Fowler, C. A., & Xie, X. (2016). Involvement of the speech motor system in speech perception. In P. van Lieshout, B. Maassen, and H. Terband (eds.). *Speech motor control in normal and disordered speech: Future developments in theory and methodology* (pp. 1-24). Rockville, MD: ASHA Press.

Chapter 1

Involvement of the Motor System in Speech Perception

Carol A. Fowler[a,b] and Xin Xie[a]

[a]University of Connecticut, Storrs

[b]Haskins Laboratories, New Haven, CT

Correspondence to Carol A. Fowler: carol.fowler@haskins.yale.edu

Abstract

We review some of the findings of motor system activation and involvement in speech perception. We address the findings in relation to Liberman's motor theory of speech perception and some other more recent proposals that provide an account of those findings. We conclude that although findings show convincingly that motor system activation occurs during speech perception (and is not limited to audiovisual speech perception), specific claims—for example, of analysis by synthesis—are not supported. We then review a broader array of findings of motor system activation and involvement in perception and cognition quite generally. We suggest that the roles of motor system involvement in speech are shared with the roles of the motor system's involvement in these broader domains. Very speculatively, we suggest that one role relates to the need for within- and between-person perceptuomotor coupling. We propose some lines for further research to test those speculations and to further explore the roles of motor activation in speech.

In this chapter, we review research findings showing that there is speech motor system involvement in perceiving speech. These findings are most relevant to the motor theory of speech perception proposed by Alvin M. Liberman and colleagues (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; Liberman & Whalen, 2000). Accordingly, we discuss the findings in the context of the claims of that theory, showing that they support some, but not all, of them.

However, recent findings of motor system involvement in speech perception have not always been interpreted as evidence favoring Liberman's motor theory. Rather, they have led to new ideas about the motor system and speech perception (e.g., Hickok & Poeppel, 2007; Schwartz, Basirat, Ménard, & Sato, 2012; Skipper, Nusbaum, & Small, 2005, 2006; Skipper, van Wassenhove, Nusbaum, & Small, 2007). Having reviewed a sampling of the research demonstrating motor system involvement, we review some of those newer ideas as well.

Another development over the last two decades is a collection of findings of motor system involvement quite generally in perception and cognition. Accordingly, understanding the role or roles of the speech motor system in speech perception requires looking beyond the domain of speech to other domains in which findings of "embodiment" have been reported. We provide a sampling of these findings before offering some speculations of our own on the role(s) of the speech motor system in speech perception. Finally, we suggest directions for future research that may test those speculations and provide further understanding of the speech motor system and speech perception.

The Motor Theory of Speech Perception

The motor theory was first developed (e.g., Liberman, 1957) from behavioral findings (Liberman, Delattre, & Cooper, 1952; Liberman, Delattre, Cooper, & Gerstman, 1954) suggesting that listeners' identifications of consonants corresponded more closely to the vocal tract gestures of the speaker than to the consequent acoustic signals. In one finding (Liberman et al., 1952), when different consonantal articulations were found to produce the same acoustic "cue" because of coarticulation with different vowels, listeners heard different consonants corresponding to the different articulations. The second finding (Liberman et al., 1954) was complementary. In this case, two very different acoustic patterns that in natural speech would be caused by coarticulation of the same consonantal constriction-release gestures with different vowel gestures were heard as the same consonant. In both findings, "when articulation and the sound wave go their separate ways" (Liberman, 1957, p. 121), perception tracks articulation.

Liberman and colleagues developed a motor theory of speech perception to explain these striking findings. From the perspective of Liberman and Mattingly (1985), the motor theory is "in two ways motor" (p. 7). First is a claim consistent with the early findings described earlier and with many convergent findings obtained over the ensuing decades (see, e.g., Galantucci, Fowler, & Turvey, 2006, for a review). It is that listeners perceive speakers' phonetic gestures of the vocal tract. Ostensibly, the gestures have the invariance that the acoustic signal lacks. The second claim is that to achieve a motor percept, listeners recruit their own speech motor systems in perceptual processing.

A special perceptual system is required for speech perception in the view of Liberman and colleagues because, in contrast to sources of other acoustic signals, speech is coarticulated. Coarticulation gives rise to a very complex relation between the acoustic speech signal and the phonetic segments that provide the public substrate for linguistic communications. Because gestures for segments overlap temporally, there are no segment-sized units in the acoustic signal corresponding to phonetic segments (the "segmentation problem"). Correspondingly, there are no invariant acoustic patterns that have been found to be in 1:1 correspondence with phonetic segments (the "invariance problem"). Among acoustic signals having mechanical causes in the environment, for Liberman and colleagues, the segmentation and invariance problems arise only for speech perception, necessitating a brain specialization. The nature of the specialization proposed by Liberman and Mattingly (1985) is a phonetic module (cf. Fodor, 1983) in which the speech production system responsible for producing coarticulated speech (or, in Liberman and Mattingly's 1985 proposal, an "innate vocal tract synthesizer") is involved in decoding coarticulated speech.

In the theory, the nature of motor system involvement is a process of analysis by synthesis. An input acoustic signal is analyzed for its informative cues. The cues lead to a hypothesis about how the cues were produced. The hypothesis is tested by the synthesis component. Listeners' perceptions conform to the actions simulated by the synthesizer.

In the motor theory (e.g., Liberman & Whalen, 2000), there is another reason for a link between perception and production in speech; that reason is "parity." Essentially, there has to be a common phonological code for perception and production. This is both because language users are at once speakers and listeners who use the same phonology for talking and listening, and because, for between-person communication to succeed, talkers and listeners have to use a sufficiently common code. In the theory, the code is gestural.

A Caveat

Before providing an evidence-based assessment of the motor theory's second claim of motor system involvement in speech perception, we remark (cf. Fowler, 1996) that the first claim of the motor theory (that listeners perceive gestures) does not logically entail the second.

Moreover, Liberman and Mattingly (1985) notwithstanding, a theory claiming that listeners perceive speech gestures is not thereby a motor theory.

As noted, many convergent findings confirm that speech gestures are perceived. However, by some accounts (e.g., Fowler, 1986, 1996), listeners perceive phonetic gestures because that is what the acoustic signal provides information about, not because listeners recruit the speech motor system in perception. *Phonetic gestures* are linguistically significant actions of the vocal tract (e.g., Browman & Goldstein, 1992) that are the immediate causes of structure in the acoustic signal. Because different gestures cause distinctive structuring of the acoustic signal, the signal provides information for the gestures. In this sense, perceiving speech (and other causes of acoustic signals; e.g., Rosenblum, 2005) is like perceiving across the perceptual modalities. In the case of vision, for example, perceivers intercept informative patterning in reflected light that is informative about events in the environment because the events cause distinctive patternings in the light. Although visual perceivers intercept patterned light, that is not what they perceive. Rather, they perceive what in the environment the light patterning informs about. (For example, when a person walking causes distinctive patterning in reflected light that an observer intercepts, the observer sees the person walking, not the patterning in the light.) In both visual and auditory perception, including speech perception, perceivers perceive distal causes of the proximal signals that stimulate their sense organs. In the case of speech (perceived visually, auditorily, or haptically), the distal causes are phonetic gestures of the vocal tract.

Theories of visual perception agree that when motor action occurs in an observer's line of sight, the observer sees the motor action; this is analogous to the first claim of the motor theory of speech perception. However, the theories are not, thereby, deemed motor theories. To our knowledge, no theorist has made a claim analogous to the motor theory's second claim. For example, for a human to see another person walking, it is not considered necessary that the observer's own locomotor system be involved in perceptual processing, even though walking is coarticulated in the sense that multiple limbs move in overlapping timeframes. In the same way, to perceive talking auditorily does not logically require speech motor system assistance. We show shortly, however, that there is motor system involvement in speech perception (and there is motor system activation in perceiving walking; Takahashi, Kamibayashi, Nakajima, Akai, & Nakazawa, 2008).

Evidence for Motor Involvement in Speech Perception

In the following three sections of the chapter, we review evidence from studies of behavior and brain that there is motor system involvement in speech perception, but we also acknowledge some challenges to those findings. Following that, we interpret the findings of motor involvement in the context of the motor theory and of more recent proposals of motor system involvement in speech perception.

Behavioral Evidence for Motor Involvement in Speech Perception

Following is a sampling of behavioral evidence suggesting motor system recruitment in speech perception. Sams, Möttönen, and Sihvonen (2005) explored several variants of the well-known McGurk effect (e.g., McGurk & MacDonald, 1976) in which a video of a face speaking one syllable is dubbed with another appropriately selected acoustic syllable, and listeners report hearing a syllable that reflects integration of information obtained cross-modally. For example, a video of a speaker mouthing /ka/ dubbed with acoustic /pa/ characteristically leads to reports of /ka/ or /ta/. In one variant of this procedure, Sams et al. eliminated the video but had the listener him- or herself silently mouth /ka/ or /pa/ in synchrony with acoustic /pa/ presented in noise. The result was qualitatively the same as in the McGurk effect, although it was somewhat weaker.

Thus, silent simulation by the listeners of the articulatory actions for /k/ affected the consonant that they experienced hearing in a /k/-ward direction.

Experiments using articulatory perturbations provide converging evidence of speech motor influences on speech perception with the findings of Sams et al. (2005). A striking example is reported by Nasir and Ostry (2009). The researchers perturbed the jaw in the direction of protrusion as participants produced words that included the vowel /æ/. This perturbation had no measurable or audible acoustic consequences, but, even so, most participants compensated for it and adapted to the mechanic perturbation. Consequently, the jaw trajectory after compensation was closer to its pre-perturbation course than it was before adapting. A comparison of participants' perceptual classification of vowels along a *head* to *had* continuum before and after adaptation revealed a boundary shift. Fewer vowels were identified as /æ/ after adaptation. No shifts in identification were found for control participants who underwent the same procedure except that no perturbations were applied during the perturbation phase. Moreover, among participants in the experimental group, the size of the adaptive response to perturbation significantly correlated with the size of the perceptual shift. Clearly, therefore, the perceptual shift was tied to the adaptive motor learning. The finding shows that changes in the way that speakers produce a vowel also changes how they extract acoustic information for the vowel in speech perception—a finding mirrored in a very different experiment described next.

Ménard and Schwartz (2014) found individual differences in how speakers of French produced vowels differing in height. Some talkers distributed the vowels /i/, /e/, and /ɛ/ so that /i/ and /e/ were closer in F1 (and, therefore, in tongue height or tongue body constriction degree) than were /e/ and /ɛ/. Other talkers showed the opposite pattern. Remarkably, the grouping was maintained in perceptual studies. That is, those speakers who produced /i/ and /e/ with more similar constriction degrees than for /e/ and /ɛ/ also showed perceptual categories with the same F1 spacing. A similar finding was reported for English by Bell-Berti, Raphael, Pisoni, and Sawusch (1979). Like the finding of Nasir and Ostry (2009), these findings link listeners' own styles of talking to their extraction of articulatory information from speech acoustics. That is, listeners are not merely extracting information about the talker's articulations from the acoustic signal; their extraction of information appears to be biased by their own ways of talking.

Brain Evidence for Motor Involvement in Speech Perception

Here, we provide a sampling of neurophysiological evidence that reveals motor involvement in speech perception. In a study by Fadiga, Craighero, Buccino, and Rizzolatti (2002), participants listened to words and nonwords in which medial consonants were either a geminate /f/ (e.g., *baffo*)—a nonlingual (labiodental) consonant—or a geminate /r/ (e.g., *birra*)—a lingual consonant. During speech listening, the left motor cortex of participants experienced transcranial magnetic stimulation (TMS) in an area that stimulated tongue muscles, and the excitability of the muscles was assessed by motor-evoked potentials (MEPs) recorded from the tongue. The researchers found that TMS gave rise to greater MEPs as participants listened to words and nonwords having medial lingual rather than nonlingual consonants.

At the same time, other studies provide support for an influence in the complementary direction. Whereas the study by Fadiga et al. (2002) showed that listening to speech potentiates corresponding motor responses, these studies show that potentiating motor areas affects perception. Meister, Wilson, Deblieck, Wu, and Iacoboni (2007) used repetitive TMS (rTMS) to disrupt activity in the ventral premotor cortex temporarily. Participants' performance in a phoneme discrimination task decreased in accuracy after rTMS was applied compared with a baseline condition, but performance was unaffected in a control color discrimination task. This study implied that disruption of premotor cortex hinders speech perception. Moreover, D'Ausilio et al. (2009) applied (nonrepetitive) TMS either to the tongue or the lip motor area during a phoneme recognition task and measured the response time and accuracy of participants' classification of consonant sounds presented in noise (labial /b/ and /p/ or dental /d/ and /t/). Critically, the TMS pulses were delivered to the corresponding motor area just prior to auditory stimulus presentation in order to prime the area for subsequent phoneme perception. It was found that TMS of the tongue motor area led to faster and more accurate (relative to a no TMS baseline) classification of dental compared with labial consonants, whereas TMS of the lip motor area led to faster and more accurate classification of labial than dental consonants. Thus, stimulation of specific motor areas facilitates speech perception in a corresponding way.

Neuroimaging studies have shown that motor and premotor cortices can become active during acoustic speech perception. Wilson, Saygin, Sereno, and Iacoboni (2004) reported neural activation in the superior ventral premotor cortex during passive listening to meaningless monosyllabic speech sounds compared with rest, whereas a signal change in the same region was not evident when participants listened to a nonspeech sound (a bell). Wilson and Iacoboni (2006) examined neural responses in precentral motor areas during passive listening to native versus nonnative phonemes. The stimuli consisted of 25 English consonants and 25 consonants from a variety of languages other than English, all in an /aCa/ format. Although activation for speech perception was observed in the same area reported by Wilson et al., the activation was significantly smaller in size in response to native relative to nonnative phonemes, presumably because familiar consonants require less processing than less familiar consonants. The finding that motor areas were active was interpreted as reflecting the sensorimotor nature of speech perception.

Other research (Pulvermüller et al., 2006) shows that neural activation during speech perception is specific to consonant category in precentral gyrus. The researchers examined event-related functional magnetic resonance imaging (fMRI) during three kinds of tasks: passive listening to spoken syllables, including labial /p/ and lingual /t/ sounds; silent articulation of the same syllables; and nonverbal movement of the lips or tongue. Brain activation patterns indicated that the same precentral regions were activated by these three tasks if the same articulator was targeted, but there was differentiation of lip-related activation from tongue-related activation.

The foregoing studies presented speech stimuli acoustically only. In recent research, Skipper and colleagues (e.g., Skipper et al., 2005, 2006, 2007) have used audiovisual presentation and have found signatures of audiovisual integration. For example, given an audio /pa/ coupled with a video /ka/, participants reported hearing /ta/. Interestingly, the activation pattern in the frontal motor area with that audiovisual pairing correlated with the pattern produced by an audio /ta/–video /ta/ (AV /ta/) more than with patterns produced by either AV /pa/ or AV /ka/. In contrast, early activation in auditory and visual areas correlated respectively with activity evoked by an audiovisual /pa/ and an audiovisual /ka/ stimulus. Thus, only at the late stage of neural responses did activation patterns in these sensory areas become similar to those elicited by /ta/. The researchers interpreted this result as showing analysis by synthesis in audiovisual speech perception as reflected by cross-modal integration in the motor areas after modality-specific recognition in visual and auditory areas. That is, they concluded that “activity in areas of the motor system associated with speech production during observation of AV speech is a hypothesis about a particular (phonetic) interpretation of the stimulus properties rather than an accurate or veridical representation of stimulus properties” (Skipper et al., 2007, p. 2396).

Some Challenges

We have reviewed evidence for motor involvement in speech perception. However, that interpretation has been challenged. Menenti, Pickering, and Garrod (2012) noted that much current evidence on motor involvement during speech perception derives from artificial paradigms used in unnatural settings (something that can be said, however, about most speech perception research). For example, effects are most likely to appear under adverse listening conditions such as degraded speech (e.g., D'Ausilio, Bufalari, Salmas, & Fadiga, 2012) or isolated segments of speech presented out of context. Even researchers who argue for motor recruitment allude to certain prerequisites (e.g., lack of contextual support or challenging listening conditions) for motor effects to occur (e.g., Osnes, Hugdahl, & Specht, 2011). We note, however, that motor activation occurs under conditions that are not adverse—for example, in research by Fadiga et al. (2002) and in behavioral studies that we have described. Moreover, as D'Ausilio et al. (2012) acknowledged, noisy listening is typical outside of the laboratory; accordingly, motor activation in speech perception may be typical as well.

Another challenge, at least to Liberman's motor theory, occurs to findings that when motor involvement occurs, it is not necessarily special to speech stimuli. In a recent fMRI study, Agnew, McGettigan, and Scott (2011) compared cortical responses with speech sounds (voiceless consonants: plosives /t/ and /k/, a fricative /f/ and an affricate /tʃ/) and nonspeech mouth sounds (ingressive click) during passive listening. The investigators found no significant differences in those premotor areas previously identified as being involved in speech perception (Pulvermüller et al., 2006; Wilson & Iacoboni, 2006) in response to speech versus nonspeech mouth sounds. The researchers suggested that this lack of a speech-specific effect in motor and premotor areas reflects a generic sensitivity to mouth sounds in this region of the cortex. (We note, however, that the clicks are speech sounds in some languages.)

In addition, motor involvement is not found in all speech tasks or in all participants. As for speech tasks, Sato, Tremblay, and Gracco (2009) found that rTMS in the left ventral premotor area had an impact on a phoneme discrimination task but not on syllable discrimination or phoneme identification (but see Evans & Davis, 2015). Other studies have shown a lack of premotor activation in tasks that do not require explicit phonological judgment (e.g., Krieger-Redwood, Gaskell, Lindsay, & Jefferies, 2013). In a very different study, Menenti, Gierhan, Segaert, and Hagoort (2011) compared neural systems involved at various levels of language production and comprehension. These levels included semantic, syntactic, lexical, and *primary processing load* (i.e., number of syllables in sentences constructed for production or heard in comprehension—ostensibly, a measure of lower level sensory and motor processing). They found no differences in the neural systems involved in production and comprehension except at the lowest level. There, they found interactions between task and activation in motor areas, with motor involvement only in language production.

As for individual differences, Szenkovits, Peelle, Norris, and Davis (2012) found that changes in activation of motor regions in the brain (including the areas discovered by Wilson et al., 2004) during a pseudosyllable perception task relative to baseline reading and repetition tasks were predicted by individuals' verbal short-term memory. Individuals with poor verbal short-term memory were less likely to recruit motor regions in speech perception.

Taken together, these findings (among others) have led to skepticism about the generality of motor involvement in speech perception across tasks, testing conditions, and even individuals. These sources of skepticism may also apply to the domains outside of speech that we review shortly. For the present, however, we are impressed by the generality of motor involvement in a great variety of tasks, settings, and cognitive domains.

Do the Foregoing Findings Especially Support Liberman's Motor Theory?

Findings reviewed in the previous sections relate to the second claim of the motor theory of Liberman and colleagues (e.g., Liberman et al., 1967; Liberman & Mattingly, 1985; Liberman & Whalen, 2000) that there is motor system involvement in speech perception manifested as a process of analysis by synthesis that is special to speech perception. Both the behavioral and the brain studies of which we provide a limited sampling provide evidence that there can be motor system involvement in speech perception and that, when there is, the involvement shows both specificity and complementarity as predicted by the theory. As for specificity, for example, the study by D'Ausilio et al. (2009) showed faster and more accurate perceptual responses to labial consonants when TMS was applied to the lip region of the motor system and faster and more accurate responses to lingual consonants when TMS was applied to the tongue region. This is as expected if the role of the motor system in speech perception involves motor activation specific to the phonetic segments being perceived. Findings by Pulvermüller et al. (2006) are compatible in showing specificity. As for complementarity, on the one hand, the study by Fadiga et al. (2002) showed that perception of words or nonwords leads to muscle activation under TMS. On the other hand, D'Ausilio et al. (2009) and Meister et al. (2007) found that motor activation via TMS affects speech perception.

The findings do not show that motor system involvement is essential to speech perception[1] **<<Footnote 1>>** (cf. Lotto, Hickok, & Holt, 2009). Indeed, research findings in the last section suggest that it need not accompany speech perception in all situations. Additionally, evidence favoring a process of analysis by synthesis is weak. Although, as noted, Skipper et al.

(2007) interpreted some of their findings as favoring a process of analysis by synthesis, they restricted that process to audiovisual speech perception. We return to this issue when we review the motor theory of Skipper and colleagues next.

Overall, the findings that we reviewed neither support nor challenge the view of Liberman et al. that, in respect to the involvement of the motor system in speech perception, speech perception is special. Moreover, we reviewed findings of Agnew et al. (2011) that mouth sounds that were nonspeech sounds for their listeners (click sounds) showed motor activation like that to speech. We reinforce the conclusion that motor activation is not special to speech below in our review of evidence far outside of the domain of speech. That research shows evidence of motor activation quite generally across cognitive and perceptual domains.

Do the Foregoing Findings Support Other Motor Theories of Speech?

Skipper and colleagues (Skipper et al., 2006, 2007) proposed their own motor theory of speech perception on the basis of findings of motor activation during audiovisual speech perception, including those that we cited earlier and other findings. Having found more robust motor system activation in audiovisual than in auditory-only speech perception (Skipper et al., 2005), they suggested that perception of vocal tract gestures occurs when visual information is available. In this case—and especially when processing by the ventral ("sensory-semantic") neural route is discouraged by use of nonsense syllables in contrast to words or words in sentences—processing in a dorsal ("sensory-motor") processing stream dominates. In the theorists' view, either visual-gestural support or top-down lexical or other contextual support is required for speech perception because of the lack of acoustic invariants available for auditory speech perception.

In Skipper and colleagues' theory, when the dorsal route dominates, a coupled forward-inverse model instantiated in the mirror-neuron system (e.g., Rizzolatti et al., 1988) achieves speech perception by a process of analysis by synthesis. The perceptual information from the acoustic signal and visual input leads to hypothesized motor commands that would have given rise to the perceptual input. These commands are used in the forward, synthesis component to predict the sensory consequences that they should give rise to and to mediate linguistic interpretation of the input.

The theory differs from that of Liberman and colleagues most notably in restricting the gestural mode of perception to audiovisual speech perception. We are skeptical that this restriction is warranted. Except for findings by Skipper et al. (2007), none of the studies of motor system involvement in speech perception that we reviewed earlier were studies of audiovisual speech perception. Additionally, we later review evidence of very general motor system involvement in cognition. A restriction in the domain of speech only to audiovisual speech perception is unlikely given that generality.

With respect to both motor theories, on grounds of implausibility, we also reject the conclusion that analysis by synthesis is centrally involved in speech perception. First, analysis by synthesis should not be required to perceive speech gestures because the acoustic signal, having been structured by gestures, provides information about them. Second, analysis by synthesis is unlikely to permit speech perception in real time.

Although it remains true that invariants supporting direct perception of speech have not been found, theorists have, in general, significantly exaggerated the indeterminacy in the mapping between the acoustic signal and speech gestures. Iskarous (2010; see also Fowler & Iskarous, 2013) pointed out that some assumptions underlying claims of indeterminacy in the acoustics-articulation mapping are not true of the vocal tract—for example, that it is without energy loss. When Iskarous replaced that assumption with a more realistic assumption of energy losses of the right sort, he showed that vowel gestures are recoverable from acoustic signals. Further investigation has shown more general recovery of speech gestural dynamics from acoustics (Iskarous, Fowler, & Whalen, 2010; Iskarous, Nam, & Whalen, 2010).

Of course, this line of investigation does not rule out a process of analysis by synthesis, but it does suggest that it may not be required. More than this, analysis by synthesis may not be feasible. The supposed indeterminacy in the mapping from acoustics to gestures that has motivated the idea of analysis by synthesis implies that hypotheses about the gestural sources of the perceptual information tested by the synthesis component are subject to error. In turn, this means that when an error is detected, a new hypothesis has to be formulated and tested. Unless such errors are very rare, listeners would, it seems, fall behind talkers producing continuous, fluent speech (see also Schwartz et al., 2012). However, if errors are very rare, then "hypotheses" are not really hypothetical and need not be tested.

In short, we do not judge the motor theory of Skipper et al. (2006, 2007) to be better supported by research findings than that of Liberman and colleagues (Liberman et al., 1967; Liberman & Mattingly, 1985; Liberman & Whalen, 2000). Both are supported in part by evidence favoring motor system activation and involvement in speech perception. More specific claims are not supported convincingly, in our view.

In a recent article, Schwartz et al. (2012) proposed an alternative theory of speech perception in which the motor system plays an important role in the development of speech but perhaps not in online speech perception (cf. Hickok & Poeppel, 2007). Schwartz and colleagues have contributed to findings showing that speech perception is shaped by listeners' sensitivity to the gestural basis of acoustic speech signals (e.g., Ménard & Schwartz, 2014). However, they rejected the first claim of the motor theory that mature listeners perceive speech gestures. One reason for this conclusion is that listeners perceive English /ɹ/ as the same /ɹ/ even though research shows that it is produced in different variants by different speakers and even by the same speakers in different contexts. The acoustic signatures of the different variants of /ɹ/, however, are very similar. This implies to these, and other, theorists that objects of speech perception cannot be gestural. Moreover, Schwartz et al. pointed out that Liberman's motor theory does not predict the very general finding that segment inventories of languages, especially vowel inventories, are clearly shaped by perceptual considerations—for example, acoustic dispersion (e.g., Lindblom, 1986). In their view, evidence suggesting perception of gestures may, instead, reflect that perceptual representations of speech segments are structured by articulatory principles in the course of language development.

Schwartz et al. (2012) proposed the Perception-for-Action-Control Theory (PACT) in which perceptual representations and speech gestures are costructured in development as learners both produce speech-like actions and perceive them so that the perceptual representations are shaped by gestural principles and speech gestures are shaped by perceptual principles. For example, in the course of producing cyclic actions of the jaw (MacNeilage, 1998) in early vocalizing, infants hear quantal or categorical (e.g., Stevens, 1989) acoustic consequences of continuous articulatory opening–closing actions. During part of the jaw opening–closing range, vowel-like sounds are produced; however, at some point and for another (small) range, a shift to frication occurs, and finally, with full closure, stop-like sounds are produced. This fosters development of categories (vowels, fricatives, stops) that are shaped by perception of acoustic consequences of gestural action.

Schwartz et al. (2012) rejected any likely role of the motor system in online speech perception. They rejected analysis by synthesis on grounds similar to those we proposed earlier. They discounted evidence of the sort we summarized that revealed online recruitment of the motor system in perception as having very small effects and as failing to show that motor recruitment is required for speech perception.

Because our concern in the present chapter is not on the motor theory's first claim that gestures are perceived, we address the arguments of Schwartz et al. (2012) in this domain very briefly. As for the case of /ɹ/, we refer the reader to Fowler (2003) and comment that if two different sets of gestures produce acoustic signals that are imperceptibly different, gesture perceivers cannot perceive them as different. However, that does not rule out that they perceive gestures.

More relevant to current concerns, we are convinced by the findings we summarized in earlier sections of the chapter that there is online recruitment of the motor system in speech perception, although recruitment may not be general to tasks or individuals. Other studies to be reviewed next show pervasive recruitment of the motor system in perception and cognition generally. Unless speech perception is special, motor recruitment in speech is likely. We agree that such recruitment is not logically required for speech perception and that no finding has shown that it is necessary. However, necessary or not, it occurs, and it is valuable to consider why it does.

Embodiment in Perception and Cognition

Over the last decades, there has been a growing literature showing motor system recruitment in many perceptual and cognitive domains. In our view, these findings establish an important context for speculations about the role of motor recruitment in speech perception.

The Motor System and Language

Pulvermüller and Fadiga (2010) reviewed studies showing motor activation in language understanding. In particular, research shows activation of the premotor and motor cortices by words the meanings of which relate to human action. For example, the words *kick* and *grasp* lead to somatotopic activations; *kick* leads to activation in a foot region, whereas *grasp* leads to activation in a hand region. The activation occurs with a short enough latency (100–250 ms after presentation of a word in print) that the investigators inferred that activation is integral to word understanding, not a consequence of it.

Language understanding has effects on movement itself. For example, Glenberg and Kaschak (2002) compared latencies of judgments that sentences did or did not make sense in conditions that crossed direction of response movement and direction of movement implied by sentence meaning. Responses were faster, for example, to sentences such as *Andy delivered the pizza to you* if the response movement was toward the body rather than away from it. The latency pattern reversed if the sentence implied a direction of movement away from the body—for example, *You delivered the pizza to Andy.*

Finally, action experience modulates activation of motor regions of the brain during language understanding. Beilock, Lyons, Mattarella-Micke, Nusbaum, and Small (2008) found that experience playing or watching hockey led to recruitment of activity in the left dorsal premotor area of the brain during comprehension of sentences about individuals engaged in hockey-related activities (compared with sentences about everyday actions). Performance making judgments about the sentences was better with the more hockey-related experience participants had, and performance on the sentence task correlated positively with left dorsal premotor activity.

This is just a sampling of findings showing that language understanding is embodied (e.g., Fischer & Zwaan, 2008). Speech perception is not special or exceptional in this regard. However, language is not special either.

Embodiment Outside of Language

Motor recruitment during perceptual and cognitive activity is not confined to language. It occurs very broadly in a variety of perceptual and conceptual or cognitive domains. Following is a review of just a small number of the extant relevant studies. For a more comprehensive review, see Barsalou, Niedenthal, Barbey, and Ruppert (2003).

Earlier we commented that, in contrast to speculation in the speech domain, no one has suggested the need for a motor theory of the perception of other human actions—for example, walking. Even so, Takahashi et al. (2008) reported finding specific motor recruitment during the observation of walking. They used TMS to activate the tibialis anterior and soleus muscles of the legs of observers who watched actors either standing or walking on a treadmill. They found enhanced MEPs in those muscles compared with a baseline just when the confederates were walking but not standing. This finding is analogous to that of Fadiga et al. (2002) described earlier in which listeners under TMS applied to the tongue motor region exhibited enhanced MEPs in the tongue while listening to words or nonwords whose production involved the tongue more compared with those using the tongue less.

Somewhat like the study by Beilock et al. (2008) on language comprehension by individuals varying in hockey experience, in an fMRI study, Calvo-Merino, Glaser, Grèzes, Passingham, and Haggard (2005) compared professional dancers' brain activation patterns during passive watching of people performing actions in which the observers were skilled or not. These observers demonstrated greater activation in premotor cortex for the trained actions than for the untrained ones. In addition, the differentiated activation was not due to mere visual exposure to the actions but was related to actual motor learning experience (Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006). Thus, activities in one's motor repertoire are evoked by their perception.

Listening experience also affects activation of motor regions. Lahav, Saltzman, and Schlaug (2007) trained nonmusicians to play a novel piano music piece and then, using fMRI, examined their brain activation as they listened to the trained piece and two untrained pieces. One untrained piece had the same notes as in the trained piece but in transposed order; the other had different notes. There was much less activation in the premotor region in the latter two conditions compared with the trained condition. This suggests that there can be motor involvement during perception of nonspeech sounds that a listener has experience producing.

A nonspeech perceptual finding somewhat analogous to that of D'Ausilio et al. (2009) shows an effect of motor activation on perception of facial expressions. Blaesi and Wilson (2010) presented observers with photographs of a face in which the expression was morphed along a continuum from a smile to a frown. The task was to classify the expressions of faces presented in random order from the continuum into one of those two categories. On half of the trials, participants clenched a pen lengthwise in their teeth that forced them to adopt an expression like that of the photographed face at the smile end of the continuum. This manipulation had a large effect on classification judgments. More judgments that the face was smiling were made in the pen condition than in the comparison (no pen) condition.

Motor recruitment also occurs in problem solving or in thought processes more generally. Goldin-Meadow and Beilock (2010) reviewed evidence that the manual gesturing that accompanies language use not only reflects thought but also that it can influence thought—both of observers of those gestures and of the gesturer him- or herself. For example, Ping and Goldin-Meadow (2008) found that children instructed in solving conservation problems learn more from instruction that includes both gesture and speech than from instruction involving only speech. This result held whether or not the objects relevant to the conservation problem being discussed were present. Similar findings have been reported for instruction in other kinds of problems as well (e.g., Church, Ayman-Nolley, & Mahootian, 2004).

In another study (Beilock & Goldin-Meadow, 2010), gestures that solvers of the Tower of Hanoi problem[2] **<<Footnote 2>>** used in describing their solution had an impact on their performance when they went on to solve a variant of the problem. In particular, gestures relating to actions appropriate to the initial solution of the problem, but not to the solution of the variant, were associated with poorer performance on the variant.

The foregoing review provides a highly impoverished overview of the many findings of embodiment in perception and cognition, now extant in the literature. In short, motor activation appears to be a default concomitant of perception and cognition, and speculations as to its role in speech perception should be developed in the context of this array of findings.

Speculations on the Roles of Motor Involvement in Perception

Assessed in the context of the literature from which we selected the earlier examples, the proposal of Liberman and colleagues (e.g., Liberman et al., 1967; Liberman & Mattingly, 1985) that the speech motor system solves a problem special to speech perception—namely, that of decoding coarticulated speech—is implausible. The proposals we offer here as to other possible roles are made with very limited confidence. Research on speech perception that is informed by the larger context of findings of embodied cognition and perception remains to be done. We suggest that the pervasiveness of evidence of motor involvement in perception and cognition reflects, at least in part, the importance of varieties of coordination or coupling (within and between individuals) that are characteristic of many animals.

Beginning with the individual, we remark that in Gibson's (1966, 1979) ecological approach to perception, perception and action are always functionally coupled. Perception guides action, and exploratory action provides useful perceptual information. Warren (2006) commented that perceptual information serves as one source among others (including the structure of the environment and the biomechanics of the body) of constraints that underlie the emergence of adaptive action. Coupling between perception and action is essential to adaptive living.

As for language within the individual, we have noted—following Liberman and Whalen (2000)—that listeners are speakers, and the language/languages (e.g., English) that they use in both roles is/are generally the same. Findings that perceiving speech disposes a speaker toward imitative production (e.g., Fowler, Brown, Sabadini, & Weihing, 2003) and, complementarily, that changes or differences in modes of production of a segment may be correlated with differences in how listeners extract information for it in perception (Ménard & Schwartz, 2014; Nasir & Ostry, 2009) may constitute reflections of that within-individual coupling between modes of language use. (For a broader perspective on within-person, production-comprehension coupling in language, see Garrod & Pickering, 2004, and Pickering & Garrod, 2013.)

As important as within-person coordination is cross-person coupling. Humans are social organisms, and much of human life involves cross-person coordination in the service of achieving joint aims. Language use can serve a coordinating function (e.g., Clark, 1996). In the case of cross-person coordination, perceiving the actions of another provides constraints on one's own. Indeed, perceiving the actions of someone else can lead to unintentional coordination (e.g., Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007; Schmidt & O'Brien, 1997).

Compatibly, in language use, hearing the speech of an interlocutor fosters coordination that supports interpersonal conversation (e.g., Garrod & Pickering, 2004; Pickering & Garrod, 2009). Garrod and Pickering (2004) reviewed evidence that at multiple levels of language ("situation models," syntax, lexical items, phonetics), the speech of one interlocutor primes compatible speech by another. That is, interlocutors borrow each other's syntax, lexicon, and dialect as they engage in conversation. In Garrod and Pickering's view, this helps to explain why conversation is easy compared with producing a monologue, even though the former requires coordination across two physically separate cognitive systems.

Indeed, much of the research on motor involvement in perception, language, and cognition more generally has suggested that motor activation promotes imitative responses. Correspondingly, ideas about the role that embodiment serves have been stimulated by the idea that embodiment is fundamentally imitative. For example, Rizzolatti and Arbib (1998) suggested that embodiment fosters a kind of empathy or action understanding. Compatibly, Pickering and Garrod (2009; see also Pickering & Garrod, 2013) suggested that embodiment supports emulation, and emulation, in turn, fosters prediction or forecasting of the future behaviors of an interlocutor.

We are as skeptical of ideas that interlocutors construct models (Pickering & Garrod, 2009, 2013) for predicting behaviors of interlocutors in the same way that we are skeptical of analysis by synthesis as an important component of speech perception. We prefer to think that interlocutors' behaviors provide prospective information for their future actions that partners perceive and that lead to embodied responses.

However, it remains true that most embodied responses that have been reported in the literature are imitative in nature. If a major function of embodiment is to foster interpersonal coupling and coordination, and if most functional coordination does not involve imitating one's partner, why are findings of imitation so pervasive?

We note that findings of embodiment do not always involve imitation (e.g., Olmstead, Viswanathan, Aicher, & Fowler, 2009) and that "imitation" can be quite abstract. Olmstead et al. (2009) had participants engage in a coordinative task (bimanual pendulum swinging) that precluded their imitation of actions described in sentences to which they listened. The investigators found that sentences relating specifically to actions of the hand and arm (but not to other actions) had nonimitative effects on the motor task, suggesting a bodily reorganization to encompass the dual tasks of pendulum swinging and comprehension of arm/hand-related sentences. As for the abstract nature of some imitative movements, as described earlier, Glenberg and Kaschak (2002) found that response movements that paralleled movement implied in a sentence were faster than movements opposite in direction. For example, movements toward the body were faster to *Andy gave you the pizza* than movements away from the body, paralleling the direction of motion of the imaginary pizza. However, in nature, if there were an Andy with a pizza, and he gave a pizza to someone (*you* in the sentence), that someone would reach out in a direction opposite to the pizza's movement to take the pizza. Accordingly, in that finding, the movement somehow simulates that of Andy's arms or of the pizza, not the participant (*you*) mentioned in the sentence.

That aside, imitation does seem to be a default embodied response at least in speech. Although in hostile settings talkers may diverge in their speaking style from others (e.g., Babel, 2010; Bourhis & Giles, 1977; Labov, 1963), in most situations that have been studied (see, e.g., Babel, 2010; Giles, Coupland, & Coupland, 1991; for a review, see Pardo, 2006), convergence is the rule. Moreover, convergence occurs in settings that are almost wholly nonsocial (e.g., Goldinger, 1998; Shockley, Sabadini, & Fowler, 2004). So, again, it is appropriate to ask the following question: If a function of motor activation in perception and cognition is to foster coordination, why is the typical response imitative, when imitation is not always or even stereotypically an adaptive kind of coordination? Future research is required to answer this question. We make some suggestions in the following section.

Here, we provide some insights drawn from research on perception and action. Marmelat and Delignières (2012) recently demonstrated that interpersonal coordination can result from reciprocal anticipation on the basis of a global and multiscale, rather than local, coordination between systems. Dyads of participants synchronized oscillatory movements using handheld pendulums. Three conditions varied the amount of available information about the movements performed by the other member of the dyad and so were expected to vary the strength of local coupling within a pair. Findings revealed a high correlation between dyad members' fractal exponents (an index of movement complexity) independent of the strength of local temporal correlations between the two participants across the three conditions. This outcome demonstrates a distinction between local coordination and coupling on a higher level, suggesting that coordination is not restricted to short-term interpersonal adaptations. If similar global coordination occurs during speech conversation, then construction of internal models for making short-term predictions (e.g., Pickering & Garrod, 2013) may be obviated (e.g., Stephen & Dixon, 2011).

Moreover, global coupling will not necessarily manifest itself as imitation. Instead, joint action of coupled actors (e.g., in conversation) at the behavioral level can be understood and modeled according to its own dynamics (e.g., Schmidt, Fitzpatrick, Caron, & Mergeche, 2011). For example, interlocutors may diverge in a hostile setting and converge in some other settings.

Future Research on Motor Activation in Speech

Research That Expands the Contexts of Speaking

In the previous section, we asked why imitation has been a default outcome of motor activation in interpersonal speech settings. One possible answer is that in most experimental settings, not much is going on. The perceptual setting in most experiments does not provide much in the way of constraint on action. However, in more realistic settings in which perceptual information is present to guide action in some particular way (cf. Warren, 2006), motor activation will be for adaptive action whether it is imitative or not.

Indeed, this is suggested by research on phonetic convergence already described. Absent any particular set of constraints on mode of speaking, speakers converge phonetically with the speech they hear (e.g., Goldinger, 1998; Shockley et al., 2004). However, in hostile contexts (e.g., Babel, 2010; Bourhis & Giles, 1977), speakers may diverge. In different contexts still—for example, in conversation with someone with a strong foreign accent—a speaker may neither converge nor diverge; rather, he or she may produce hyperarticulated, slow speech to assist the listener to understand (Kangatharan, Uther, & Gobet, 2012). Accordingly, one line of research may be to expand the contexts in which embodied responses to perceived speech are observed to explore the range of embodied responses.

Put in the context of interpersonal interaction, future studies should explore the specific conditions under which behavioral coordination emerges and the contextual influences or features that mediate emergence of behavioral coupling. For example, it is possible that some a priori similarities between speaker and listener (e.g., similar speech rates) influence the strength or stability of coordination, and this may be reflected in the activation levels of motor regions. In relation to the research by Marmelat and Delignières (2012) described earlier, this should vary the global dynamics of the interpersonal coupling.

Research That Manipulates Motor Responding

Researchers have questioned whether motor involvement is essential to speech perception. We have suggested, in any case, that it is not essentially involved in decoding the speech signal. However, there may be ways to ask what work the motor system does do in speech perception. We have suggested that it may serve a role in intra- and interpersonal coupling in language use.

One avenue to pursue relating to intrapersonal coupling may be that under exploration by Nasir and Ostry (2009), described earlier. They showed that changes in the way that speakers produce a vowel have effects on their perception of the vowel. Moreover, the greater the adaptation is to a new way of producing the vowel, the greater the perceptual impact. However, it is not entirely clear why the perceptual effects were the effects they were. Why, when the adaptation was to a protrusion perturbation of the jaw during utterance of words containing /æ/ that had no measurable or audible acoustic effects, did the perceptual effects show that fewer /æ/s were identified along a continuum varying in vowel height? Further research along these lines may help to clarify the nature of the articulation–perception relation.

One way to look at interpersonal coupling is to suppress motor activation during the kinds of experience that have led to convergence in the literature. The prediction would be that convergence would be reduced or eliminated. For example, shadowing has been used to promote phonetic convergence in some studies (e.g., Babel, 2010; Goldinger, 1998; Shockley et al., 2004). If shadowing was accompanied by rTMS, or—perhaps more realistically—if shadowing was replaced by articulatory suppression (e.g., repetitive mouthing of a syllable, such as *da*), would convergence observed in a subsequent task be eliminated or reduced compared with convergence in the absence of articulatory suppression?

Other Research Lines

We have alluded to researchers, such as Skipper and colleagues (Skipper et al., 2005, 2006, 2007) or Hickok and Poeppel (2007), who, among others, have explored brain systems supporting speech perception and have drawn inferences from their research about the role of the motor system (or the lack thereof) in speech perception. This type of research is valuable and needs to be pursued. We are not experts in this domain, however, and do not suggest specifically how it should be pursued. We do have one general suggestion, though. In our view, research focusing on the brain needs to be accompanied by behavioral studies when questions relating to the function of motor involvement in speech perception are being addressed. The irreducible unit of investigation, in our view, should be the perceiver/actor perceiving and acting in the world, not the isolated brain.

Conclusions

The central message we have intended to convey by our review of motor system activation during speech perception and more generally during perception and cognition is that there is motor system involvement in speech perception and pervasively in other cognitive domains. However, understanding its role or roles in speech requires taking a broad perspective on motor system activation. If motor activation in speech perception has a function, its function is unlikely to be one that is specific to the perceptual task confronting the perceiver of speech. In this regard, we reject the proposals of Liberman and colleagues (e.g., Liberman & Mattingly, 1985) that relate motor system involvement in speech perception to address the problem of decoding coarticulated speech. Because motor system involvement is so pervasive in perception and cognition, its functions are more likely to reflect requirements that span perceptual and cognitive activities quite generally.

We have proposed that motor activation reflects pervasive requirements for within-person and between-person perceptuomotor coupling. As to the former, perceivers perceive in the service of acting, and actors engage in exploratory activity to increase the availability of perceptual information. Within the domain of speech, language users are both perceivers and producers of the same language(s) and so within-person coupling of the perceptual and motor speech systems is as important for speech as it is for other perceptual-motor activities. As to interpersonal coupling, humans are social organisms who engage stereotypically in coordinated joint actions with others (e.g., Marsh, Johnston, Richardson, & Schmidt, 2009; Marsh, Richardson, Baron, & Schmidt, 2006). Specifically, in the domain of speech, there is ample evidence of cross-person coupling of speech actions, as reviewed earlier. More abstractly, there is the parity requirement that languages users have a sufficiently common language code for language to serve its communicative function. Among other things, this means that the language produced by a talker be perceived as including the same language forms used by the listener.

Our suggestion that perceptuomotor coupling within and between language users is a central underlying source of motor activation in speech perception is, however, a shot mostly in the dark. Other researchers may have other insights to offer. However, we urge that research be designed with the understanding that motor activation in speech perception is likely to reflect an adaptation to challenges more general than those that specifically confront the perceiver of speech.

References

Agnew, Z. K., McGettigan, C., & Scott, S. K. (2011). Discriminating between auditory and motor cortical responses to speech and nonspeech mouth sounds. *Journal of Cognitive Neuroscience, 23,* 4038–4047.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39,* 437–456.

Barsalou, L. W., Niedenthal, P. M., Barbey, A., & Ruppert, J. (2003). Social embodiment. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 43–92). San Diego, CA: Academic Press.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science, 21,* 1605–1610.

Beilock, S. L., Lyons, I. M., Mattarella-Micke, A., Nusbaum, H. C., & Small, S. L. (2008). Sports experience changes the neural processing of action language. *Proceedings of the National Academy of Sciences, USA, 105,* 13269–13273.

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

Blaesi, S., & Wilson, M. (2010). The mirror reflects both ways: Action influences the perception of others. *Brain and Cognition, 72,* 306–309.

Bourhis, R., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity and intergroup relations* (pp. 119–135). London, England: Academic Press.

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

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An FMRI study with expert dancers. *Cerebral Cortex, 15,* 1243–1249.

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology, 16,* 1905–1910.

Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International Journal of Bilingual Education and Bilingualism, 7,* 303–319.

Clark, H. (1996). *Using language.* Cambridge, England: Cambridge University Press.

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex, 48,* 882–887.

D'Ausilio, A., Pulvermüller, F., Selmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology, 19,* 381–385.

Evans, S., & Davis, M. H. (2015). Hierarchical organization of auditory and motor representations in speech perception: Evidence from searchlight similarity analysis. *Cerebral Cortex, 25*, 4772-4788.

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.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology, 61,* 825–850.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology.* Cambridge, MA: MIT Press.

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. *The Journal of the Acoustical Society of America, 99,* 1730–1741.

Fowler, C. A. (2003). Speech production and perception. In A. Healy & R. Proctor (Eds.), *Handbook of psychology: Vol. 4. Experimental psychology* (pp. 237–266). New York, NY: Wiley.

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

Fowler, C. A., & Iskarous, K. (2013). Speech perception and production. In A. Healy & R. Proctor (Eds.), *Handbook of psychology* (2nd ed., Vol. 4, pp. 236–263). New York, NY: Wiley.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13,* 361–377.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8,* 8–11.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston, MA: Houghton Mifflin.

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

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

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9,* 558–565.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279.

Goldin-Meadow, S., & Beilock, S. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science, 5,* 664–674.

Hickok, G., Holt, L. L., & Lotto, A. J. (2009). Reply to Wilson: What does the motor cortex contribute to speech perception? *Trends in Cognitive Sciences, 13,* 330–331.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8,* 393–402.

Iskarous, K. (2010). Vowel constrictions are recoverable from formants. *Journal of Phonetics, 38,* 375–387.

Iskarous, K., Fowler, C. A., & Whalen, D. H. (2010). Locus equations are an acoustic expression of articulatory synergy. *The Journal of the Acoustical Society of America, 128,* 2021–2032.

Iskarous, K., Nam, H., & Whalen, D. H. (2010). Perception of articulatory dynamics from acoustic signatures. *The Journal of the Acoustical Society of America, 127,* 3717–3728.

Kangatharan, J., Uther, M., & Gobet, F. (2012). The effect of physical appearance and accent on the elicitation of vowel hyperarticulation by British English native speakers in speech to foreigners. *The Journal of the Acoustical Society of America, 132,* 2002.

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., & Jefferies, E. (2013). The selective role of premotor cortex in speech perception: A contribution to phoneme judgements but not speech comprehension. *Journal off Cognitive Neuroscience, 25*, 2179-2188.

Labov, W. (1963). The social motivation of a sound change. *Word, 19,* 273–309.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience, 27,* 308–314.

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

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

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 and Applied, 68,* 1–13.

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

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

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). Orlando, FL: Academic Press.

Lotto, A., Hickok, G., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences, 13,* 110–114.

MacNeilage, P. (1998). The frame-content theory of evolution of speech production. *Behavioral and Brain Sciences, 21,* 499–546.

Marmelat, V., & Delignières, D. (2012). Strong anticipation: Complexity matching in interpersonal behavior. *Experimental Brain Research, 222,* 137–148.

Marsh, K. L., Johnston, L., Richardson, M. J., & Schmidt, R. C. (2009). Toward a radically embodied, embedded social psychology. *European Journal of Social Psychology, 39,* 1217–1225.

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

McGurk, H., & MacDonald, J. (1976, December 23). Hearing lips and seeing voices. *Nature, 264,* 746–748.

Meister, I., Wilson, S., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology, 17,* 1692–1696.

Ménard, L., & Schwartz, J. (2014). Perceptuo-motor biases in the perceptual organization of the height feature in French vowels. *Acta Acustica united with Acustica, 100,* 676–689.

Menenti, L., Gierhan, S., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening as revealed by functional MRI. *Psychological Science, 22,* 1173–1182.

Menenti, L., Pickering, M. J., & Garrod, S. C. (2012). Toward a neural basis of interactive alignment in conversation. *Frontiers in Human Neuroscience, 6,* 1–9.

Nasir, S., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences, USA, 106,* 20470–20475.

Olmstead, A. J., Viswanathan, N., Aicher, K. A., & Fowler, C. A. (2009). Sentence comprehension affects the dyamics of bimanual coordination: Implications for embodied cognition. *Quarterly Journal of Experimental Psychology, 62,* 2409–2417.

Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage, 54,* 2437–2445.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America, 119,* 2382–2393.

Pickering, M. J., & Garrod, S. (2009). Prediction and embodiment in dialogue. *European Journal of Social Psychology, 39,* 1162–1168.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36,* 329–347.

Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology, 44,* 1277–1287.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience, 11,* 351–360.

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

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science, 26,* 867–891.

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.

Rosenblum, L. D. (2005). The primacy of multimodal speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 51–78). Malden, MA: Blackwell.

Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research, 23,* 429–435.

Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language, 111,* 1–7.

Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human Movement Science, 30,* 834–884.

Schmidt, R. C., & O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology, 9,* 189–206.

Schwartz, J., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics, 25,* 336–354.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing speech. *Perception & Psychophysics, 66,* 422–429.

Skipper, J., Nusbaum, H., & Small, S. (2005). Listening to talking faces: Motor cortical activation during speech perception. *NeuroImage, 25,* 76–89.

Skipper, J., Nusbaum, H., & Small, S. (2006). Lending a helping hand to hearing: Another motor theory of speech perception. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 250–285). New York, NY: Cambridge University Press.

Skipper, J., van Wassenhove, V., Nusbaum, H., & Small, S. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex, 17,* 2387–2399.

Stephen, D., & Dixon, J. A. (2011). Strong anticipation: Multifrctal cascade dynamics modulate scalingin synchronization behaviors. *Chaos, Solitons & Fractals, 44,* 160–168.

Stevens, K. (1989). On the quantal nature of speech. *Journal of Phonetics, 17,* 3–45.

Szenkovits, G., Peelle, J. E., Norris, D., & Davis, M. H. (2012). Individual differencs in premotor and motor recruitment during speech production. *Neuropsychologia, 50,* 1380–1392.

Takahashi, M., Kamibayashi, K., Nakajima, T., Akai, J., & Nakazawa, K. (2008). Changes in corticospinal excitability during observation of walking. *NeuroReport, 19,* 727–731.

Warren, W. (2006). The dynamics of perception and action. *Psychological Review, 113,* 358–389.

Wilson, S. (2009). Speech perception when the motor system is compromised. *Trends in Cognitive Sciences, 13,* 329–330.

Wilson, S., & Iacoboni, M. (2006). Responses to nonnative phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage, 33,* 316–345.

Wilson, S., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7,* 701–702.

<<Footnotes>>

[1]See Meister et al. (2007), already described, who interpreted their findings in this way. However, also see the debate that ensued after this finding (Hickok, Holt, & Lotto, 2009; Wilson, 2009).

[2]This is a problem involving four disks of different sizes that fit over pegs. There are three pegs, and the solver begins with the four disks stacked on the left-most peg in order of size, with the smallest disk on top. He or she has to move the disks one at a time so that, in the end, they are stacked in the same size order on the last peg. However, a constraint is that a larger disk must never sit above a smaller one.