) speech perception involves access to the speech motor system (e.g., Liberman et al., 1967).

We will argue that the first claim is difficult to evaluate because it has several readings: (1a) that speech perception is special with respect to audition, in that its objects are not the proximal acoustic patterns but the distal gestures that generated the acoustic patterns; (2a) that speech is special with respect to audition, in that it implies recruitment of the motor system in perception; and (3a) that speech is produced and processed by a piece of neural circuitry that represents a specialization in the biological sense. We will argue that unless (1a) and (2a) are interpreted very narrowly, they are disconfirmed by the available evidence and that evidence for (3a) is difficult to obtain. However, we will argue that (2) and (3), the most radical claims of the theory, should not be dismissed.

As for (2), we will review some of the evidence relevant to the claim and argue that although each piece of evidence can be individually explained by alternative accounts, these accounts differ for each piece of evidence, whereas (2) provides a single coherent account of all of the findings.

As for (3), the core claim of the theory, we will review an extensive body of evidence compatible with the claim.

Preparation of this article was supported by NICHD Grant HD-01994 and NIDCD Grant DC-03782 to Haskins Laboratories. We thank L. Goldstein, R. Millikan, D. Shankweiler, M. Studdert-Kennedy, and D. Whalen for helpful discussions and comments. I. Mattingly provided helpful comments and discussion at an early stage of preparation of the manuscript. C. Carello provided the artwork in Figure 3. Correspondence concerning this article should be addressed to B. Galantucci, Haskins Laboratories, 300 George St., New Haven, CT 06511 (e-mail: bruno.galantucci@haskins.yale.edu).

Our review will cover findings of perception–motor links, both at the neural and at the behavioral levels, in domains of increasing generality. We will begin with the specific domain of speech perception, then will consider the more general domain of animal communication, and finally, will move beyond communication, covering the domain of perception in general. In this context, we will review both findings that motor competence[3] is accessed in perception and findings that the motor system itself is involved in perceptual tasks. Our conclusion will be that the evidence supports (3) in its most general sense.

This article has five sections. In the first section, we will describe the successive versions of the motor theory of speech perception, illustrating how the theory has undergone progressive abstraction and has progressively placed more emphasis on speech as an evolutionary adaptation special to humans. In the next three sections, we will discuss each of the three main claims of the theory individually. In the final section, we will offer a summary evaluation of the motor theory of speech perception and remark that the core lesson learned by Liberman and colleagues from half a century of empirical research is, in the end, rather simple: Cognition, like any product of evolution, cannot be understood in isolation but needs to be understood as embedded in a meaningful ecological context and embodied in biologically plausible perceiving–acting systems. We will conclude by arguing that the pursuit of such a broad scientific perspective places the motor theory of speech perception in close connection with much of the theorizing and research in contemporary cognitive science.

THE MOTOR THEORY OF SPEECH PERCEPTION

Speech Perception as Association Learning

Liberman and colleagues first developed their motor theory (Liberman, 1957; Liberman, Delattre, & Cooper, 1952) to explain some surprising experimental findings. The experiments had been stimulated by an unexpected failure of a reading machine, intended for the blind, in which an acoustic alphabet composed of arbitrary sounds substituted for an orthographic one (see Liberman, 1996, chap. 1). The failure of participants to learn appeared to be due to their inability to perceive alphabetic sequences at practically useful rates. At those rates, the participants could not identify the individual sounds in the sequence; rather, the sounds merged into a holistic blur. The problem, it seemed to Liberman and colleagues, was that the sequences of discrete sounds exceeded the temporal resolving power of the auditory system. This outcome led to the question of why listeners can perceive speech at the rapid rates that they do and to research designed to reveal how the acoustic signal encodes the consonants and vowels of spoken words.

Liberman and colleagues (Liberman, 1957, 1996; Liberman et al., 1952; Liberman, Delattre, Cooper, & Gerstman, 1954) used the sound spectrograph to explore the acoustic structure of speech and the Pattern Playback[4] to test their hypotheses about how the acoustic signal specifies the phonetic segments of syllables, words, and sentences. In this way, they discovered that phonetic segments are coarticulated. That is, vocal tract gestures for successive consonants and vowels overlap temporally. For Liberman, coarticulation was a very important feature of speech because, if information for the phonetic segments overlaps, information for each segment can span a longer interval of time, and the ear can resolve the segments temporally. In Liberman's words, speech is not an acoustic "alphabet" or "cipher," but an efficient "code" (Liberman et al., 1967).

However, coarticulation has other consequences as well: The acoustic speech signal is highly context sensitive, and it has no discrete phone-sized segmental structure. For example, Liberman and colleagues (Liberman et al., 1952; Liberman et al., 1954) found that both of the main acoustic cues for stop consonants—the bursts of energy generated during release of the stop constriction and the formant transitions that occur as the constriction for the consonant gives way to that for the following vowel—are highly context sensitive.

As for energy bursts, Liberman et al. (1952) found that an invariant burst, centered at 1440 Hz, sounded like /p/ before the vowels /i/ (*ee*) and /u/ (*oo*) but like /k/ before /a/ (*ah*). Liberman et al. (1952) recognized that due to coarticulation, getting a stop burst centered at 1440 Hz required a labial constriction before /i/ and /u/ but a velar constriction before /a/. In other words, information about coarticulation caused the same bit of acoustic signal to be identified as different phonetic segments.

As for formant transitions, Liberman et al. (1954) found that the second formant transitions of the two-formant synthetic syllables /di/ and /du/ are markedly different (see Figure 1). The transition of /di/ is high and rising to the level of the high second formant for /i/. That for /du/ is low and falling to the low level of the second formant for /u/. In these synthetic syllables, there is no invariant acoustic structure for the /d/s, but they sound alike. Liberman (e.g., 1957) recognized that there was something common to the two /d/s—namely, the way in which they are produced. Both are produced by a constriction gesture of the tongue tip against the alveolar ridge of the palate. Whereas, in the /pi/–/ka/–/pu/ example, the same bit of acoustic signal—which, due to coarticulation, has to be produced by different constriction gestures—causes different percepts in the context of different vowels, in the /di/–/du/ example, acoustically very different transitions underlie an invariant percept when the underlying consonantal articulation is the same.

These findings led to a generalization. When acoustic patterns are different but the articulatory gestures that would have caused them in natural speech are the same, or vice versa, perception tracks articulation (Liberman, 1957).

The earliest version of the motor theory reflected Liberman's training as a behaviorist (see Liberman, 1996). Liberman (1957) proposed that infants mimic the speech they hear and that this leads to associations between articulation and its sensory consequences, on the one hand,

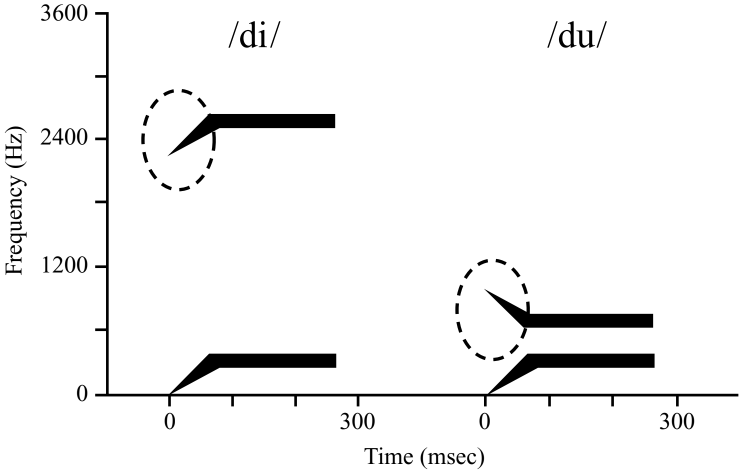

Figure 1. Spectrographic patterns for the two two-formant synthetic syllables /di/ and /du/. Note the difference in formant transitions, marked by the dotted ovals.

and the acoustic signals that the movements generate, on the other hand. Accordingly, for the syllables /di/ and /du/, a tongue tip constriction and its sensory consequences are associated with the context-sensitive acoustic information for /d/ in each syllable. Perhaps by a process of acquired similarity, whereby associating different acoustic signals for the syllables to the same response makes the syllable-initial consonants sound alike, the /d/s come to sound the same to listeners (cf. Lawrence, 1949, 1950). Another process, of acquired distinctiveness, may explain how similar acoustic signals (such as similar energy bursts) with different underlying articulations come to sound distinct. Eventually, overt mimicry is short-circuited. For skilled perceivers, the consequence of the memory representations established by mimicry is that "the articulatory movements and their sensory effects mediate between the acoustic stimulus and the event we call perception" (Liberman, 1957, p. 122).

As we will see in the next section, this version of the motor theory did not endure.

Speech Perception as a Human Evolutionary Achievement

Beginning in the 1960s, there was a paradigm shift in experimental psychology from behaviorism to cognitive psychology, and language was at its core (Chomsky, 1959; Skinner, 1957). Consistent with this paradigm shift, two lines of research began to challenge the association-learning mechanism invoked by Liberman (1957). On the one hand, there was research on animal communication that demonstrated the existence of many species-typical innate communicative behaviors (e.g., Nottebohm, 1970). On the other hand, there was research on prelinguistic infants that demonstrated that babies were able to detect most phonetic contrasts at birth (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971).[5] Influenced by these findings (see Liberman, 1996, chap. 1), Liberman and colleagues changed their explanation for the motoric nature of the speech percept. Whereas the explanation offered in the 1950s invoked ontogenetic learning processes (Liberman, 1957), the new account invoked phylogenetic adaptations unique to our species (Liberman et al., 1967). In particular, Liberman et al. (1967) argued that the adaptations consisted of the skills to coarticulate speech and to perceive coarticulated speech. Given that neither skill would be useful without the other, they concluded that the two skills had to have coevolved. To Liberman and colleagues, the simplest way to ensure the coevolution of the two skills was for them to be linked via a single mechanism (cf. Alexander, 1962, whose similar argument will be presented below). Liberman et al. (1967) suggested either of two possibilities for such a mechanism. They proposed analysis by synthesis (cf. Halle & Stevens, 1962; Stevens, 1960), in which the listener analyzed the acoustic input, guessed at how it was produced by the speaker, synthesized a virtual acoustic signal based on the guess, and matched the virtual to the actual signal. Given a sufficiently close match, the listener achieved a percept that corresponded to the invariant motor commands sent to the musculature underlying the vocal tract actions that produced the acoustic signal. These, like articulation and its sensory consequences in the early motor theory, were presumed to be invariant for a given phoneme. A second, undeveloped proposal was that neural networks for production and perception allowed for crosstalk (cf. Liberman, Cooper, Studdert-Kennedy, Harris, & Shankweiler, 1968).

Speech Perception as the Output of a Phonetic Module

The notion of *module*, introduced in Fodor's (1983) *Modularity of Mind*, offered a new context within which the speech mechanism proposed by Liberman and colleagues in 1967 gained credibility (Liberman & Mattingly, 1985). A module, in Fodor's view, is a neural system forged by evolution to perform a specific task that requires *eccentric* processing—that is, processing special to its particular domain. According to Fodor, there are many such modules, including, for example, neural systems underlying sound localization in audition and depth and color perception in vision. The speech mechanism proposed in the 1960s by Liberman et al. (1967) was now seen by Liberman and Mattingly (1985), as well as by Fodor, as a *phonetic module*—that is, as just another example of a common kind of system.

Findings of duplex perception (Mann & Liberman, 1983; Whalen & Liberman, 1987) provided evidence for a phonetic module. In one version of the finding, most of a three-formant syllable (the *base*) is presented to the listener's left ear. The remaining part of the syllable—either of two third-formant transitions that, when integrated with the base, sound like /da/ or /ga/—is presented to the right ear. Listeners' perceptions are duplex. They hear unambiguous /da/ or /ga/ in the left ear, and they hear the transition as a nonspeech "chirp" in the right ear, a percept like the one elicited by the third-formant transition when presented alone. Moreover, when a continuum of transitions between those for /da/ and /ga/ is presented in the ear opposite the base, speech discriminations look categorical; chirp discriminations do not. A finding that the same acoustic fragment is perceived in two ways at the same time suggested to Liberman and colleagues that two perceptual systems underlie the distinct percepts (but see Fowler & Rosenblum, 1990, for an alternative explanation). That one percept is phonetic and the other is *homomorphic*[6] (Liberman & Mattingly, 1989) with the acoustic signal suggests that one of these perceptual systems is the phonetic module; the other is the auditory system.

In the modular version of their theory, Liberman and Mattingly (1985) retained from the previous version the idea that the objects of speech perception are motoric, but they developed a new understanding of these objects.

In the 1967 version of the theory, the motoric objects were understood by Liberman and colleagues as invariant motor commands to muscles that moved anatomical structures—that is, the individual vocal tract articulators. However, this idea faced two challenges. First, there was evidence for considerable context sensitivity in the muscle activity driving the articulators' movements and, by implication, in the motor commands that underlay that activity (e.g., MacNeilage, 1970). Second, three related theoretical developments made the idea that the phonetic invariances were to be found in motor commands to anatomical structures ever more implausible. Turvey (1977) developed a theory of action in which the motor system was to be understood in terms of functional units—called *coordinative structures* (cf. Easton, 1972)—rather than anatomical structures. Next, Fowler, Rubin, Remez, and Turvey (1980) extended Turvey's theory to the domain of speech production. Finally, Browman and Goldstein's *articulatory phonology* (Browman & Goldstein, 1986) identified coordinative structures as fundamental linguistic units that they called *phonetic gestures* (see note 2).

Liberman and Mattingly (1985) adopted Browman and Goldstein's (1986) phonetic gestures as motoric objects of speech perception, but, due to their belief that coarticulation irreparably distorts gestures when they are implemented in the vocal tract, they were forced to make a distinction between the gestures *intended* by the speaker at a prevocal, linguistic level and the actual movements that occur in the speaker's vocal tract. And given that, for speech to serve its communicative function, speakers and listeners must converge on one and the same linguistic currency (see the Speech Perception as Parity Preserving section below). Liberman and Mattingly identified intended gestures, and not actual vocal tract actions, as the fundamental objects of speech perception.[7]

Speech Perception as Parity Preserving

A theoretical concept that Mattingly and Liberman developed in the last years of their theorizing was that of *parity* (Mattingly & Liberman, 1988). In its last formulation, this time by Liberman and Whalen (2000), the notion of parity was interpreted in three ways. One is that listeners and talkers have to converge on what counts as a linguistic action. As Liberman and Whalen put it, "/ba/ counts but a sniff does not" (p. 189). A second is that prototypically, phonetic messages sent and received must be the same. The third one is that the production and perception specializations for speech must have coevolved and have done so because, for Liberman and colleagues, they are most likely one and the same specialization.[8]

Liberman and colleagues recognized that parity is a very important constraint on the design of language: If it is broken, speech does not serve its communicative function. For this reason, Liberman (1996) adopted parity as a theoretical touchstone. Any theory of speech must explain how the parity requirement is met.

SPEECH PROCESSING IS SPECIAL

As we noted above, the evaluation of the claim that speech processing is special is difficult because the claim has at least three readings, which, for convenience, we will reiterate here: (1a) is that speech perception is special with respect to audition in that its objects are not the proximal acoustic patterns; (2a) is that speech is special with respect to audition in that it implies recruitment of the motor system in perception; and (3a) is that speech is produced and processed by a piece of neural circuitry that represents a specialization in the biological sense.

To complicate matters, the three readings are difficult to interpret themselves, primarily because the term *special* is ambiguous. For example, (1a) cannot be interpreted as meaning that speech perception is the *only* auditory process whose objects are distal properties, because, as Liberman and colleagues understood (e.g., Liberman & Mattingly, 1985), there are other well-known auditory processes whose objects are distal properties, such as processes underlying sound localization.[9] However, if we take (1a) in a weaker sense, as meaning that speech perception is *one* of the auditory processes whose objects are distal properties, the aptness of the term *special* seems to depend on how common the state of affairs described by the claim is. Recent evidence suggests that listeners quite generally perceive distal properties auditorily (Carello, Anderson, & Kunkler-Peck, 1998; Carello, Wagman, & Turvey, 2005; Kunkler-Peck & Turvey, 2000).[10] Moreover, even the best evidence in favor of (1a), duplex perception (Mann & Liberman, 1983; Whalen & Liberman, 1987), has been replicated in a domain (perception of slamming doors) in which inferences from duplex perception to a special perceiving mechanism are highly implausible (Fowler & Rosenblum, 1990).

(2a) is also open to different interpretations. The claim can be read in a strong sense, to mean that audition is the *only* perceptual system that implies recruitment of the motor system. Or it can be read in a weaker sense, to mean that speech perception is special because, *within audition*, it is the only process that implies recruitment of the motor system. There is plenty of evidence that in its strong sense, (2a) is false. In fact, as we will show below, motor recruitment in perception is general and widespread. Moreover, some of the evidence concerns auditory processes that have nothing to do with speech (Kohler et al., 2002), suggesting that (2a) is likely to be false even in its weaker form.

As for (3a), we will not take a stand here on whether there is a bit of neural hardware specifically dedicated to speech. In our view, the only evidence that would support (3a) in full would be findings that some circuit of the nervous system is active *if and only if* speech is perceived or produced. At the moment, such evidence is difficult to obtain. Research conducted under a weaker criterion—that is, a finding that some circuit of the nervous system is active *if* speech is either perceived or produced—suggests that speech is processed by a neural circuit different from the circuit that processes nonvocal sounds (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Whalen et al., 2006). However, the neural circuit that processes speech seems to be the same as the one that processes other *nonphonetic* vocal sounds (e.g., laughs or coughs). In other words, (3a) may also be false; the neural specialization for speech, if it exists, may be for vocal sounds in general (or even more generally, for the acoustic consequences of action; Kohler et al., 2002), not for speech per se.[11]

PERCEIVING SPEECH IS PERCEIVING GESTURES

A controversial claim of the motor theory of speech perception is that the objects of speech perception are the speakers' vocal tract gestures and not the acoustic patterns that the gestures generate in the air. Given that under normal circumstances, the acoustic patterns and the gestures that caused them necessarily co-occur, what would count as evidence that perceptual objects are gestural and not acoustic? Here, we will consider four diverse kinds of evidence (see Goldstein & Fowler, 2003, for a more extensive review).

The first was proposed by Liberman (1957). He asked, "when articulation and sound wave go their separate ways, which way does perception go?" (p. 121). His answer was that perception goes with articulation. Liberman knew that the question as posed was not wholly accurate. As we remarked above, articulation is the cause of the sound wave, and hence, they cannot "go their separate ways." However, due to coarticulation, sometimes they seem to, providing an opportunity to assess the nature of perceptual speech objects. The relevant evidence was provided by the findings reviewed earlier (Liberman et al., 1952; Liberman et al., 1954) that very different second-formant transitions can signal /d/ in the synthetic syllables /di/ and /du/ and that identical stop bursts can signal /p/ and /k/ before different vowels.

A second kind of evidence is provided by the fact that some gestures may be specified by information other than that in air pressure waves. In particular, information present in reflected light or skin pressure patterns may specify some phonetic gestures (e.g., labial gestures). When it does, we can ask whether speech perception is responsive to these additional sources of information. The answer to the question is yes. For example, seeing a speaker produce one syllable while listening to a different syllable can affect how the heard syllable is perceived (the *McGurk effect*; e.g., Massaro, 1987; McGurk & MacDonald, 1976). Another example is the finding that listeners perceive speech in noise more accurately when they can see the speaker than when they cannot (Sumby & Pollack, 1954). These effects have been replicated when the syllable is not seen but is perceived haptically (Fowler & Dekle, 1991; Gick, personal communication, August 25, 2004).

A third kind of evidence derives from the prediction that if listeners perceive gestures, speech imitation should be very fast, because speech percepts may serve as instructions for imitation. Imitative responses to speech are very fast (Fowler, Brown, Sabadini, & Weihing, 2003; Kozhevnikov & Chistovich, 1965; Porter & Castellanos, 1980; Porter & Lubker, 1980). In order to explain why these findings imply perception of gestures, we need to undertake a brief detour.

Two of the tasks commonly used in the reaction time literature are the *simple* and the *choice* tasks. In *simple* tasks, participants produce the same detection response to different stimuli. For example, they might hit the same response button with their right hand when they see either a square or a rectangle. In *choice* tasks, participants make different responses to different stimuli. For example, they might hit a button with their right hand when they see a square and hit a button with their left hand when they see a rectangle. According to Luce (1986), choice reaction times typically exceed simple times by about 100–150 msec, because there is an element of choice in the choice task that is absent in the simple task. That is, to make a response in the simple task, participants merely have to detect the stimulus, whereas in the choice task, they have to identify the geometric shape and choose which button to press on that basis.

However, the element of choice in the choice task can be reduced when the stimulus provides nonarbitrary information useful for the response. In our example, if the square (which calls for a right-hand response) is presented on the right and the rectangle (which calls for a left-hand response) is presented on the left, choice response times are shorter than response times to stimuli presented in the center (Umiltà, Rubichi, & Nicoletti, 1999; see also Hietanen & Rama, 1995).

In choice tasks that use spoken stimuli and spoken responses, the element of choice may be similarly reduced if the stimuli provide nonarbitrary information useful for the responses. Several studies have shown that this is the case (Fowler et al., 2003; Kozhevnikov & Chistovich, 1965; Porter & Castellanos, 1980; Porter & Lubker, 1980). For example, Fowler et al. (2003) presented the speech of a model speaker producing an extended /a/ vowel, followed at an unpredictable time by one of the three CV syllables /pa/, /ta/, or /ka/. In the simple task, the participants shadowed the /a/ vowel, and as soon as they detected the model shifting to one of the CV syllables, they always produced the same designated syllable: /pa/ for one third of the participants and /ta/ and /ka/ for the remaining thirds. In the choice task, the stimuli were the same, but now the participants produced an imitative response. That is, they shadowed /a/, and as soon as they detected the model shifting to /pa/, /ta/, or /ka/, they produced /pa/, /ta/, or /ka/, respectively. Fowler et al. (2003) found a 26-msec difference between the two tasks, a result that replicates earlier findings by Porter and Castellanos and by Porter and Lubker (see also Kozhevnikov & Chistovich, 1965). These differences are far less than the 100- to 150-msec differences between simple and choice tasks reported by Luce (1986). Moreover, the simple response times are in the range of those of the simple task studies summarized by Luce (1986). In other words, the results by Fowler et al. (2003), as well as those by Porter and colleagues, suggest that the element of choice in the choice task had been reduced. This reduction is understandable if perceiving speech is perceiving gestures.[12] In that case, what is perceived provides instructions for the required response—indeed, reducing the element of choice. However, if perceiving speech is perceiving acoustic objects per se, the element of choice remains substantial; the stimuli are acoustic, but the responses are actions.

A second finding of the study by Fowler et al. (2003), reinforcing the interpretation that perceiving speech is perceiving gestures, was obtained in the simple task. On one third of the trials in this task, the CV syllables produced by the model matched the participant's designated syllable. For example, the model's CV might have been /pa/, and the designated response was also /pa/. On two thirds of the trials, the model's CV mismatched the participant's designated response. Response times on matching trials were significantly shorter than those on mismatching trials. Again, if perceiving speech does not imply perceiving gestures, this difference is difficult to explain. However, if listeners do perceive gestures, the model's matching syllable may have served as a goad for an imitative response.

A final kind of evidence tests the prediction that if listeners perceive gestures, their parsing of the speech signal should be sensitive to the acoustic effects of coarticulated gestures. That is, the signal should be processed so that acoustic information for a given gesture is used as information for that gesture even when its acoustic consequences overlap with the acoustic consequences of another gesture. Figure 2 illustrates this idea metaphorically. Figure 2A shows the acoustic consequences of two gestures that overlap in time, captured in a space in which the x-axis is time and the y-axis reflects the relative prominence of the gestures' acoustic consequences. As Figure 2A illustrates, during the time in which the two gestures overlap (the gray region in the figure), the acoustic consequences for the two gestures are mixed. In Figure 2B, the mixing of the acoustic consequences of the two gestures is eliminated by situating the acoustic signal in a space that captures the signal's gestural causes.

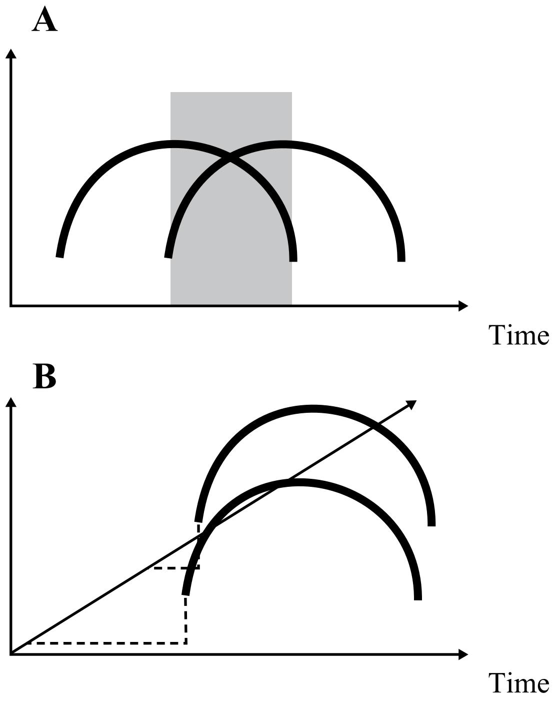

Figure 2. Metaphorical illustration of the concept of gestural parsing. Panel A shows the acoustic consequences of two gestures that overlap in time, captured in a space in which the *x*-axis is time and the *y*-axis reflects the relative prominence of the gestures' acoustic consequences. As panel A illustrates, during the time in which the two gestures overlap (the gray region in the figure), the acoustic consequences for the two gestures are inextricably mixed. In panel B, the mixing of the acoustic consequences of the two gestures is eliminated by situating the acoustic signal in a space that captures the signal's gestural causes.

There is considerable evidence that listeners situate the acoustic signal in a space that captures its gestural causes (summarized by Fowler, 2006). On the one hand, there is evidence that listeners perceptually separate the acoustic information for, say, Gesture A in the acoustic domain of Gesture B from the acoustic information for Gesture B (*perceptual separation*). On the other hand, there is evidence that listeners ascribe the acoustic consequences of

Gesture A in the acoustic domain of Gesture B to Gesture A (*acoustic ascription*).

Mann and Repp (1980) have provided an example of perceptual separation. They found that listeners report more *s*s along an /s/ to /ʃ/ (*sh*) continuum when the consonant precedes /u/, a lip-rounded vowel, than when it precedes /a/. Coarticulatory lip-rounding in the consonant has the acoustic consequence of lowering its frequency spectrum, lowering the high frequencies of /s/ toward the lower frequencies of /ʃ/. The finding of more *s* responses preceding /u/ suggests that listeners are separating the spectrum-lowering effects of anticipatory lip rounding from the acoustic consequences of the consonant (Mann & Repp, 1980).

Whalen (1984) has provided an example of acoustic ascription. He presented listeners with /s/ or /ʃ/ consonants followed by /u/ or /a/ vowels and asked them to identify the vowels. Half of the syllables were cross-spliced so that, for example, an /u/ vowel was spliced onto an /s/ or /ʃ/ consonant that had been produced in the context of an /a/. The other half of the syllables were spliced so that, for example, an /u/ vowel was spliced onto an /s/ or /ʃ/ consonant that had been produced in the context of a different /u/ vowel. Response times to identify the vowels were longer for cross-spliced than for spliced syllables. In other words, listeners used coarticulatory information for the vowel in the domain of the consonant as information for the vowel. They were misled when the information was misleading—that is, in cross-spliced syllables. In short, listeners ascribed to the vowels the acoustic consequences of the vowels present in the consonants, the same acoustic consequences that Mann and Repp (1980) showed were separated from the /s/ and /ʃ/ consonants.

The different classes of research findings reviewed in this section can be given alternative explanations in theories that negate the claim that the vocal tract gestures are the objects of speech perception. For example, Massaro (1998) has proposed that the McGurk effect is the result of prototypes in memory that specify the optical and acoustic cues for syllables, and Lotto and Kluender (1998) have invoked spectral contrast to account for some cases of apparent perceptual separation (*compensation for coarticulation*). As for the findings of Liberman et al. (1952) and Liberman et al. (1954), we are not aware of an alternative explanation. However, one might propose some as yet unspecified perceptual context effects that could explain the data. Similarly, although we are not aware of alternative accounts for the imitative responses to speech, an account that does not assume that perceiving speech is perceiving gestures might be devised. However, it is very unlikely that such an account would invoke prototypes in memory, spectral contrast, or the as yet unspecified context effects that might explain the findings of Liberman et al. (1952) and Liberman et al. (1954).

Indeed, all of these explanations, actual and imagined, are likely to differ from one another. In contrast, the hypothesis that gestures are the objects of speech perception provides a unified account of all of the findings: Perceiving speech is perceiving phonetic gestures.

THE MOTOR SYSTEM IS RECRUITED FOR PERCEIVING SPEECH

There is little direct evidence for motor system or motor competence involvement in speech perception—not because evidence has been sought but not found but, rather, because it has not frequently been sought. In the following, we first will review the available evidence within the domain of speech. Next, we will assess the likelihood of motor involvement in speech perception in an indirect way, by situating the motor theory of speech perception in the larger context provided by findings that (1) perceptual–motor links mark other natural communication systems, (2) human perception of motion is informed by motor competence, and (3) the motor system itself is recruited for perceptual tasks.

Motor Involvement in Speech Perception

One finding of motor involvement in speech perception has been provided by Cooper (1979). Many earlier studies had reported a *selective adaptation* effect (see the pioneering study by Eimas & Corbit, 1973). Repeated presentation of a syllable such as, for example, /pa/ leads to fewer identifications of ambiguous syllables as /pa/ along, say, a /ba/-to-/pa/ continuum. Cooper showed not only that perception of speech is affected by selective adaptation, but also that production of speech is affected as well. He found small but reliable reductions in the voice onset times of /pi/ and /ti/ syllables produced by speakers after adapting to acoustically presented /pi/, a finding that suggests a perception–production link in speech.

Bell-Berti, Raphael, Pisoni, and Sawusch (1979) have provided additional evidence for a production–perception link involving the English vowels /i/, /ɪ/ (the vowel in *bit*), /e/ (the vowel in *bait*), and /ɛ/ (the vowel in *bet*). The phonetic differences among the vowels can be described in two ways. The vowels decrease in "height" in the series as listed above. Alternatively, /i/ and /e/ are tense vowels; /ɪ/ and /ɛ/ are lax. Within the tense and lax pairs, the vowels differ in height. Bell-Berti et al. found individual differences in the production of the vowels. Consistent with the phonetic distinction in height, 4 speakers showed a gradual decrease in activity of the genioglossus muscle (a muscle of the tongue that affects tongue height) in the series of four vowels as listed. Consistent with the phonetic distinction between lax and tense vowels, 6 speakers showed comparable levels of activity for /i/ and /e/ that were higher than the activity levels for the two lax vowels.

In a perception test, the 10 participants partitioned into the same two groups. Listeners identified vowels along an /i/ to /ɪ/ continuum. In one test, the vowels along the continuum were equally likely to occur. In the other, called the *anchoring test*, the vowel at the /i/ end of the continuum occurred four times as frequently as other members of the continuum. The latter manipulation tends to decrease /i/ identifications for the ambiguous members of the continuum. The participants who showed a height distinction in their production of the four vowels showed much larger anchoring effects than did the 6 speakers who produced /i/ with more activity in the genioglossus muscle and, presumably, a higher tongue than for /ɪ/. The authors concluded that for the second group of listeners, /i/ and /ɪ/ were not adjacent vowels (differing in height), whereas they were for members of the first group.[13] Whether or not this is the appropriate account, it is telling that the participants grouped in the same way as talkers as they did as listeners. This provides some evidence suggesting that how a listener perceives speech is informed by how the listener, as a speaker, produces it.

Interactions between action and perception have also been demonstrated in tasks in which exposure to visible speech gestures is involved. Kerzel and Bekkering (2000) found a compatibility effect in speech production. The participants in their experiment saw a video of a face mouthing /ba/ (as in *box*) or /da/ (as in *doll*) on each trial. At a variable interval after the visual presentation of this task-irrelevant material, they saw either of two symbols (in one experiment, ## or &&) that they had learned to associate with the spoken responses /ba/ and /da/. The participants' task was to respond as quickly as possible to these symbols (by saying /ba/ or /da/), and they were told to ignore the video clips. Nevertheless, the results showed that there was an effect of the irrelevant visible speech gesture. In particular, the latency to produce the syllables cued by the symbols was affected. The /ba/ responses were faster when the face mouthed /ba/ than when it mouthed /da/. Likewise, /da/ responses were faster when the face mouthed /da/ than when it mouthed /ba/. Kerzel and Bekkering invoked the motor theory of speech perception and suggested that perceiving the mouthed gestures activates a corresponding motor code that interacts with the codes activated by the simultaneous planning of the same action, as elicited by the relevant cue.

In addition to this slim amount of behavioral evidence suggesting a production–perception link in speech, there is some neural evidence, which will be reviewed below in the Neural Evidence: Mirror Neurons in Primates section.

Perceptual–Motor Links as Marks of Parity-Achieving Systems

The requirements for parity that the speech module is meant to satisfy are specific instances of general requirements that constrain the development and preservation of interindividual communication systems.[14] That is, for any communicating conspecifics, (1) there must be convergence on what count as communicative messages, and (2) messages sent and received must be the same. Moreover, as for parity in speech, meeting these requirements is more easily guaranteed if the production and perception systems of communicating animals coevolve.

Studies conducted on the acoustic communication systems for mate recognition in crickets and frogs (Doherty & Gerhardt, 1983; Hoy, Hahn, & Paul, 1977) suggest that the mechanisms that support the achievement of communication parity in these species are similar in kind to those proposed by the motor theory of speech perception. In particular, these studies provide evidence supporting the idea that there exists a linkage between the systems underlying the production of sounds (in this case, in one animal) and those underlying their perception in its mate. Doherty and Gerhardt and Hoy et al. agree that this linkage is due to genetic coupling. Their main argument relies on the fact that when two species are bred to create hybrids with different calls, the mates receptive to the call—the hybrid females here—show a strong preference for the call of the males coming from the same breed of hybrids, relative to the calls of other hybrids or the original breeds. Tight genetic linkages must preserve communicational parity. Alexander (1962) further proposed the existence of a common neural mechanism that supports production of calls by the sender and perception of them by the receiver. In his words,

> In the evolution of any communicative system, whenever change of any sort occurs, there must be a change in two respects: the signal and the receiver. In the case of cricket stridulations, this means that the song of the male and the ability of the female to respond to it (correctly) must evolve together as a unit . . . . But the kind of differences that occur among the songs of closely related species usually do not in any way involve the structure of the stridulatory apparatus (at least externally). Likewise, the differences in the ability of the female to *respond* (properly) probably do not in any way involve the auditory apparatus itself. In both cases the difference seems to reside in the central nervous system. Indeed . . . song differences among closely related species always (and usually only) involve those unalterable components of the patterns that must derive from the central nervous system. Is it possible that in some or many cases the song difference—perhaps even the particular difference in the structure of the central nervous system itself—is exactly the same as the difference which causes the response difference? . . . If there is a linkage—or an identity—here, it would represent an interesting simplification of the process of evolutionary change in a communicative system—something of an assurance that the male and the female or the signaler and the responder—really will evolve together and possibly an increased likelihood through this that the entire system will persist. (p. 465)

Although the idea of direct genetic coupling has been challenged in recent times by proponents of a coevolutionary process that relies on genetic correlations, rather than on genetic coupling (Boake, 1991; Butlin & Ritchie, 1989; Jarvis & Nottebohm, 1997), there is substantial agreement about the fact that the production and perception systems of communicating animals must coevolve in order to preserve the communicating species (e.g., Blows, 1999). Moreover, in a way that is remarkably similar to linguistic differentiation among humans, whenever the parity constraints between the motor system of the sender and the perceptual system of the receiver are violated, there is the potential for speciation (Ryan & Wilczynski, 1988; Shaw, 2000). In Ryan and Wilczynski's words,

> Mate recognition requires congruence between the structure of the signal and the response properties of the sensory system that decodes the signal. This occurs in visual, olfac-

> tory, and electrosensory modalities and has been especially well documented in acoustic mate recognition systems. This congruence is necessary for efficient communication, and during evolution must be maintained by correlated changes in the signal and the receiver. By promoting assortative mating, these correlated differences in signal and receiver can restrict genetic exchange and promote genetic divergence among populations. Thus divergence in courtship signals can be an important component of the speciation process. (p. 1786)

Finally, there is ample evidence in favor of the idea that linkages between the perceptual and motor systems exist within the nervous system of the same animal. For example, for songbirds, such as zebra finches (Williams & Nottebohm, 1985), canaries (Burt, Lent, Beecher, & Brenowitz, 2000; Nottebohm, Stokes, & Leonard, 1976), and white-crowned sparrows (Whaling, Solis, Doupe, Soha, & Marler, 1997), as well as for other birds, such as parrots (Plummer & Striedter, 2000), the neural motor centers that control song production have been shown to be sensitive to acoustic stimulation, particularly to the songs that are specific to the species of the bird (but see Dave, Yu, & Margoliash, 1998, for a cautionary note about the interpretation of these results).

Over different taxa and through different pieces of neural circuitry, one sees the same design principle: The system that produces a signal of communicative value is connected to the system that perceives the signal. Similar evidence of linkages between perception and the motor system is available for monkeys and humans and will be reviewed below.

Motor Competence in Perception

We now will move outside the realm of communication systems to review the evidence for the involvement of motor competence in perception. The review will cover the following aspects of motor competence: (1) motor competence with regard to anatomical constraints on body movements, (2) motor competence with regard to the general dynamical constraints on biological motion, and (3) motor competence with regard to the subtle individual differences in performing body movements.

Motor competence about general anatomical constraints. Research by Shiffrar and Freyd (1990, 1993) suggests that knowledge of anatomic constraints affects what people perceive. Their research made use of the phenomenon of apparent motion. Apparent motion may be seen if, for example, light flashes are presented in alternation on the left and right sides of a display. If the timing relations are appropriate, viewers report seeing a light that moves smoothly back and forth across the screen. The motion path is typically the most direct, shortest path between the two locations at which light flashes or other stimuli were presented. Shiffrar and Freyd's results were different.

Shiffrar and Freyd (1990) presented photographs of human figures in alternation. Example pictures are shown schematically in Figures 3A and 3B. To get from the arm posture in Figure 3A to the arm posture in Figure 3B, only one body movement is possible, that illustrated in Figure 3C. However, a shorter path would be the movement shown in Figure 3D, which is impossible because of anatomical constraints on the joints of the arm. Shiffrar and Freyd (1990) found that at short stimulus onset asynchronies (SOAs) between picture presentations, viewers did see the shorter path; however, at longer SOAs, they saw the longer, anatomically possible path shown in Figure 3C. At both long and short SOAs in which pictures of inanimate objects (e.g., a clock) were presented, the motion path was always the shortest, most direct path.

In later research, Shiffrar and Freyd (1993) eliminated the possibility that viewers see longer motion paths with longer SOAs by including stimuli in which the shorter path was the anatomically possible one. In that case, the short path was seen at all SOAs.

Motor competence about general dynamical constraints. Further evidence of exploitation of motor competence in perception has been provided by the work of Viviani and colleagues (e.g., Kandel, Orliaguet, & Viviani, 2000; Viviani, Baud-Bovy, & Redolfi, 1997; Viviani & Mounoud, 1990; Viviani & Stucchi, 1989, 1992a, 1992b). In evidence that we will review below, Viviani and colleagues found that perception of the properties of an event—for example, a motion trajectory—is sensitive to the law that specifies the motion when the trajectory is the product of a biological motor system. On the basis of findings such as these, Viviani and Stucchi (1992b) argued that "If a visual pattern can only correspond to a specific motor behavior, and if humans are genetically equipped to produce that behavior, one can speculate that specialized motor competencies are called upon in the perception of the pattern" (p. 233). As was mentioned above, the kind of specialized motor competence Viviani and Stucchi referred to is a law that governs biological movements in a two-dimensional space. The law, called the *two-thirds power law*, expresses a constraint on the relation between the kinematics and the geometry of biological movements, so that velocities are slower the more curved the motion paths.[15]

Viviani and colleagues (e.g., Viviani & Stucchi, 1989) have investigated the consequences for perception of the two-thirds power law by presenting traces over time of a two-dimensional movement in which the geometry of the trajectory (the radius of curvature), the kinematics of the motion (the velocity profile), or both were manipulated. The results were clear: Compliance with the law affected the perception of the trace. For example, if the trace followed a circular trajectory but had a velocity profile that under the two-thirds power law would characterize an elliptical biological motion, the motion was perceived as elliptical (Viviani & Stucchi, 1989).

Moreover, uniform kinematics of a visible planar motion were perceived as nonuniform if they violated the two-thirds power law, and nonuniform kinematics of a planar motion were perceived as uniform if they abided by the law (Viviani & Stucchi, 1992a). The latter phenomenon is consistent and robust. Velocity profiles with differences

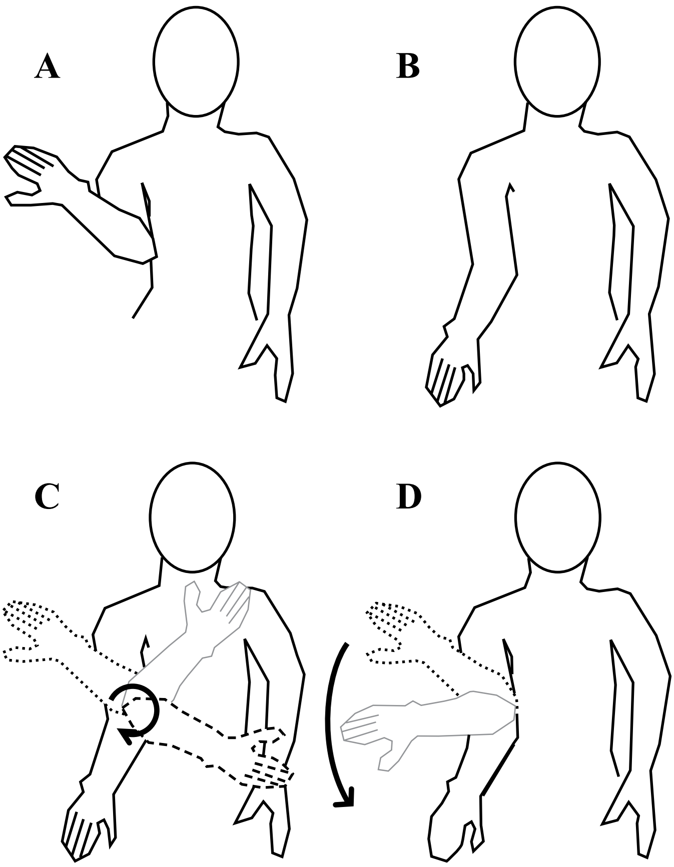

Figure 3. Illustration of an anatomical constraint on the rotation of the arm. Panels A and B illustrate the end points of an arm movement. Given the anatomical constraints on the joints of the arm, the movement can occur along the trajectory illustrated in panel C, but not along the trajectory illustrated in panel D.

between the slowest and the fastest velocities on the order of 200% were judged uniform, and the misjudgments remained when the participants were shown examples of true uniform velocities.

Compliance of visible planar trajectories with the two-thirds power law also had consequences for simple pursuit tracking, whether or not the trajectories were predictable. Motions consistent with the law were pursued more accurately than motions inconsistent with the law (Viviani, Campadelli, & Mounoud, 1987; Viviani & Mounoud, 1990).

The law's influence extends to kinesthetic perception (Viviani et al., 1997). For passive movement of an arm along an elliptical trajectory, adherence to the law affected both verbal judgments of trajectory direction and active reproduction of the passive trajectory by the other arm.

Motor competence about specific individual signatures. Viviani and colleagues identified the motor competence manifest in their perceptual tasks as a law expressing a general dynamical constraint on biologically produced movements. However, motor competence may also include constraints that are not general but, rather, are idiosyncratic to the movements of an individual. If, as Viviani and colleagues suggested, motor competence is called upon in perception of movement, perceptual performance may be enhanced in the case of movements produced by the same individual who perceives them, because the maximal amount of motor competence is available to support perception.

Several studies (recently reviewed by Knoblich & Flach, 2003) suggest that such enhancement indeed occurs. For example, Repp (1987) found that participants were more accurate in recognizing recordings of their own hand-clapping than they were in recognizing the clapping of other, familiar persons. Similarly, Knoblich and Prinz (2001) found that individuals looking at a dynamic display of a symbol being traced over a two-dimensional surface distinguished between symbols that they had produced themselves in a previous session (without seeing the outcome of their productions) and the same symbols produced by others. The ability was not affected significantly by familiarity with the symbols being traced (e.g., Roman vs. Arabic script for European participants) and, consistently with the findings by Viviani and colleagues, depended on the presence of information about the velocity profile of the motions.

There are also indications that not only are humans particularly skilled at distinguishing their own productions from those of other people, but also, when they perceive themselves, they are more accurate in predicting movement outcomes.[16] Knoblich and Flach (2001) found that when watching video clips representing an arm throwing a dart toward a target, participants were better at predicting the outcome of a throw when the video clips were of their own arm, rather than of that of someone else. Similar results have been obtained for participants' predictions about whether or not a new stroke will follow a dynamically presented handwritten symbol (Knoblich, Seigerschmidt, Flach, & Prinz, 2002).

In sum, there is reason to believe that perception is particularly attuned to the general anatomical and dynamical constraints on biological movements, as well as to the specific subtleties of individual movements. In other words, the same conclusion that Liberman and colleagues (e.g., Liberman et al., 1967) drew specifically for speech, that speech motor competence must inform speech perception, can be drawn, on very different bases, for motor competence and perception quite generally.

Motor System Involvement in Perception

We now will turn to evidence that not just motor competence, but the motor system itself, is recruited in perception. We first will review neural evidence for motor system recruitment in perception and then will review neural and behavioral evidence that a divide between perception and action is difficult, if not impossible, to identify.

Neural evidence: Mirror neurons in primates. The discovery of *mirror neurons* (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992) provides direct neural evidence for motor system involvement in perception. Since the discovery, research on mirror neurons has grown quickly, and reviews have recently been provided elsewhere (Rizzolatti & Craighero, 2004; Rizzolatti, Fogassi, & Gallese, 2001). Here, we will provide a brief chronological overview of the findings that are most relevant for the purposes of this article.

In the late eighties and early nineties, Rizzolatti and colleagues (e.g., Rizzolatti et al., 1988) were investigating the activity of individual neurons in the premotor cortex (area F5) of monkeys engaged in hand manipulations of objects. Area F5 is involved primarily in the control of hand movements and includes neurons that code goal-directed actions (e.g., grasping, tearing, holding, etc.) in a highly specific way, responding selectively to subtle details of the movements to be performed. While recording the activity of the motor neurons in area F5 of the monkey's brain, they discovered that some neurons that fired when the monkey performed a specific grasping movement of the hand also fired when a human experimenter was performing a similar grasping movement in front of the monkey (Di Pellegrino et al., 1992). Not only were these neural activities in the premotor cortex correlated with abstract visual properties of specific movements, but also they were correlated with these movements in a way that was independent of who was performing the action.

Soon after the discovery of the mirror neuron system in monkeys, evidence for a corresponding system in humans was obtained for finger, hand, and arm movements in experiments in which transcranial magnetic stimulation was used (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Strafella & Paus, 2000). Similar results have been obtained using PET data (Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Rizzolatti et al., 1996) and fMRI data (Iacoboni et al., 1999).

In Rizzolatti and Arbib's (1998) words, "taken together, the human and monkey data indicate that, in primates, there is a fundamental mechanism for action recognition. . . . Individuals recognize actions made by others because the neural pattern elicited in their premotor areas during action observation is similar to that internally generated to produce that action" (p. 190).

Recently, researchers have shown that Rizzolatti and Arbib's (1998) "fundamental mechanism for action recognition" has ties with general audition, as well as with speech perception.

As for general audition, Kohler and colleagues (Kohler et al., 2002) found neurons in the premotor cortex of monkeys that respond not only when the monkey performs a specific action (e.g., breaking a peanut) or sees the action performed by someone else, but also when the monkey merely hears the sound that is caused by the specific action (i.e., the cracking noise of the peanut's shell).

As for speech perception, there is now evidence that perceiving speech involves neural activity of the motor system. Two recent studies involving the use of transcranial magnetic stimulation of the motor cortex have demonstrated activation of speech-related muscles during the perception of speech. Fadiga and his colleagues (Fadiga, Craighero, Buccino, & Rizzolatti, 2002) found that when listeners hear utterances that include lingual consonants, they show enhanced muscle activity in the tongue. Watkins and colleagues (Watkins, Strafella, & Paus, 2003) found that both while listening to speech and while seeing speech-related lip movements, people show enhanced muscle activity in the lips. Complementarily, two fMRI studies (Pulvermüller et al., 2006; S. M. Wilson, Saygin, Sereno, & Iacoboni, 2004) demonstrated that there is overlap between the cortical areas active during speech production and those active during passive listening to speech.

The meshing of perception and action. Another kind of neuron found in the ventral premotor cortex of monkeys is called the *canonical neuron*, a type of neuron that responds both when the monkey grasps an object and when it sees the same graspable object (Murata et al., 1997). Visibly different objects that can be handled in a similar fashion evoke similar responses. In other words, canonical neurons are responsive to the actions that an object potentially affords, even when acting on the object is not required. Chao and Martin (2000) found compatible evidence in humans. Using fMRI, they found that the left ventral premotor area and the left posterior parietal region showed more activity when humans viewed tools than when they viewed animals, faces, or houses. They suggest that these regions link the visible properties of objects, such as tools, with the hand and finger movements involved in making use of them.

Moreover, visuo-tactile neurons in area F4 have been shown to have receptive fields that code visual space in a way that is more easily explained by assuming motor body-dependent coordinates, rather than visual body-independent space coordinates (Rizzolatti, Fadiga, Fogassi, & Gallese, 1997).

In brief, there is reason to believe that the original domain of the mirror neuron system may be extended beyond action execution and action recognition to the domain of the general perception of objects and space in motor terms. In other words, the perceptual relationship between an animal and its surrounding physical world is reflected in the nervous system in a way that is intimately intertwined with the neural means for preparing to produce compatible actions.

The action-effect and common-coding principles. Another recent body of evidence that is compatible with the idea of a meshing between perception and action in the nervous system is that collected by Prinz and his colleagues. Prinz (1997) proposed that "'event codes' and 'action codes' should be considered the functional basis of percepts and action plans, respectively. It is held that they share the same representational domain and are therefore commensurate" (p. 133). This idea, called the *common-coding principle*, contrasts with the more traditional view that perception and action are coded in incommensurate formats and, consequently, each has to be translated into the format of the other (e.g., Massaro, 1990; Posner, 1978; Sternberg, 1969). To solve the problem of incommensurability, Prinz invoked the *action-effect principle*: Planned actions are represented in terms of the effects they produce in the world (cf. Bernstein, 1967; Pribram, 1971). Therefore, their representation becomes indistinguishable from those of any other perceived events, because both are about the distal world.

Prinz's (1997) proposal leads to the prediction that when simultaneous activation of the perceptual and motor codes occurs, the two codes may interact. Prinz and his colleagues have found substantial evidence for this prediction (see Hommel, Müsseler, Aschersleben, & Prinz, 2001, for a review).

One piece of evidence is the well-known stimulus–response compatibility effect. Responses to stimuli located in spatial positions compatible with the effectors that make the responses tend to be executed more quickly than responses to stimuli located in noncompatible spatial positions (Fitts & Deininger, 1954). For example, responses to a sound presented to the right ear are faster with the right hand than with the left hand (Simon, Hinrichs, & Craft, 1970). Prinz (1997) considers the compatibility effect a direct demonstration of the common coding between perception and action. Compatible locations induce faster actions of the effectors because they activate part of the very same codes (e.g., *right*) that planned actions require.

In its original formulation, the notion of compatibility was in terms of body-side correspondences, as above. However, Hommel (1993) extended the notion of compatibility beyond anatomical correspondences to correspondences between environmental events. He demonstrated that inversion of the compatibility effect can be induced not only by manipulating the effectors' spatial setting (e.g., by doing the task with the hands crossed; cf. Simon et al., 1970), but also when the focus of the task is shifted from the movements of the effectors to the events that are caused by these movements. In his experiment, each of the participants' hands operated a switch that turned on a light on the opposite side of space. If the experimental instructions stressed that the movements of the hands on the switch were the relevant responses for the task, the participants showed a typical compatibility effect: Each hand responded more quickly to go signals presented auditorily on the same side of space. The location of the lights had no impact on responses. However, when the instructions emphasized that the consequences of the hand movements—the lights switching on—were the relevant responses for the task, the compatibility effect reversed: Each hand responded more quickly to go signals presented on the opposite side of space. This finding shows that the symmetries underlying the compatibility effect can arise in task space as well as in body space and that the symmetries can differ one from the other. The results are fully compatible with the action-effect principle.

Following a similar line of reasoning, Mechsner, Kerzel, Knoblich, and Prinz (2001) demonstrated that given appropriate visual feedback, participants can produce bimanual movements that ordinarily are very difficult. In one experiment, they used a bimanual task in which the participants moved each hand to control the circular movement of a flag. The participants could not see their hands, and their task was to make the flags move either in phase or in antiphase. When the cranks that regulated the movements of the flags were set so that the participants had to produce complex polyrhythms to synchronize the movements of the flags (e.g., three full cycles of one hand for every four full cycles of the other), the participants were successful at the task after a brief period of practice. This is surprising because complex polyrhythms are extremely difficult to perform for untrained participants (e.g., Treffner & Turvey, 1993). The authors explained the findings in terms of the action-effect and common-coding principles. They suggested that the difficulty of performing movement patterns does not depend on their intrinsic motor difficulty but on the ease with which their perceivable consequences can be controlled.

Other evidence in favor of the common-coding principle has come from Stürmer, Aschersleben, and Prinz (2000), who had participants produce a grasping gesture (first close the hand from a neutral posture, then open it) or a spreading gesture (first open, then close). These responses were cued by abrupt color changes of a hand that was visible on a computer monitor, with different colors signaling which gestures to perform. Crucially, this change in color occurred—with different onset times—while the hand on the monitor was producing a task-irrelevant gesture, starting and ending from a neutral half-open position. The irrelevant gesture reproduced one of the two gestures that the participants were cued to perform, and the participants were told to ignore it, selecting their responses only on the basis of the color change. The participants' response latencies were shorter when their movements matched the irrelevant gestures, demonstrating that an interaction took place between their perceiving and their acting.

A complementary result was obtained by Niedenthal, Brauer, Halberstadt, and Innes-Ker (2001), who demonstrated that preventing a motor response consistent with what is being perceived causes interference in perception. In particular, they presented dynamic images of faces that morphed from a happy to a sad expression or vice versa. The task of the participants was to stop the movie where they first detected an expression that was different from the one displayed initially; alternatively, they stopped the movie when they first detected the offset of the original expression. One group of participants performed the task with a pen held between lips and teeth, to prevent facial mimicry. The finding was that the participants in whom mimicry was prevented detected both the onset of a new expression and the offset of the original expression at a later point than did those in whom mimicry was permitted to occur.

The results by Niedenthal et al. (2001) can be related to the finding that when humans are preparing to execute an action on an object, they show facilitation in the processing of visual stimuli that are congruent with the intrinsic *motor properties* of the object they are about to handle (Craighero, Fadiga, Rizzolatti, & Umiltà, 1999). That is, preparedness to act seems to imply a state of perceptual selectivity toward properties of the environment that are congruent with the actions to be executed. For example, when participants are prepared to grasp a bar with a given spatial orientation upon the appearance of a go signal, the processing of the go signal is faster when its orientation matches that of the bar to be grasped (Craighero et al., 1999). Crucially, this facilitatory effect on perception is present not only in the latency to grasp the bar, but also in alternative responses, such as the latency of eye blinking. This indicates that the effect does not reflect stimulus–response compatibility between the go signal and the action to be produced.

In conclusion, there is substantial behavioral evidence that the architecture of cognition is permeated by linkages between the perceptual and the motor systems (Prinz & Hommel, 2002). Moreover, the theoretical perspective proposed by Prinz and colleagues introduces a new possible explanation for a motor theoretical account of perception. Whereas for most of the theorists reviewed earlier, perception is grounded in motor competence and/or processes that are enclosed within the anatomical boundaries of the perceiver, in Prinz and colleagues' proposal, perception shares common coding with action, because both are grounded outside of the physical boundaries of the perceiver, in the distal world.

CONCLUSION

We have reviewed the three main claims of the motor theory of speech perception to determine whether they deserve further consideration. Our review suggests that two claims—namely, that perceiving speech is perceiving gestures and that perceiving speech involves the motor system—warrant extended scientific scrutiny but the claim that speech is special, to the extent that it can be evaluated, does not.

Here, we will conclude with a general remark about the intellectual enterprise undertaken by Liberman and colleagues.

The main lesson learned by Liberman and colleagues in 50 years of empirical research is, in the end, rather simple: Cognition, like all products of evolution, cannot be understood in isolation (e.g., Clark, 1997). Rather, understanding cognition requires comprehending that it is *both* embedded in a meaningful ecological context *and* embodied in living perception–action systems (Bernstein, 1996; Dewey, 1896; Gibson, 1979; James, 1892; Pillsbury, 1911; Sperry, 1952; Thelen & Smith, 1994; Turvey & Shaw, 1995).[17] The concept of parity, developed by Liberman and colleagues (Liberman & Whalen, 2000; Mattingly & Liberman, 1988) in their later theorizing, captures well the extent to which Liberman and colleagues attempted to implement in their thinking the lesson they learned from their experiments. In fact, the concept of parity, the threefold nature of which we illustrated earlier in the article, can be seen as an attempt to integrate, through a set of simple constraints, the two contexts within which cognition must be understood. On the one hand, parity is intended by Liberman and colleagues as two abstract constraints on the speaker–listener linguistic interaction—that is, constraints arising from the meaningful ecological context within which spoken communicative acts are embedded (cf. Pickering & Garrod, 2004). On the other hand, parity is intended to be an abstract constraint on the symmetric coevolution of the machinery for producing and perceiving speech—that is, a constraint on the embodiment of spoken communication.

The concept of parity also captures another important aspect of the intellectual enterprise undertaken by Liberman and colleagues. As we documented earlier in the article, in order to understand the facts of speech perception ever better, Liberman and colleagues had to broaden the scope of their scientific perspective progressively, making it increasingly abstract. Such broad scope and abstractness may seem unjustified in a theory meant to address some specific facts about speech perception (e.g., Ohala, 1996), and this may well have been the reason for the skeptical reception of the theory within the field of speech. However, we suspect that it is exactly because of its broad scope and abstractness that the theory has had a positive reception outside of its own field. Indeed, today, the theory is more closely connected with research and theorizing in the broad context of cognitive science (e.g., Fadiga et al., 2002; Kerzel & Bekkering, 2000; Rizzolatti & Arbib, 1998; Viviani, 2002; S. M. Wilson et al., 2004) than it is with research and theorizing in the field of speech.

It is our hope that this recent connection, which the present article is meant to highlight, will not only enhance recognition of the valuable contributions of the motor theory of speech perception to the history of science, but also lead to new research that will develop the theory further.

REFERENCES

Alexander, R. D. (1962). Evolutionary change in cricket acoustical communication. *Evolution*, **16**, 443-467.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, **403**, 309-312.

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Sawusch, J. R. (1979). Some relationships between speech production and perception. *Phonetica*, **36**, 373-383.

Bernstein, N. (1967). *The coordination and regulation of movements*. New York: Pergamon.

Bernstein, N. (1996). On dexterity and its development (M. L. Latash, Trans.). In M. L. Latash & M. T. Turvey (Eds.), *Dexterity and its development* (pp. 1-244). Mahwah, NJ: Erlbaum.

Blows, M. W. (1999). Evolution of the genetic covariance between male and female components of mate recognition: An experimental test. *Proceedings of the Royal Society of London: Series B*, **266**, 2169-2174.

Boake, C. R. B. (1991). Coevolution of senders and receivers of sexual signals: Genetic coupling and genetic correlations. *Trends in Ecology & Evolution*, **6**, 225-227.

Boersma, P. (1998). *Functional phonology*. The Hague: Holland Academic Graphics.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, **3**, 219-252.

Burt, J. M., Lent, K. L., Beecher, M. D., & Brenowitz, E. A. (2000). Lesions of the anterior forebrain song control pathway in female canaries affect song perception in an operant task. *Journal of Neurobiology*, **42**, 1-13.

Butlin, R. K., & Ritchie, M. G. (1989). Genetic coupling in mate recognition systems: What is the evidence? *Biological Journal of the Linnean Society*, **37**, 237-246.

Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science*, **9**, 211-214.

Carello, C., Wagman, J. B., & Turvey, M. T. (2005). Acoustic specification of object properties. In J. Anderson & B. Anderson (Eds.), *Moving image theory: Ecological considerations* (pp. 79-104). Carbondale: Southern Illinois University Press.

Chao, L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, **12**, 478-484.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal behavior*. *Language*, **35**, 26-58.

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

Cooper, W. (1979). *Speech perception and production: Studies in selective adaptation*. Norwood, NJ: Ablex.

Craighero, L., Fadiga, L., Rizzolatti, G., & Umiltà, C. (1999). Action for perception: A motor–visual attentional effect. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 1673-1692.

Dave, A. S., Yu, A. C., & Margoliash, D. (1998). Behavioral state modulation of auditory activity in a vocal motor system. *Science*, **282**, 2250-2254.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, **3**, 357-370.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, **55**, 149-179.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, **91**, 176-180.

Doherty, J. A., & Gerhardt, H. C. (1983). Hybrid tree frogs: Vocalizations of males and selective phonotaxis of females. *Science*, **220**, 1078-1080.

Easton, T. A. (1972). On the normal use of reflexes. *American Scientist*, **60**, 591-599.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, **4**, 99-109.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, **171**, 303-306.

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.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, **73**, 2608-2611.

Fitts, P. M., & Deininger, R. L. (1954). S–R compatibility: Correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology*, **48**, 483-492.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, **68**, 178-183.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory & Language*, **49**, 396-413.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Crossmodal contributions to speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 816-828.

Fowler, C. A., & Galantucci, B. (2005). The relation of speech perception and production. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 633-652). Malden, MA: Blackwell.

Fowler, C. A., & Rosenblum, L. D. (1990). Duplex perception: A comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 742-754.

Fowler, C. A., Rubin, P. E., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In G. Butterworth (Ed.), *Language production* (pp. 373-420). New York: Academic Press.

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, **20**, 269-337.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159-207). Berlin: Mouton de Gruyter.

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography: 2. Observation compared with imagination. *Experimental Brain Research*, **112**, 103-111.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, **51**, 347-356.

Halle, M., & Stevens, K. (1962). Speech recognition: A model and a program for research. *IEEE Transactions on Information Theory*, **8**, 155-159.

Hayes, B., Kirchner, R., & Steriade, D. (Eds.) (2004). *Phonetically based phonology*. New York: Cambridge University Press.

Hietanen, J. K., & Rama, P. (1995). Facilitation and interference occur at different stages of processing in the Simon paradigm. *European Journal of Cognitive Psychology*, **7**, 183-199.

Hommel, B. (1993). Inverting the Simon effect by intention: Determinants of direction and extent of effects of irrelevant spatial information. *Psychological Research*, **55**, 270-279.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral & Brain Sciences*, **24**, 849-937.

Hoy, R. R., Hahn, J., & Paul, R. C. (1977). Hybrid cricket auditory behavior: Evidence for genetic coupling in animal communication. *Science*, **195**, 82-84.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, **286**, 2526-2528.

James, W. (1892). *Psychology: The briefer course*. New York: Holt.

Jarvis, E. D., & Nottebohm, F. (1997). Motor-driven gene expression. *Proceedings of the National Academy of Sciences*, **94**, 4097-4102.

Kandel, S., Orliaguet, J. P., & Viviani, P. (2000). Perceptual anticipation in handwriting: The role of implicit motor competence. *Perception & Psychophysics*, **62**, 706-716.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, **9**, 718-727.

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

Kelso, J. A. S., Vatikiotis-Bateson, E., Tuller, B., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 812-832.

Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 634-647.

Knoblich, G., & Flach, R. (2001). Predicting the effects of actions: Interactions of perception and action. *Psychological Science*, **12**, 467-472.

Knoblich, G., & Flach, R. (2003). Action identity: Evidence from self-recognition, prediction, and coordination. *Consciousness & Cognition*, **12**, 620-632.

Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 456-465.

Knoblich, G., Seigerschmidt, E., Flach, R., & Prinz, W. (2002). Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception. *Quarterly Journal of Experimental Psychology*, **55A**, 1027-1046.

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

Kozhevnikov, V. A., & Chistovich, L. A. (1965). *Speech, articulation and perception*. Washington, DC: U.S. Department of Commerce.

Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Erlbaum.

Kunkler-Peck, A., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 279-294.

Lacquaniti, F., Terzuolo, C., & Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. *Acta Psychologica*, **54**, 115-130.

Lawrence, D. H. (1949). Acquired distinctiveness of cues: 1. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, **39**, 770-784.

Lawrence, D. H. (1950). Acquired distinctiveness of cues: 2. Selective association in a constant stimulus situation. *Journal of Experimental Psychology*, **40**, 175-188.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, **29**, 117-123.

Liberman, A. M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of speech code. *Psychological Review*, **74**, 431-461.

Liberman, A. M., Cooper, F. S., Studdert-Kennedy, M., Harris, K. S., & Shankweiler, D. P. (1968). On the efficiency of speech sounds. *Zeitschrift für Phonetik, Sprachwissenschaft & Kommunikationsforschung*, **21**, 21-32.

Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, **65**, 497-516.

Liberman, A. M., Delattre, P., Cooper, F. S., & Gerstman, L. (1954). The role of consonant–vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General & Applied*, **68**, 1-13.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, **243**, 489-494.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, **4**, 187-196.

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, **60**, 602-619.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review*, **77**, 182-196.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition*, **14**, 211-235.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]–[s] distinction. *Perception & Psychophysics*, **28**, 213-228.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1990). An information-processing analysis of perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action* (pp. 133-166). Berlin: Springer.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Mattingly, I. G., & Liberman, A. M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. G. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function: Neurological bases of hearing* (pp. 775-793). New York: Wiley.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.

Mechsner, F., Kerzel, D., Knoblich, G., & Prinz, W. (2001). Perceptual basis of bimanual coordination. *Nature*, **414**, 69-73.

Mohanan, K. P. (1993). Fields of attraction. In J. Goldsmith (Ed.), *The last phonological rule* (pp. 61-116). Chicago: University of Chicago Press.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of Neurophysiology*, **78**, 2226-2230.

Niedenthal, P., Brauer, M., Halberstadt, J., & Innes-Ker, A. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression, *Cognition & Emotion*, **15**, 853-864.

Nottebohm, F. (1970). Ontogeny of bird song. *Science*, **167**, 950-956.

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in canary, *Serinus canarius*. *Journal of Comparative Neurology*, **165**, 457-486.

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America*, **99**, 1718-1725.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral & Brain Sciences*, **27**, 169-226.

Pierrehumbert, J. (1990). Phonological and phonetic representations. *Journal of Phonetics*, **18**, 375-394.

Pillsbury, W. B. (1911). The place of movement in consciousness. *Psychological Review*, **18**, 83-99.

Plummer, T. K., & Striedter, G. F. (2000). Auditory responses in the vocal motor system of budgerigars. *Journal of Neurobiology*, **42**, 79-94.

Porter, R. J., Jr., & Castellanos, F. X. (1980). Speech-production measures of speech perception: Rapid shadowing of VCV syllables. *Journal of the Acoustical Society of America*, **67**, 1349-1356.

Porter, R. J.[, Jr.], & Lubker, J. F. (1980). Rapid reproduction of vowel–vowel sequences: Evidence for a fast and direct acoustic–

motoric linkage in speech. *Journal of Speech & Hearing Research*, **23**, 593-602.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.

Pribram, K. H. (1971). *Languages of the brain: Experimental paradoxes and principles in neuropsychology*. Englewood Cliffs, NJ: Prentice Hall.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, **9**, 129-154.

Prinz, W., & Hommel, B. (Eds.) (2002). *Common mechanisms in perception and action: Attention and performance XIX*. Oxford: Oxford University Press.

Pulvermüller, F., Huss, M., Kheri, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, **103**, 7865-7870.

Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America*, **81**, 1100-1109.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, **21**, 188-194.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal movements. *Experimental Brain Research*, **71**, 491-507.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, **27**, 169-192.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1997). The space around us. *Science*, **277**, 190-191.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., & Fazio, F. (1996). Localization of grasp representations in humans by PET: 1. Observation versus execution. *Experimental Brain Research*, **111**, 246-252.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, **2**, 661-670.

Ryan, M. J., & Wilczynski, W. (1988). Coevolution of sender and receiver: Effect on local mate preference in cricket frogs. *Science*, **240**, 1786-1788.

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

Shaw, K. L. (2000). Interspecific genetics of mate recognition: Inheritance of female acoustic preference in Hawaiian crickets. *Evolution*, **54**, 1303-1312.

Shiffrar, M., & Freyd, J. (1990). Apparent motion of the human body. *Psychological Science*, **1**, 257-264.

Shiffrar, M., & Freyd, J. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science*, **4**, 379-384.

Simon, J. R., Hinrichs, J. V., & Craft, J. L. (1970). Auditory S–R compatibility: Reaction time as a function of ear–hand correspondence and ear–response-location correspondence. *Journal of Experimental Psychology*, **81**, 97-102.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Sperry, R. W. (1952). Neurology and the mind–brain problem. *American Scientist*, **40**, 291-312.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, **30**, 276-315.

Stevens, K. N. (1960). Toward a model for speech recognition. *Journal of the Acoustical Society of America*, **32**, 47-55.

Strafella, A. P., & Paus, T. (2000). Modulation of cortical excitability during action observation: A transcranial magnetic stimulation study. *NeuroReport*, **11**, 2289-2292.

Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: A study of imitation. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1746-1759.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212-215.

Sussman, H. (1989). Neural coding of relational invariance in speech: Human language analogs to the barn owl. *Psychological Review*, **96**, 631-642.

Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

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

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 211-263). Hillsdale, NJ: Erlbaum.

Turvey, M. T., & Shaw, R. E. (1995). Toward an ecological physics and a physical psychology. In R. Solso & D. Massaro (Eds.), *The science of the mind: 2001 and beyond* (pp. 144-169). Oxford: Oxford University Press.

Umiltà, C., Rubichi, S., & Nicoletti, R. (1999). Facilitation and interference components in the Simon effect. *Archives Italiennes de Biologie*, **137**, 139-149.

Viviani, P. (2002). Motor competence in the perception of dynamic events: A tutorial. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action: Attention and performance XIX* (pp. 406-442). Oxford: Oxford University Press.

Viviani, P., Baud-Bovy, G., & Redolfi, M. (1997). Perceiving and tracking kinesthetic stimuli: Further evidence of motor–perceptual interactions. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 1232-1252.

Viviani, P., Campadelli, P., & Mounoud, P. (1987). Visuo-manual pursuit tracking of human two-dimensional movements. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 62-78.

Viviani, P., & Flash, T. (1995). Minimum-jerk, two-thirds power law, and isochrony: Converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 32-53.

Viviani, P., & Mounoud, P. (1990). Perceptuomotor compatibility in pursuit tracking of two-dimensional movements. *Journal of Motor Behavior*, **22**, 407-443.

Viviani, P., & Stucchi, N. (1989). The effect of movement velocity on form perception: Geometric illusions in dynamic displays. *Perception & Psychophysics*, **46**, 266-274.

Viviani, P., & Stucchi, N. (1992a). Biological movements look uniform: Evidence of motor–perceptual interactions. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 603-623.

Viviani, P., & Stucchi, N. (1992b). Motor–perceptual interactions. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior: 2. Advances in psychology* (Vol. 87, pp. 229-248). Amsterdam: North-Holland.

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.

Whalen, D. H. (1984). Subcategorical mismatches slow phonetic judgments. *Perception & Psychophysics*, **35**, 49-64.

Whalen, D. H., Benson, R. R., Richardson, M., Swainson, B., Clark, V. P., La, S., et al. (2006). Differentiation of speech and nonspeech processing within primary auditory cortex. *Journal of the Acoustical Society of America*, **119**, 575-581.

Whalen, D. H., & Liberman, A. M. (1987). Speech-perception takes precedence over nonspeech perception. *Science*, **237**, 169-171.

Whaling, C. S., Solis, M. M., Doupe, A. J., Soha, J. A., & Marler, P. (1997). Acoustic and neural bases for innate recognition of song. *Proceedings of the National Academy of Sciences*, **94**, 12694-12698.

Williams, H., & Nottebohm, F. (1985). Auditory responses in avian vocal motor neurons: A motor theory for song perception in birds. *Science*, **229**, 279-282.

Wilson, M. (2001). Perceiving imitatible stimuli: Consequences of isomorphism between input and output. *Psychological Bulletin*, **127**, 543-553.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, **131**, 460-473.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004).

Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, **7**, 701-702.

NOTES

1. According to the ISI Web of Knowledge (February 2005), the article by Liberman et al. (1967) has been cited 1,236 times, and that by Liberman and Mattingly (1985) has been cited 497 times.

2. By *vocal tract* (or *phonetic*) *gestures*, we mean coordinated actions of vocal tract articulators that achieve some linguistic goal. It is important to note that according to this definition, gestures are not movements of individual articulators. For example, a bilabial consonant such as /b/ corresponds to one gesture—the complete closure of the lips—achieved by the coordination of the movements of three articulators: the jaw and the two lips (Kelso, Vatikiotis-Bateson, Tuller, & Fowler, 1984).

3. By *motor competence*, we refer to the set of laws that manifest themselves in coordinated action (e.g., Haken, Kelso, & Bunz, 1985; Kawato, 1999; Kelso, 1995; Kugler & Turvey, 1987; Viviani & Flash, 1995). By *motor system*, we refer to the physical manifestations of those laws in the body. Although the distinction may not withstand future investigations, some of the empirical findings that we will review below suggest it.

4. This is a device that transforms painted spectrographic patterns back to sound.

5. Although both of these influences were recognized by Liberman (1996, chap. 1), he did not provide specific references. Hence, we decided to provide two illustrative references. However, we could not find appropriate references predating 1967, the time period Liberman was referring to in his chapter.

6. That is, the percept mirrors the structure in the acoustic signal. This contrasts with the phonetic percept, which is *heteromorphic* with respect to the acoustic signal but is *homomorphic* with respect to the articulation.

7. Within the gestural perspective on speech perception, there are alternative views. According to an account that we favor, the very distinction between intended gestures and actual vocal tract actions is problematic, because it reflects the traditional division between the discrete mental categories identified by formal phonology and the gradient physical entities that are the subject matter of phonetics. That is, traditionally, *cognitive* phonology is held distinct from *physical* phonetics (e.g., Pierrehumbert, 1990). Thus, the utterance that one intends and the utterance that one effects are logically separable. To the extent that this phonology–phonetics gap exists, theories of speech perception (including that of Liberman and Mattingly) are forced to emphasize that the objects of perception are intended (cognitive phonology), not actual (physical phonetics). By the same token, if the phonology–phonetics gap does not exist, if it were to be dissolved through conceptual advances, theories of speech perception would more simply equate the objects of speech perception with actual gestures. In other words, intended and actual gestures would be the same.

The aforementioned conceptual advances are to a significant extent in progress (Browman & Goldstein, 1986; Goldstein & Fowler, 2003). Articulatory phonology is at the forefront. In developing the low-dimensional cognitive phonology and the high-dimensional physical phonetics as complementary aspects of a single complex dynamical system, articulatory phonology promises to dissolve the phonological–phonetics gap (Gafos, 2002; Goldstein & Fowler, 2003; Saltzman & Munhall, 1989). More general pursuits of a phonetically informed phonology can also be noted. Their intent likewise is to minimize the division between phonology and phonetics (e.g., Boersma, 1998; Hayes, Kirchner, & Steriade, 2004; Mohanan, 1993). Presuming that these related enterprises are successful, future accounts of speech perception may have no difficulty with the claim that hearers perceive speakers' actual gestures.

8. Other theorists acknowledge the existence of a linkage between speech perception and production. However, whereas for the motor theory the linkage reflects biological coevolution of the production and perception systems, for these theorists it reflects a bias for language communities to select articulations that have auditorily distinctive acoustic consequences (e.g., Diehl, Lotto, & Holt, 2004). These accounts are not mutually exclusive.

9. That is, the acoustic structure underlying our perception of sound in space includes time of arrival differences, intensity differences, and spectral differences at the two ears. However, we do not perceive these differences; rather, we perceive a distal property: the location of the sounding event in space.

10. For example, Carello et al. (1998) demonstrated that people can perceive the length of a dowel by the sound that it makes when dropped on the floor.

11. This idea would not have appealed to Liberman and colleagues because, for them, the phonetic module was especially adapted to the production and perception of coarticulated speech.

12. We do not mean to imply that this pattern of results for imitative responses is special to speech (M. Wilson, 2001). In particular, Fowler et al. (2003) predicted that their results would replicate in other imitative tasks. For example, they suggested tasks in which the stimuli are visual (e.g., video clips of finger movements) and the responses are manual (finger movements) or tasks in which the stimuli are nonspeech mouth sounds (e.g., lip smacking) and the responses are vocal. In these examples, as for speech, the reason for predicting that choice response times should approach simple response times is the same. Perceivers perceive distal events (finger movements, lip smacks, speech gestures) and not the proximal stimulation (reflected light and acoustic signals) that stimulates the sense organs.

13. These results may be at odds with a strict reading of the parity constraint that phonetic messages sent and received must be the same. Liberman and colleagues never specified what should count as *sameness*. Our opinion is that sameness of phonetic messages sent and received ought to be interpreted as *sufficient equivalence* (Fowler & Galantucci, 2005). If that were not the case, in fact, speakers of different dialects could not communicate with one another, and nonnative speakers could not be understood by native speakers.

14. Liberman and colleagues often noted the ubiquity of the parity requirement for communication systems (e.g., Liberman, 1996; Liberman & Mattingly, 1985). However, they never extended the scope of their theory beyond the domain of speech perception.

15. As presented, the law is formulated in its simplest form (Lacquaniti, Terzuolo, & Viviani, 1983). Over the years, the law has been revised and refined in order to account for new empirical observations (Viviani & Flash, 1995). However, the changes are not directly relevant to the argument made here. For our present purposes, any proposed law, as long as it correctly captures properties that are unique to biological movements, works well to predict perceptual performance and judgments. (Note that such a law does not necessarily generalize to nonbiological movements. For example, the motion of an object that is subject to a gravitational field—say, the earth rotating around the sun—follows a pattern that is opposite to that specified by the two-thirds power law, accelerating when the radius of curvature increases.)

16. M. Wilson and Knoblich (2005) have recently generalized this conclusion. They proposed that not only does perceiving oneself facilitate predicting the outcome of one's own actions but also, more generally, perceiving conspecifics' actions in motor terms facilitates predicting the outcome of their actions.

17. Although the core of Liberman and colleagues' theorizing is coherent with the tradition of thought identified here, their enterprise, it should be remarked, was not driven by it. It seemed to be largely unknown by Liberman and his colleagues or of little interest to them. The motivation for their motor theory of speech perception is to be found in the data provided by their experiments: Explanations derived from a nonembodied, nonembedded theory of cognition were not able to handle the facts.

(Manuscript received September 17, 2004;
revision accepted for publication June 15, 2005.)