The Handbook of Speech Perception
Edited by: David Pisoni & Robert Remez
Blackwell Publishing
pp. 633–652

26 The Relation of Speech Perception and Speech Production

CAROL A. FOWLER AND
BRUNO GALANTUCCI

26.1 Introduction

For the most part, speech perception and speech production have been investigated independently. Accordingly, the closeness of the fit between the activities of speaking and of perceiving speech has not been frequently addressed. However, the issue is important, because speakers speak intending to be understood by listeners.

We will focus on two central domains in which it is appropriate to explore the relation of speech production to speech perception: the public domain in which speakers talk, and listeners perceive what they say; and the private domain in which articulatory mechanisms support talking, and perceptual mechanisms support listening to speech.

In the public domain, we will suggest that the fit between the activities of talking and listening must be close and that, in fact, languages could not have arisen and could not serve their functions if the fit were not close. In respect to the private domain, we will focus specifically on a proposal identified with the motor theory of speech perception (e.g., Liberman, 1996; Liberman & Mattingly, 1985) that articulatory mechanisms are brought to bear on speech perception.

The public and private domains of speech are not unrelated. Speech is an evolutionary achievement of our species, and it is likely that the required tight coupling of the public activities of talking and of perceiving talk that shaped the evolution of language involve the evolution of the mechanisms that serve language use.

As for evidence for our claim that, in the public domain, the fit between talking and listening is close, we will sample research findings suggesting that primitive objects of speech perception are gestures. If this is the case, then gestures constitute a public currency of both perceiving and producing speech.

As for evidence relating to the motor theory's claim that, in the private domain, there is coupling of mechanisms that support talking and listening, research findings are limited; however, the evidence is buttressed by numerous findings of couplings between mechanisms supporting action and mechanisms supporting its perception. We review some of that evidence in a later section.

26.2 The Relation between Production and Perception in Public Language Use

26.2.1 *Theoretical context*

Language exhibits duality of patterning (Hockett, 1960). That is, it has syntactic structuring of words in sentences and phonological structuring of consonants and vowels in words. Although the meaningful utterances that the first level of structuring yields surely are the main foci of language users' attention, our focus here will be on the phonological level. This is because, at this level, languages provide the forms that speakers use to make their linguistic messages public.

Underlying the effectiveness of public language is a "bottom-line" requirement, namely that listeners must, in the main, accurately perceive the language forms that talkers produce. We will refer to this as achievement of "parity" (cf. Liberman & Whalen, 2000) here, a relation of sufficient equivalence between phonological messages sent and received. Because achievement of parity is essential to communicative efficacy, we expect properties of languages to be shaped by this requirement. We propose two such properties.

First, the forms should be the public actions of speakers, or they should be isomorphic with those actions. That is, if language forms are the very parts of our language system that permit its public use, and if, in public use of language, talkers intend to convey these forms to listeners by some kind of public action, successful communication would be fostered if the public actions were the forms themselves or were isomorphic with them. The second parity-fostering property is related to the first. It is that language forms should be preserved throughout a communicative exchange. That is, talkers should intend to convey a message composed of a sequence of phonological forms, their public actions should count as producing those forms for members of the language community, and the forms should be conveyed to listeners by acoustic signals that constitute information about them. In turn, listeners should perceive and recognize those language forms.

In general, linguists and psycholinguists do not agree that languages have these parity-fostering properties. In particular, they do not identify activities of the vocal tract as phonological forms. Rather, forms are components of linguistic competence in the mind of the language user. For example:

> Phonological representation is concerned with speakers' implicit knowledge, that is with information in the mind. (Pierrehumbert, 1990, p. 376)

> [Phonetic segments] are *abstractions*. They are the end result of complex perceptual and cognitive processes in the listener's brain . . . They have no physical properties. (Repp, 1981, p. 1462)

> Auditory coding of the signal is followed by processes that map the auditory representation onto linguistic units such as phonetic features, phonemes, syllables or words. (Sawusch & Gagnon, 1995, p. 635)

Notare language forms considered isomorphic with vocal tract activities. For example, MacNeilage and Ladefoged (1976, p. 90) remark that:

> there has been . . . an increasing realization of the inappropriateness of conceptualizing the dynamic processes of articulation itself in terms of discrete, static, context-free linguistic categories such as "phoneme" or "distinctive feature." This development does not mean that these linguistic categories should be abandoned – as there is considerable evidence for their behavioral reality (Fromkin, 1971). Instead it seems to require that they be recognized . . . as too abstract to characterize the actual behavior of the articulators themselves. They are, therefore, at present better confined to primarily characterizing earlier premotor stages of the production process . . . and to reflecting regularities at the message level.

This dichotomy between message level forms and physical implementations of speech remains today, a quarter century after publication of MacNeilage and Ladefoged's paper. For example, it is apparent in the comprehensive model of language production of Levelt, Roelofs, & Meyer (1999). There, phonological forms are abstract and featurally underspecified representations,[1] whereas the phonetic forms that drive articulation are the articulatory gestures of Browman and Goldstein (e.g., 1992; also see below Section 26.2.2.1).

The apparent mismatch between properties of phonological segments (or phonemes) and articulatory actions occurs in part because talkers coarticulate when they speak; that is, they overlap vocal tract activities for consonants and vowels temporally and spatially. Coarticulation is identified, for the most part, as destructive of some essential properties of phonological segments, in particular, their discreteness, their static nature and their context-invariance. Coarticulated consonants and vowels are analogous to smashed Easter eggs in Hockett's (1955) famous metaphor, they are distortions of phonetic segments according to Ohala (e.g., 1981), and they eliminate the possibility of articulatory or acoustic invariants corresponding to consonants and vowels according to Liberman and Mattingly (1985; see also Liberman, 1996).

These characterizations signify not only that the public actions that count as speaking are not language forms and are not isomorphic with them, but also, therefore, that language forms are not preserved throughout a communicative exchange. They are present as components of the talker's plan to produce an utterance, and, if the listener recovers the phonological message, they are known to the listener as well. However, they are not preserved in the talker's public actions. This means that the acoustic signal cannot provide certain information about the forms. Accordingly, the listener has to reconstruct the forms from such things as "auditory cues" (Sawusch & Gagnon, 1995) or acoustic and optical cues (e.g., Massaro, 1998). In this perspective, language forms reside in the minds of language users, not in the intermediate media – vocal tract, air, ear – that support communication.

We ask whether this perspective is realistic in effectively characterizing the fit between talking and listening as poor. We think that it is not. Rather, because research in speech production and perception is generally undertaken independently, researchers have not confronted the issue of their mutual fit. Here we begin with the hypothesis that languages do have the parity-fostering characteristics

listed above and ask whether we can eliminate the barriers that current thinking about speaking and listening has erected in the way of this hypothesis.

26.2.2 *Evidence*

26.2.2.1 The nature of phonological forms

By most accounts, phonological forms are collections of featural attributes. They are essentially timeless; they are discrete one from the other, and they are context-free. In all of these respects, they appear quite different from the articulatory actions of speaking and from the consequent acoustic signals. Articulatory actions are dynamic and overlapping, and the specific movements that constitute production of a consonant or a vowel are context-sensitive due to coarticulation. The acoustic speech signal likewise undergoes constant change, there are no phone-sized segments apparent in it, and the acoustic structure specifying a given consonant or vowel is highly context-sensitive.

By one account, however, atoms of phonological competence are not different in these ways from actions of the vocal tract during speech. This is the account provided by articulatory phonology (e.g., Browman & Goldstein, 1992, 1995).

Eliminating the differences between descriptions in the two domains requires adjustments in how we think about both the elements of phonological competence and articulatory actions. A crucial move in this direction is to assume that elements of phonological competence have their primary home in the vocal tract, not in the mind. They are linguistically significant actions of the vocal tract, called *gestures*. Gestures are not movements of individual articulators, but rather are coordinated actions usually of two or more articulators. The actions create and release constrictions. An example is the bilabial closure that occurs in production of English /b/, /p/, and /m/. Gestures are atoms of phonological competence as well (Browman & Goldstein, 1992).

Making this move necessarily eliminates the discrepancies between the language forms of phonological competence and the actions that implement them in speaking. Gestures are dynamic as they are produced and as they are perceived and known. Moreover, although they are produced in overlapping time frames, they are discrete. That is, in the syllable /bi/, for example, a constriction is made at the lips. Overlapping with that, temporally, the tongue body forms a constriction at the palate for /i/. Both constrictions are made; they are discrete in the sense of being distinct one from the other. Moreover, at a coarse-grained level of description, the two gestures are context-free. The lips always make contact for /b/; a palatal constriction is always made for /i/, regardless of the coarticulatory context. The synergistic relations among the articulators that contribute to a gesture allow the coarse-grained gestural action to be invariantly achieved, even though, at a finer-grained level of description, due to coarticulation, the movements that achieve the gesture are context-sensitive (cf. Abbs & Gracco, 1984; Kelso et al., 1984). Thus the jaw may contribute more to lip closure in the context of a coarticulating close vowel such as /i/ than in the context of an open vowel such as /a/.

For present purposes, the important achievement of articulatory phonology is in showing that languages can have the parity-fostering properties suggested

above. According to articulatory phonology, phonological forms are in fact the public actions in which speakers engage when they talk, and consequently, they may be preserved throughout a communicative exchange. They are the atoms of talkers' plans to speak and of their vocal tract activity. Moreover, because they are the immediate causes of structure in the acoustic speech signal, and because distinctive gestures structure the signal distinctively, the signal can provide information for the gestures. If listeners use this information as such and track gestures (Fowler, 1986, 1996), then phonological language forms are preserved throughout a communicative exchange.

Identifying phonological atoms as public gestures does not preclude their serving the roles that features have served in more traditional phonologies. For example, whether a speaker produces the word *big* or the word *dig* depends on the oral constriction gesture for the initial consonant of the word. Gestures minimally distinguish words (Browman & Goldstein, 1992).

We now consider evidence that the primitives of listeners' perceptions of speech are language forms like those proposed by articulatory phonologists. We consider four kinds of evidence.

26.2.2.2 *Evidence that listeners perceive phonological gestures*

The earliest findings that listeners perceive articulatory gestures was obtained by Liberman and his colleagues at Haskins Laboratories. Two findings provide complementary evidence. One (Liberman et al., 1954) is that, in two-formant synthetic syllables, /di/ and /du/, the critical information specifying that the initial consonant of each syllable is /d/ is physically quite different. It is a high rise in the frequency of the second formant transition in /di/ and a low fall in frequency in /du/. Separated from the remainder of the syllables, the transitions sound quite distinct, and neither sounds like /d/. In context, they sound alike. There is something alike about /di/ and /du/ when they are naturally produced. Both /d/ gestures are achieved by a constriction of the tongue tip against the alveolar ridge of the palate. The transitions, produced after release of the constrictions, are acoustically dissimilar because of coarticulation by the following vowel. In this instance, the same gesture, which has two distinct acoustic consequences (i.e., the distinct second formant transitions), is perceived as the same consonant. Perception tracks articulation.

The second finding (Liberman, Delattre, & Cooper, 1952) was obtained when voiceless stop consonants were cued by stop bursts rather than by formant transitions. In this case, an invariant burst, centered at 1440 Hz was identified predominantly as /p/ before /i/ and /u/ but as /k/ before /a/. Due to coarticulation, to produce the same stop burst in the different contexts requires a labial constriction before /i/ and /u/, but a velar constriction before /a/. Here, different gestures, giving rise because of coarticulation to the same bit of acoustic structure, are perceived as different.

Both findings appear to show that "when articulation and the sound wave go their separate ways" (Liberman, 1957, p. 121), perception tracks articulation.

There is another kind of finding showing that speech perception tracks articulation. This finding concerns how listeners parse the acoustic speech signal to recover phonological forms. They parse the signal along gestural lines.

Due to coarticulation, different gestures can have converging effects on common acoustic dimensions. For example, production of an unvoiced consonant may cause a high falling fundamental frequency (F0) pattern on a following vowel (e.g., Silverman, 1987). This may occur because the vocal folds are tensed to keep them apart during consonant production (e.g., Löfqvist et al., 1989). When they are adducted for the following vowel, the tension will raise F0 at vowel onset. (For other accounts, see Kingston & Diehl, 1994.) Likewise, a high vowel is associated with a higher F0 than a low vowel, an outcome that also is likely to have a cause in production constraints (Whalen & Levitt, 1995; Whalen et al., 1995). These segmental effects are superimposed on the intonation contour of the larger utterance in which they are produced. However, neither is heard as part of the intonation contour or even as pitch (e.g., Fowler & Brown, 1997; Pardo & Fowler, 1997; Silverman, 1987). Rather, the F0 contour caused by the voiceless consonant contributes to the perception of voicelessness (e.g., Pardo & Fowler, 1997; Silverman, 1986); that caused by vowel height contributes to the perception of vowel height (e.g., Reinholt Peterson, 1986). Listeners parse acoustic speech signals along gestural lines.

The sight of a human face mouthing one syllable dubbed onto a different, acoustically presented, syllable can lead listeners to hear something different than they hear in the absence of the video display. For example, acoustic /ma/ dubbed onto a face mouthing /da/ will be reported most frequently as /na/, a percept that integrates the visible alveolar gesture with the acoustically specified voicing and nasality (McGurk & MacDonald, 1976).

An analogous effect occurs when the haptic feel of consonantal gestures is substituted for the visible face (Fowler & Dekle, 1991). Although explanations for the original McGurk effect have been proposed that do not invoke perception of gestures as the common currency allowing audio-visual integration of phonetic information (e.g., Massaro, 1998), we believe that the haptic findings do require a gestural account. This interpretation leads to the prediction that, when print replaces the visible or felt facial gestures, the McGurk effect should disappear because print is not immediately caused by vocal tract gestures. Under conditions like those of the haptic experiment, Fowler and Dekle found no effect of print on acoustic speech perception. Given this set of findings, and the fact that visually and haptically perceived faces provide information about the same gestural events, we interpret these data sets as converging evidence in favor of the claim that listeners perceive gestures.

Another source of evidence that listeners perceive gestures concerns the rapidity of imitation. Canonically, choice response times exceed simple response times by 100 to 150 ms (Luce, 1986). In a characteristic choice task, participants might push one button when a green light flashes and a different button if a blue light flashes. In the simple task, they hit the same response button whenever any light flashes, whether it is green or blue.

When stimuli and responses are spoken utterances, the difference between choice and simple response latencies can become quite small, with both sets of latencies near those of rather fast simple response times (Porter & Castellanos, 1980; Porter & Lubker, 1980). This suggests that the element of choice in the choice task has been reduced. In the choice task as implemented by Porter and

colleagues, a model speaker produced extended /a/ and then, after an unpredictable interval, shifted to something else, say, /ba/, /da/, or /ga/. Participants shadowed the model's disyllable. In the simple task, participants were assigned a syllable (say, /ba/): as in the choice task, they shadowed the model's extended /a/, but when the model shifted to /ba/, /da/, or /ga/, the participants produced their designated syllable, no matter what the model uttered. Porter and Castellanos (1980) found a 50 ms difference between simple and choice response times; Porter and Lubker (1980) found an even smaller difference. These results suggest that the participant's production of syllables in the choice task benefits from hearing the same syllable as the signal that prompts responding as if the signal serves as instructions for the required response. We have recently replicated Porter and Castellanos' experiment (Fowler et al., 2003), and we found a 26 ms difference between choice and simple response times with average simple responses times around 160 ms.

If listeners perceive gestures, these results are easy to understand. In the choice task, perceiving the model's speech is perceiving instructions for the required phonetic gestural response. We obtained two additional outcomes consistent with this interpretation. First, in the simple task, on one-third of the trials, the syllable that the participant produced was the same as the model's syllable. In our task, the responses were /pa/, /ta/, and /ka/. For a participant whose designated syllable was /pa/, the response matched the model's on trials when the model said /pa/. If listeners perceive gestures and the percept serves as a goad for an imitative response, responses should be faster on trials in which the model produced the participant's designated syllable than on other trials. We found that they were. Second, in a subsequent experiment, in which only choice responses were collected, we nearly doubled the voice onset times (VOTs) of half the model's syllables by manipulating the speech samples, and we asked whether our participants produced longer VOTs on those trials than on trials with original model VOTs. They did, and we concluded that our participants were, in fact, perceiving the model's speech gestures (specifically, the particular phasing between oral constriction and laryngeal devoicing), which served as a goad for an imitative response.

26.2.3 *Conclusion regarding the fit between the public actions of talking and listening to speech*

We find no convincing barriers to the idea that phonological forms are public actions, and we find evidence in its favor. If phonological forms are the public actions of speakers when they talk – that is, if phonological forms are gestures, and if, as the evidence suggests, listeners perceive gestures – then languages do meet the requirements of parity. Language forms are the public actions of the vocal tract during speech, and they are preserved throughout a communicative exchange.

We next consider the relations between the mechanisms that support talking and listening to speech.

26.3 The Relation between Mechanisms for Talking and Listening

26.3.1 *The motor theory*

In Liberman and Mattingly's theory (1985) there is another way in which speech perception and speech production are related, aside from their sharing public language forms. Parity in communication may also be fostered if the mechanisms for producing and perceiving speech are related. Liberman and Mattingly (1985, 1989) suggested that such a relation is realized in a speech module, that is, a dedicated piece of neural circuitry that evolved as a specialization for producing and perceiving speech.

However, the existence of such a module was not inferred solely on the basis of the theoretical constraints imposed by the requirements of parity. Its existence was suggested by empirical observations and by the flood of research findings that accompanied the development of speech technology.

A first step along the path that led to the proposal of the speech module was the realization that speech is not an acoustic alphabet. This conclusion was based on findings that, when a linguistic message is produced as a sequence of discrete acoustic units, it cannot be perceived at practically useful rates (Liberman et al., 1967). Notably, such a conclusion does not hold in general. In the visual modality, for example, speech can be rendered alphabetically, both in production (i.e., writing) and in perception (reading). The essence of the discovery by Liberman and colleagues was that an acoustic analog of an orthographic alphabet was not workable (cf. also Harris, 1953), even with carefully designed alphabets and extensive training of alphabet learners.

This conclusion suggested that a dedicated mechanism to handle speech may exist: If speech cannot be replaced by an acoustic alphabet, then its perception may require machinery different from the kind of machinery that handles print and perhaps different from the machinery that handles other acoustic sequences, such as Morse code.

The second step occurred when a more thorough understanding of the speech signal became available due to the advent of spectrograms. One of the earliest discoveries was that, in real speech, acoustic information about phonemes is not temporally discrete. A given bit of acoustic signal can contain information about several phonemes and, conversely, one phoneme can influence the acoustic signal for a period of time longer than its length as conventionally measured (that is, as a discrete acoustic interval). These discoveries deepened the puzzle of the relation between the physical instantiation of speech and its linguistic units. The speech signal is continuous, and it codes phonetic information in a highly parallel fashion, yet the phonetic percept is discrete and sequential. However, what appears to be a complicated puzzle for the scientist may be an optimal solution for nature. On one side, if the physical instantiation of speech were a sequential signal made of discrete units, then speech would be a highly inefficient communication bearer in the acoustic medium.[2] On the other side, if linguistic units, specified in parallel in the signal, were to blend in perception, then listeners would not perceive the particulate units that are the atoms of open and productive phonological systems

(Studdert-Kennedy, 2000). A transformation from continuous-parallel information to discrete, sequential units in perception and a transformation going the other way in production seemed to be central to the design of human languages. But, how are these transformations achieved?

This was another indication that a dedicated mechanism exists to handle speech: the perceptual system for speech must be capable of transforming continuous-parallel acoustic structure into discrete-sequential phonological entities rapidly and accurately. No other acoustic percept imposes such requirements on the auditory system; this peculiarity calls for a specialized module.

The third step was closely related to the second. Speech is a physically continuous-parallel signal because, when it is produced, the vocal tract massively coarticulates the discrete phonetic units. In the view of Liberman and colleagues (e.g., Liberman et al., 1967), coarticulation is necessary to evade limits on the temporal resolving power of the ear. The capacity to coarticulate must have coevolved with that of decoding the effects of coarticulation in the acoustic signal, because neither capability would be useful without the other. Perhaps, then, these capabilities are both grounded in a common mechanism, identified by Liberman and Mattingly (1985) as a phonetic module.

The module evolved as a cortical structure shared between speech perception and production; its primary goal is to make motor knowledge about the effects of coarticulation available to the perceptual system. It is a compact evolutionary solution to the problem of coding discrete-sequential messages in a continuous-parallel acoustic signal, and it provides a rationale for the observation that perception tracks articulation.

In a later development, Liberman and Mattingly (1985) adopted as perceptual objects the phonetic gestures of Browman and Goldstein's (1986) articulatory phonology (see Section 26.2.2.1 above). That is, Liberman and Mattingly (1985) proposed that listeners perceive gestures, not individual movements of individual articulators. However, they preserved their earlier idea that coarticulation in speech destroys the discrete character of phonetic units in the acoustic signal, gestures no less than classical consonants and vowels (contrary to our earlier proposal). Accordingly, the gestures that listeners perceive, in the theory, are intended, not actual, gestures. The phonetic module enabled recovery of intended gestures from highly encoded speech signals.

Still later, Liberman and Mattingly (1989) explored the consequences of having postulated a phonetic module. Specifically, they attempted to locate the module within the architecture of the auditory system. To this end, they distinguished open and closed modules. Open modules (also called horizontal systems) are all-purpose devices that provide information about the energy distribution patterns detected by sensory systems. They are open in the sense that they can adaptively adjust to new environmental situations. The percepts they render are *homomorphic* with (that is, have the same form as) the proximal stimulation that causes them. In the case of the auditory system, for example, pitch is the homomorphic percept for frequency, and loudness is the homomorphic percept for intensity. Closed modules (also called vertical systems) are special-purpose devices that provide information about the distal structure that is behind the proximal energy distribution patterns detected by the sensory systems. They are closed in the sense that, being highly specialized for a particular kind of stimulation, they cannot

adapt to new environmental situations. The percepts they render are *heteromorphic* with respect to proximal stimulation in that they have the same form as the distal events that cause the proximal stimulation. For example, speech perception and sound localization yield heteromorphic percepts (phonetic gestures and the location of a sounding source, respectively).

If speech percepts are attributed to a closed module, the question arises about their relationship with other auditory percepts, those coming from the open module as well as those coming from other closed modules. Liberman and Mattingly proposed that closed modules serially precede open modules, preempting the information that is relevant for their purposes and passing along whatever information is left to open modules.[3] This particular architectural design leads to the possibility of *duplex perception*, that is, the phenomenon that occurs when information left after preemption by the closed modules gives rise to homomorphic percepts: Homomorphic and heteromorphic percepts are simultaneously produced in response to the same stimulus.

26.3.2 *Evidence especially favoring a motor theory of speech perception*

The motor theory makes three closely related claims. It claims that listeners perceive intended gestures, that perception is achieved by a module of the nervous system dedicated to speech production and perception, and that speech perception recruits the speech motor system.

We have reviewed some of the evidence suggesting that gestures are perceived. Here we review evidence relating to the other two claims, focusing largely on the third.

The strongest behavioral evidence for a dedicated speech processing system is provided by findings of duplex perception. In one version of this finding, listeners are presented with a synthetic /da/ or /ga/ syllable, where /da/ and /ga/ were synthesized to be identical except for the third formant transition, which falls for /da/ and rises for /ga/. If the part of the syllable that is the same for /da/ and /ga/ (called the "base") is presented to one ear and the distinguishing transition is presented to the other, listeners integrate the information across the ears and hear /da/ or /ga/, depending on the transition. However, at the same time, they also hear the transition as a pitch rise or fall (e.g., Mann & Liberman, 1983). The finding that part of the signal is heard in two different ways at the same time suggests that two different perceptual systems are responsible for the two percepts. One, a phonetic processor, integrates the base and the transition and yields a phonetic percept, /da/ or /ga/. The other yields a homomorphic percept. Presumably this is an auditory processor, an open module.

This interpretation has been challenged on a variety of grounds (e.g., Fowler & Rosenblum, 1990; Pastore et al., 1983). We will not review those challenges here. Rather, we note that the motor theoretical interpretation of duplex perception would be buttressed by evidence favoring the third claim of the theory, that there is motor system or motor competence involvement in perceiving speech. This is because, generally, theorists do not claim motor involvement in auditory perception.

In fact, evidence for motor involvement in speech perception is weak. However, apparently this is because such evidence has rarely been sought, not because many tests have yielded negative outcomes. We have found three sets of supportive behavioral data and some suggestive neuropsychological data.

Following a seminal study by Eimas and Corbit (1973), there were many investigations of "selective adaptation" in speech perception. Listeners heard repeated presentations of a syllable at one end of an acoustic continuum, say /pa/, and then identified members of, say, a /pa/ to /ba/ continuum. After hearing repeated /pa/ syllables, listeners reported fewer /pa/s in the ambiguous region of the continuum. Eimas and Corbit suggested that phonetic feature detectors (a detector for voicelessness in the example) were being fatigued by the repetitions, making the consonant with that feature less likely to be perceived than before adaptation. Although this account was challenged (e.g., by Diehl, Kluender, & Parker, 1985), for our purposes, the interpretation is less important than the finding by Cooper (1979) that repeated presentations of a syllable such as /pi/ had weak but consistent effects on *production* of the same syllable or another syllable sharing one or more of its features. For example, VOTs of produced /pi/s and /ti/s were reduced after adaptation by acoustic /pi/. This finding implies a perception-production link of the sort proposed by the motor theory.

Bell-Berti et al. (1978) provided further behavioral evidence for a motor theory. The vowels /i/, /ɪ/, /e/, and /ɛ/ of English can be described as differing in either of two ways. They decrease in height in the series as listed above. Alternatively, /i/ and /e/ are described as tense vowels; /ɪ/ and /ɛ/ are their lax counterparts. Within the tense vowel pair and the lax pair, vowels differ in height. Bell-Berti et al. found that speakers differed in how they produced the vowels in the series in ways consistent with each type of description. Four of their ten speakers showed activity of the genioglossus muscle (a muscle of the tongue affecting tongue height) that gradually decreased in the series of four vowels as listed above suggesting progressively lower tongue heights. The remaining six speakers showed comparable levels of activity for /i/ and /e/ that were much higher than activity levels for the two lax vowels. This suggested use of a tense-lax differentiation of the vowels.

In a perception test, the ten participants partitioned into the same two groups. Listeners identified vowels along an /i/ to /ɪ/ continuum under two conditions. In one, the vowels along the continuum were equally likely to occur. In the other condition, an anchoring condition, the vowel at the /i/ end of the continuum occurred four times as frequently as the other continuum members, a manipulation that decreases /i/ responses. The magnitude of this anchoring effect differed in the two groups of talkers; across the ten participants, the effect magnitude had a bimodal distribution. Participants who had shown progressively decreasing levels of genioglossus activity in their production of the four vowels showed considerably larger effects of anchoring than the six speakers who produced /e/ with more genioglossus activity than /ɪ/. The authors speculated that the difference occurred because, for the second group of listeners, /i/ and /ɪ/ are not adjacent vowels, whereas they are for members of the first group. Whether or not this is the appropriate account, it is remarkable that the participants grouped in the same way as listeners as they had as talkers. This provides evidence suggesting that speech percepts include information about motor production of speech.

Kerzel and Bekkering (2000) provide additional behavioral findings that they interpret as consistent with the motor theory. They looked for compatibility effects in speech production. On each trial, participants saw a face mouthing /bʌ/ or /dʌ/. At a variable interval after that, they saw either of two symbol pairs (in one experiment, ## or &&) that they had learned to associate with the spoken responses /ba/ and /da/. Kerzel and Bekkering found an effect of the irrelevant visible speech gesture on latencies to produce the syllables cued by the symbols such that /ba/ responses were faster when the face mouthed /bʌ/ than when it mouthed /dʌ/. Likewise /da/ responses were facilitated by visible /dʌ/. Kerzel and Bekkering argued that these effects had to be due to stimulus (visible gesture)-response compatibility, not stimulus-stimulus (that is, visible gesture-visible symbol) compatibility, because the symbols (## and &&) bear an arbitrary relation to the visible gestures whereas the responses do not. Their interpretation was that the visible gestures activated the speech production system and facilitated compatible speech actions, an account consistent with the motor theory. It has yet to be shown that acoustic speech syllables, rather than visible speech gestures, have the same effect.

There is some recent neuropsychological evidence providing support for a motor theory of speech perception. Calvert and colleagues (1997) reported that auditory cortical areas activate when individuals view silent speech or speech-like movements. Moreover, the region of auditory cortex that activated for silent lipreading and for acoustic speech perception was the same. More recently, they (MacSweeney et al., 2000) replicated the findings using procedures meant to ensure that fMRI scanner noise was not the source of the auditory cortical activation.

Using transcranial magnetic stimulation of the motor cortex, Fadiga et al. (2002) found enhanced muscle activity in the tongue just when listeners heard utterances that included lingual consonants. Conversely, using PET, Paus et al. (1996) found activation of secondary auditory cortex, among other brain regions, when participants whispered nonsense syllables with masking noise to prevent their hearing what they produced.

26.3.3 *The larger context in which the motor theory can be evaluated and supported*

The motor theory's claim that a linkage exists in speech mechanisms supporting speech production and perception receives additional support when it is considered in a larger context of research. Liberman (e.g., 1996) proposed that a production-perception link was a special solution to a special problem: the necessity of parity achievement in human spoken communication. We propose to deny that the link is special to the mechanisms used to implement speech (Fowler & Rosenblum, 1990). Rather, we suggest that linkages between motor and perceptual systems are blueprints of the architecture of cognition, above and beyond speech, and even above and beyond communication devices. Let us consider the evidence.

26.3.3.1 *Above and beyond speech*

Some species that use acoustic signals to recognize mates have linkages between the systems underlying the production of sounds in one animal and those

underlying their perception by its mate. For example, evidence has been reported for a genetic coupling in crickets and frogs of the mechanisms for sound production by males and for sound perception by females (Doherty & Gerhardt, 1983; Hoy, Hahn, & Paul, 1997). Although the exact nature of the genetic mechanisms that support the linkages is still debated (Boake, 1991; Butlin & Ritchie, 1989; Jarvis & Nottebohm, 1997), there is agreement that the production and perception systems have coevolved (Blows, 1999). The motor system of the sender and the perceptual system of the receiver have shaped one another, permitting mate recognition and thus, the possibility of preserving the species or, when the parity constraint is significantly violated, of differentiating them by speciation (Ryan & Wilczynski, 1988).

The existence of linkages between production and perception of mating signals is confirmed also, at an anatomo-physiological level, for songbirds such as zebra finches (Williams & Nottebohm, 1985), canaries (Nottebohm, Stokes, & Leonard, 1976), and white-sparrows (Whaling et al., 1997), and for other birds such as parrots (Plummer & Striedter, 2000). In these animals, the neural motor centers that underlie song or sound production are sensitive to acoustic stimulation. The neural centers that support sound production in parrots are different from those that support song production for the songbirds. However, for all of these birds, there is an increase in auditory responsivity of the motor nuclei as the similarity between the acoustic stimulation and the song or sound produced by the bird itself or its conspecifics increases. The fact that over very different taxa, and through different mechanisms, a linkage is present between perception and production of acoustic communication signals is a first strong suggestion that motor-perceptual interactions may be more general than the special adaptation proposed by the motor theory of speech perception.

26.3.3.2 *Above and beyond communication systems*

Although it is suggestive, the evidence we summarized above is limited in its scope, because the production-perception linkages in crickets, frogs, and birds are all in the domain of animal communication. The emergence of these perception-action linkages might be considered a special solution to a common special problem, that of achieving parity in communication systems. But other evidence suggests that linkages between motor and perceptual systems are ubiquitous and are not specific to the requirements of communication.

Viviani and colleagues (see Viviani & Stucchi, 1992 for a review) have demonstrated experimentally that the motor system is brought to bear on visual and haptic perception of movements (Kandel, Orliaguet, & Viviani, 2000; Viviani, Baud-Bovy & Redolfi, 1997; Viviani & Mounod, 1990; Viviani & Stucchi, 1989, 1992). They infer that motor competence is brought to bear on perception whenever the two-thirds power law, a law that they consider a signature of biological motion, manifests itself as a constraint shaping perception of motion.[4]

For example, Viviani and Stucchi (1992) presented observers with a light spot moving along various continuous trajectories on a computer screen. Participants were asked to adjust the velocity profile of the motion to make it look uniform. In line with previous observations by Runeson (1974), Viviani and Stucchi found that participants judged as uniform motions that were, by objective measurement,

highly variable.[5] This occurred even when observers were shown examples of uniform motion. The velocity profiles of the motions chosen as uniform all closely fit the two-thirds power law; thus, viewers perceive as uniform motions that are uniform only in their close obedience to the laws that govern biological movements. Similar effects of adherence to the two-thirds power law in perceptual performance are shown in visual judgments of motion trajectories (Viviani & Stucchi, 1989), pursuit tracking of two-dimensional movements (Viviani & Mounod, 1990), and motoric reproductions of haptically felt motions (Viviani, et al., 1997).

Other evidence for linkages between the motor and perceptual systems comes from experiments that manipulate stimulus-response compatibility and show facilitation or inhibition of motor performance due, respectively, to a compatible or incompatible perceptual stimulus that signals the initiation of the response movement (for a review, see Hommel et al., 2001).

In particular, Stürmer, Aschersleben & Prinz (2000), extended to the domain of hand gestures results like those obtained in the speech domain by Kerzel and Bekkering (2000). Their participants had the task of producing either a grasping gesture (first close the hand from a half-open position then return to half-open) or a spreading gesture (first open from a half-open position then return to half-open). The go signal for the movement was presented on a video, and it consisted of a color change on a model's hand, with different colors signaling the different gestures to be performed. Initially the hand was skin-colored; then it changed to red or blue. Along with the go signal, at varying latencies relative to the go signal, the model's hand produced either of two gestures that the participants were performing. Although participants were told to ignore the irrelevant information, they were faster to produce their responses when the movement matched the one presented on the computer screen. This finding is consistent with studies of speech reviewed earlier (Fowler, et al., 2003; Porter & Castellanos, 1980; Porter & Lubker, 1980) showing that responses are facilitated when stimuli cuing them provide instructions for their production. It is interesting that the same effect occurs for nonspeech (hand gestures in the research of Stürmer et al.) as well as for speech gestures.

26.3.3.3 Neuroimaging

Following Rizzolatti and colleagues' discovery of "mirror neurons"[6] in the premotor cortex of monkeys (see for example, Rizzolatti, 1998; Rizzolatti & Arbib, 1998; Rizzolatti et al., 1996) evidence has accumulated that a neural system exists in primates, including humans, for matching observed and executed actions (for a review see Decety & Grezes, 1999).

For example, using fMRI, Iacoboni et al. (1999) found that two cortical regions selectively engaged in finger-movement production – the left frontal operculum (area 44) and the right anterior parietal cortex – showed a significant increase in activity when the movements were imitations of movements performed by another individual compared to when the movements were produced following non-imitative spatial or symbolic cues. Remarkably, one of the two regions – area 44 – includes Broca's area, one of the important cortical regions for the production of speech.

Strafella & Paus (2000), used transcranial magnetic stimulation to demonstrate that perceiving handwriting is accompanied by an increase in the activity of the muscles of the hand (first dorsal interosseus); perceiving arm movements is accompanied by an increase in activity of muscles of the arm (biceps).

The foregoing is just a sampling of the evidence for perception-production linkages either in behavior or in the mechanisms supporting perception and action. In the context of these findings, the specific claim of the motor theory of motor recruitment in speech perception accrues considerable credibility.

26.4 Conclusions

For speech to serve its public communication function, listeners must characteristically perceive the language forms that talkers produce. We call this a requirement for achieving parity, following Liberman and colleagues (cf. Liberman & Whalen, 2000). We suggested that properties of language have been shaped by the parity requirement, and a significant example of this shaping is language's use of phonetic gestures as atoms of phonological competence, of speech production, and of speech perception. We provided evidence that gestures are perceived.

A review of behavioral and neuropsychological evidence within the speech domain, within the study of communication systems more generally, and in the larger domain of perception as well, uncovers evidence for linkages between mechanisms that support motor performance and those that support perception. This body of evidence is in favor of one claim of the motor theory while disfavoring another one. It favors the motor theory's claim of a link between speech production and perception mechanisms. Although within the speech domain there is weak evidence for this claim, its plausibility increases significantly when we consider the ubiquity of evidence for perception-production links across the board in cognition. This ubiquity, in turn, disfavors the motor theory's claim that a linkage is special to speech in providing a special solution to a special perceptual problem.

A next question to be addressed by future research is why perception-production linkages are so pervasive. Do these links reflect a general solution to a general problem? Rizzolatti and Arbib (1998) suggest that the perception-action links that mirror neurons support in primates provide an empathic way of recognizing the actions of others, possibly another kind of parity achievement. Perhaps achieving parity between the world as perceived and the world as acted upon is the main function of cognitive systems (Gibson, 1966), whether or not mirror neurons underlie them.

We began this chapter by remarking that the closeness of the fit between the activities of speaking and perceiving speech has not been frequently addressed. We now conclude by asking why, given the ubiquity of such linkages, cognitive science generally continues to investigate perception and action as if they were independent, especially in the domain of speech. We do not know the answer to this question. However, we do know that, when the assumption of logical independence between perception and action is made, then the *problem* of their relation arises. We have tried to show that there is another possible approach, in which the apparent problem becomes a resource. We assumed logical dependence between perception and action, and we found that this assumption forced us to

sharpen our understanding of language as a unitary phenomenon, significantly eroding the barriers that have been erected between its physical instantiations and its abstract nature.

ACKNOWLEDGMENTS

Preparation of the manuscript was supported by NIH grants HD-01994 and DC-03782 to Haskins Laboratories.

NOTES

1 That is, phonological forms are represented as bundles of features, but the forms are underspecified in that their predictable features are not represented.
2 According to Liberman et al. (1967), it would have a rate of transmission equivalent to that of Morse code, approximately ten times slower than normal speech.
3 Notice that this architecture does not specify the relationship between two or more closed modules. Liberman and Mattingly suggested that closed modules are not arranged in parallel because that would make preemption cumbersome, but they left to empirical investigation the question of the relations among them.
4 In brief, the law states that when humans make curved movements, the angular velocity of their movements is proportional to the two-thirds power of the curvature.
5 The difference between minima and maxima was above 200%.
6 These are neurons that respond both when an action such as grasping an object is performed and when the same action by another animal is perceived.

REFERENCES

Abbs, J. & Gracco, V. (1984). Control of complex gestures: Orofacial muscle responses to load perturbations of the lip during speech. *Journal of Neurophysiology*, 51, 705–23.

Bell-Berti, F., Raphael, L. R., Pisoni, D. B., & Sawusch, J. R. (1978). Some relationships between speech production and perception. *Phonetica*, 36, 373–83.

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-Biological Sciences*, 266, 2169–74.

Boake, C. R. B. (1991). Coevolution of senders and receivers of sexual signals: Genetic coupling and genetic correlations. *Trends in Ecology & Evolution*, 6, 225–7.

Browman, C. & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219–52.

Browman, C. & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–80.

Browman, C. & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. Port & T. van Gelder (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (pp. 175–93). Cambridge, MA: MIT Press.

Butlin, R. K. & Ritchie, M. G. (1989). Genetic coupling in mate recognition systems: What is the evidence? *Biological Journal of the Linnean Society*, 37, 237–46.

Calvert, G., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P., Woodruff, P. W. R., Iverson, S., D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276, 593–6.

Cooper, W. E. (1979). *Speech Perception and Production: Studies in Selective Adaptation*. Norwood, NJ: Ablex Publishing Company.

Decety, J. & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences*, 3, 172–8.

Diehl, R., Kluender, K., & Parker, E. (1985). Are selective adaptation effects and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception and Performance*, 11, 209–20.

Doherty, J. A. & Gerhardt, H. C. (1983). Hybrid tree frogs: Vocalizations of males and selective phonotaxis of females. *Science*, 220, 1078–80.

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

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.

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

Fowler, C. A. (1996). Listeners do hear sounds not tongues. *Journal of the Acoustical Society of America*, 99, 1730–41.

Fowler, C. A. & Brown, J. (1997). Intrinsic F0 differences in spoken and sung vowels and their perception by listeners. *Perception & Psychophysics*, 59, 729–38.

Fowler, C. A. & Dekle, D. J. (1991). Listening with eye and hand: Crossmodal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 816–28.

Fowler, C. A. & Rosenblum, L. D. (1990). Duplex perception: A comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 742–54.

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

Fromkin, V. (1971). The nonanomalous nature of anomalous utterances. *Language*, 47, 27–52.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton-Mifflin.

Harris, C. (1953). A study of the building blocks in speech. *Journal of the Acoustical Society of America*, 25, 962–9.

Hockett, C. (1955). *A Manual of Phonetics*. Bloomington, IN: Indiana University Press.

Hockett, C. (1960). The origin of speech. *Science*, 203, 89–96.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849–78.

Hoy, R. R., Hahn, J., & Paul, R. C. (1977). Hybrid cricket auditory-behavior: Evidence for genetic coupling in animal communication. *Science*, 195, 82–4.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286, 2526–8.

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

Sciences of the United States of America, 94, 4097–102.

Kandel, S., Orliaguet, J.-P., & Viviani, P. (2000). Perceptual anticipation in handwriting: The role of implicit motor competence. *Perception & Psychophysics*, 62, 706–16.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally-specific articulatory cooperation following jaw perturbation during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 812–32.

Kerzel, D. & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus-response compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 634–47.

Kingston, J. & Diehl, R. (1994). Phonetic knowledge. *Language*, 70, 419–54.

Levelt, W. J. M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29, 117–23.

Liberman, A. M. (1996). *Speech: A Special Code*. Cambridge, MA: Bradford Books.

Liberman, A. M. & Mattingly, I. (1985). The motor theory revised. *Cognition*, 21, 1–36.

Liberman, A. M. & Mattingly, I. (1989). A specialization for speech perception. *Science*, 243, 489–94.

Liberman, A. M. & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4, 187–96.

Liberman, A. M., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–61.

Liberman, A. M., Delattre, P., & Cooper, F. (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 and Applied*, 68, 1–13.

Löfqvist, A., Baer, T., McGarr, N., & Seider Story, R. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America*, 85, 1314–21.

Luce, R. D. (1986). *Response Times*. New York: Oxford University Press.

MacNeilage, P. & Ladefoged, P. (1976). The production of speech and language. In E. C. Carterette & M. P. Friedman (eds.), *Handbook of Perception: Language and Speech* (pp. 75–120). New York: Academic Press.

MacSweeney, M., Amaro, E., Calvert, G., Campbell, R., David, A. S., McGuire, P., Williams, S. C. R., Woll, B., & Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. *Neuroreport*, 11, 1729–33.

Mann, V. & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Perception & Psychophysics*, 14, 211–35.

Massaro, D. (1998). *Perceiving Talking Faces*. Cambridge, MA: MIT Press.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 747–8.

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in canary, serinus-Canarius. *Journal of Comparative Neurology*, 165, 457–86.

Ohala, J. (1981). The listener as a source of sound change. In C. Masek, R. Hendrick, R. Miller, & M. Miller (eds.), *Papers from the Parasession on Language and Behavior* (pp. 178–203). Chicago: Chicago Linguistics Society.

Pardo, J. & Fowler, C. A. (1997). Perceiving the causes of coarticulatory acoustic variation: Consonant voicing and vowel pitch. *Perception & Psychophysics*, 59, 1141–52.

Pastore, R., Schmuckler, M., Rosenblum, L., & Szczesiul, R. (1983). Duplex perception for musical stimuli. *Perception & Psychophysics*, 33, 469–74.

Paus, T., Perry, D., Zatorre, R., Worsley, K. & Evans, A. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience*, 8, 2236–46.

Pierrehumbert, J. (1990). Phonological and phonetic representations. *Journal of Phonetics*, 18, 375–94.

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

Porter, R. & Castellanos, F. X. (1980). Speech production measures of speech perception: Rapid shadowing of VCV syllables. *Journal of the Acoustical Society of America*, 67, 1349–56.

Porter, R. & Lubker, J. (1980). Rapid reproduction of vowel-vowel sequences: Evidence for a fast and direct acoustic-motoric linkage. *Journal of Speech & Hearing Research*, 23, 593–602.

Reinholt Peterson, N. (1986). Perceptual compensation for segmentally-conditioned fundamental-frequency perturbations. *Phonetica*, 43, 31–42.

Repp, B. (1981). On levels of description in speech research. *Journal of the Acoustical Society of America*, 69, 1462–4.

Rizzolatti, G. (1998). Recognizing and understanding motor events. *International Journal of Psychophysiology*, 30, 6.

Rizzolatti, G. & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21, 188–94.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131–41.

Runeson, S. (1974). Constant velocity: Not perceived as such. *Psychological Research-Psychologische Forschung*, 37, 3–23.

Ryan, M. J. & Wilczynski, W. (1988). Coevolution of sender and receiver: Effect on local mate preference in cricket frogs. *Science*, 240, 1786–8.

Sawusch, J. & Gagnon, D. (1995). Auditory coding, cues and coherence in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 635–52.

Silverman, K. (1986). F0 cues depend on intonation: The case of the rise after voiced stops. *Phonetica*, 43, 76–92.

Silverman, K. (1987). The structure and processing of fundamental frequency contours. Unpublished PhD dissertation, Cambridge University.

Strafella, A. P. & Paus, T. (2000). Modulation of cortical excitability during action observation: A transcranial magnetic stimulation study. *Neuroreport*, 11, 2289–92.

Studdert-Kennedy, M. (2000). Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function. In C. Knight, M. Studdert-Kennedy, & J. Hurford (eds.), *The Evolutionary Emergence of Language* (pp. 161–76). Cambridge: Cambridge University Press.

Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effect with manual gestures and postures: A study of imitation. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1746–59.

Viviani, P. & Mounoud, P. (1990). Perceptuomotor compatibility in pursuit tracking of two-dimensional movements. *Journal of Motor Behavior*, 22, 407–43.

Viviani, P. & Stucchi, N. (1989). The effect of movement velocity on form perception: Geometric illusions in dynamic displays. *Perception & Psychophysics*, 46, 266–74.

Viviani, P. & Stucchi, N. (1992). Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 603–23.

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 and Performance*, 23, 1232–52.

Whalen, D. & Levitt, A. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23, 349–66.

Whalen, D., Levitt, A., Hsaio, P., & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9-, and 12-month old French- and English-learning infants. *Journal of the Acoustical Society of America*, 97, 2533–9.

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, 12694–8.

Williams, H. & Nottebohm, F. (1985). Auditory responses in avian vocal motor neurons: A motor theory for song perception in birds. *Science*, 229, 279–82.