

Motor Theories of Perception

Carol A. Fowler, Bruno Galantucci, and Elliot Saltzman

Introduction

Motor theories of perception propose that there is recruitment of the motor system or of motor competence (i.e., knowledge) in perception. Perhaps the best known motor theory of perception is Liberman's motor theory of speech perception (see Liberman, 1996, for a history and overview of the motor theory). Within speech science, despite its prominence, the theory has been judged implausible on several grounds. However, in the larger field encompassing studies of perception, action, and their coupling, it is given more credence. It is instructive to consider why the judgments differ between speech experts and experts in the broader domain.

In the following, we outline the motor theory of speech perception and describe some of the findings underlying its development. Next we offer reasons why speech scientists have doubted especially one of its two central claims, namely, that the speech motor system participates in speech perception. Then we suggest why the reasons are not sufficient to refute the claim, and we show that it acquires credibility when it is set in the larger context of investigations of perception, action, and their coupling. In addition, we summarize research that suggests a neural system consistent with Liberman's largely undeveloped ideas about neural support for speech perception. The discovery of mirror neurons in primates (Rizzolatti and Arbib, 1998) provides an existence proof of neuronal perceptuomotor couplings.

The Motor Theory of Speech Perception

Although in alphabetic script, consonants and vowels are discrete, their expression in acoustic speech signals is not. This is because speakers coarticulate speech gestures; that is, they produce the articulatory gestures of successive consonants and vowels in a temporally and spatially overlapping manner. Gestures are linguistically significant actions of the vocal tract. More specifically, they are equivalence classes of articulatory patterns controlled with respect to linguistically significant goals defined in an abstract task space. Consequences of coarticulation are evident, for example, during production of the word *to* when activation of the vowel /u/ 's lip protrusion gesture overlaps the activation of the consonant /t/ 's lingual gesture (compare *tea*). Due to coarticulation, acoustic speech signals are highly context sensitive, and they lack a discrete segmental structure.

Liberman developed a motor theory of speech perception when he and his colleagues found that speech percepts track articulation more closely than the acoustic signals to which articulation gives rise. Two experimental findings were especially telling. One was that, in the synthesized syllables /di/ and /du/, the critical acoustic cues for /d/ were quite different, owing to the effects of coarticulation by the different vowels. Indeed, the cues were audibly distinct when presented in isolation to listeners. However, the gestures for /d/ are the same in natural productions of the two syllables, and the consonants sound alike. A complementary finding was that the same acoustic cue was identified as /p/ before /i/ and /u/, but as /k/ before /a/. Because of coarticulation, to generate the cue before /i/ or /u/ requires production of /p/, whereas to generate it before /a/ requires production of /k/.

Both findings suggested to Liberman that when articulation and acoustic patterns diverge due to coarticulation, perception tracks articulation, a central claim of the motor theory. Subsequently, many other findings (see Liberman, 1996, for a review) converged

on the same conclusion. A notable one is the McGurk effect (McGurk and MacDonald, 1976), in which a video of a speaker mouthing one syllable, say, /da/, is dubbed with a different acoustic syllable, say, /ma/. Listeners hear a syllable (/na/ in the example) that reflects integration of gestural information from both modalities.

In Liberman's view, recovery of articulation in speech perception implies recruitment of the motor system. Such motor recruitment is required because of coarticulation in speech production. Speech information must be transmitted rapidly, and the gestural overlap provided by coarticulation permits efficient packaging of consonants and vowels. However, coarticulation has other consequences, including context sensitivity in acoustic information for phonetic segments. Therefore, two specializations, one for coarticulating and one for perceiving coarticulated speech, are needed, and, because neither specialization is useful without the other, they had to coevolve. Moreover, given the motor character of the percept, and Liberman's view that this reflects recruitment of the motor system in perception, the inference was plausible that the specializations were one and the same: a phonetic module. By using gestures as a common currency for talkers and listeners, the module helps guarantee achievement of parity between them—that is, sufficient equivalence between phonological messages sent and received, a necessity for successful communication.

Speech Science: The Implausibility of the Motor Theory

Following are grounds on which the motor theory of speech perception has been judged implausible, and then some reasons why we reject each argument.

1. Many speech scientists (e.g., Ohala, 1996) deny that speech percepts have a motor character, and they have no other reasons to suppose that the speech motor system is involved in perceiving speech.
2. Liberman and colleagues wrote very little about how the speech motor system might participate in speech perception, and the mechanism that they typically alluded to (analysis by synthesis) is not obviously workable at the rates at which consonants and vowels are perceived.
3. Listeners' perception of speech gestures need not imply that the speech motor system is recruited in speech perception. This is because the acoustic signal, having been caused by the gestures, and taking distinctive forms for distinct gestures, provides information about them. Listeners perceive gestures because that is what the information in acoustic speech signals is about.

We will address the first objection here only by remarking that, in our opinion as in Liberman's, evidence in favor of perceiving motor gestures is substantial and unrefuted. For example, we know of no studies that refute the evidence we cited in the previous section in favor of the claim that, when articulation and acoustic patterns diverge, perception tracks articulation. As for the other two objections, however accurate they may be, neither refutes the motor theory's claim of motor system recruitment in speech perception. As for the second objection, even if the particular mechanisms proposed by Liberman and colleagues are not the ones that support speech perception, it does not follow that no mechanism involving

a production-perception link does the job. As for the last objection, even though acoustic speech signals provide information about speech gestures, that does not preclude a perceptual mechanism in which the speech motor system or motor competence participates in decoding the acoustic signal.

The Broader Scientific Field: The Necessity for Motor Theories?

In the broader scientific field, central theoretical ideas of Liberman's motor theory recur (e.g., Viviani and Stucchi, 1992), and there are research findings suggesting motor involvement in perception. We review one example of a theoretical view that shares critical ideas with those of Liberman and then summarize a few of the research findings.

A Related Idea

Prinz (e.g., 1997) addresses an issue that arises in the study of perceptually guided action and that is very much like the one we have labeled parity. In speech, *parity* refers to the relation between messages sent and perceived. The messages must characteristically be the same; otherwise communication fails. Prinz has raised the same issue in asking how perception can guide action under the common assumption that percepts are representations of sensory information, and planned actions are coded in purely motor terms; that is, they lack a common currency. He proposes instead that percepts and actions share a common code. Further, consistent with the motor theory's identification of gestures as the common currency of talkers and listeners, and with the hypothesis that gestures are represented in the task spaces of talkers and listeners (SPEECH PRODUCTION), Prinz's *action effect principle* invokes a common code that represents not the proximal stimulus, but the relevant distal event properties. (*Proximal* refers to the signals that stimulate the sense organs, whereas *distal* refers to the environmental events that causally structure the proximal stimuli). Prinz's research shows, for example, that when stimuli that guide responses in some tasks share distal features with responses, response times are affected.

Research Findings

The larger context of evidence, to which we alluded earlier, in which the motor theory gains plausibility includes evidence from communication systems of other animals, evidence of motor recruitment in perception of motion, and findings of mirror neurons. In each domain, we provide illustrative examples.

Communication systems of other animals. Male crickets produce mating calls to attract females. Females respond to the calls by moving toward the male, but they do not produce calls themselves. However, males and females show a remarkable symmetry. Different varieties of crickets produce different calls, and females prefer the calls of their own type. When crickets are hybridized by mating the male of one type to the female of another, the male's call exhibits components from the calls of both parental types. Remarkably, female hybrids prefer the hybrid call to the call of either parental type (Hoy, Hahn, and Paul, 1977). This suggests a genetic correspondence between neural systems supporting call production in males and call perception in females.

Evidence of perception-action coupling can be found within individuals as well as between them. In zebra finches, the neural

system supporting call production also responds to components of auditorily presented songs (Williams and Nottebohm, 1985). A major path for song production in the zebra finch brain begins at a "higher vocal center" (HVC), which projects to the robustus archistriatum (RA) and from there to the tracheosyringeal portion of the hypoglossal nerve (nXIIIts). nXIIIts innervates the muscles of the syrinx. The HVC and nXIIIts both respond to tone bursts, and motor neurons in nXIIIts are differentially responsive to different components of perceived songs. Hauser (1996) concludes that "in order for birds to perceive the proper acoustic features of a song syllable, the percept must be converted into a series of motor actions required to produce the sound" (pp. 148-149).

Evidence in humans for motor recruitment in perception outside the speech domain. The tangential velocity of curved movements made by humans is proportional to curvature according to a two-thirds power law, decreasing with increases in curvature (e.g., Viviani and Stucchi, 1989). Viviani and Stucchi have shown that observers' judgments of the shapes of ellipses being drawn on a computer screen (judgments as to whether the major axis is oriented vertically or horizontally) are affected not only by the form's shape, but also by its velocity profile. When ellipses were drawn with constant velocity—a profile characteristic, in natural drawing, of a circular form—perceivers' judgments were poor. Tracings of ellipses that adhered to the two-thirds power law were judged accurately. An implicit proprioceptive-motor, rather than visual, task (Viviani, Baud-Bovy, and Redolfi, 1997) provided similar results. Blindfolded participants' right arms were moved in elliptical trajectories that did or did not preserve the two-thirds power law. With the left arm, participants tried to reproduce the movement of the right arm. Shapes of reproduced trajectories were more accurate when ellipses traced by the right arm conformed to the two-thirds power law than when they did not. Together, these data show that motor competence, here knowledge about velocity constraints on biological movements, is brought to bear on perception of motion.

Other evidence for linkages between the motor and perceptual systems comes from experiments that manipulate the similarity between a stimulus-response pair and measure its facilitatory or inhibitory influence on motor performance (see Prinz, 1997, for a review). For example, Stürmer, Aschersleben, and Prinz (2000) had participants produce a grasping gesture (first close the hand, then open it) or a spreading gesture (first open then close). The task-relevant stimuli for the movements were color changes on a hand that was displayed on a computer monitor, with different colors signaling each task. The visible hand also produced a task-irrelevant gesture on each trial, starting and ending from a neutral half-open position. In one case, it closed and then opened; in the other, it opened and then closed. Although participants were told to ignore the irrelevant information, selecting their responses only on the basis of the color change, their response latencies were faster when their movements matched the irrelevant ones. That perception of a hand gesture interacts with the execution of a similar or dissimilar hand gesture provides strong evidence that the perceptual and the motor systems share a common currency.

Mirror neurons. The foregoing evidence, like the evidence underlying the motor theory of speech perception, suggests access to the motor system or to motor knowledge in perception. Recent findings of mirror neurons may reveal part of a neural mechanism that permits and promotes such access.

Rizzolatti and colleagues (see Rizzolatti and Arbib's, 1998 review) have found neurons in the premotor cortex of the monkey (area F5) that respond both when the monkey performs a given action and when it perceives a similar action performed by another monkey or by a human. Many mirror neurons are quite specific in

firing during the performance of, say, one manual grasping movement but not another. Many of them exhibit the same specificity in the observed actions that stimulate them to fire.

There is evidence for mirror neurons in humans. Fadiga et al. (1995) used transcranial magnetic stimulation (TMS) of the motor area, in which stimulation provoked muscle activity in the fingers. During TMS, participants observed several events or situations: someone grasping an object, the stationary object itself, someone tracing shapes in the air with the arm, or the dimming of a light. The investigators found more TMS-induced muscle activity in the fingers when participants were observing grasping than when they were observing any other of the events. The modulation of muscle activity was specific to the actions observed. Fadiga et al. concluded that "in humans there is a neural system matching action observation and action execution" (p. 2609).

The finding of mirror neurons reveals neural systems that underlie perception-action coupling in monkeys and perhaps in humans as well. From the perspective of the motor theory of speech perception, it is intriguing that the neurons were found in an area of the monkey brain that includes the homologue of Broca's area in humans, which is involved in language use. The findings, therefore, lend credence to the motor theory's claim of a production-perception coupling in speech.

Discussion

The motor theory of speech perception has inspired analogous theories in other domains. Yet the theory was motivated by requirements of speaking and listening that Liberman considered special to speech. Is speech special in respects that should discourage efforts to generalize some of its proposals to other domains? We suspect not.

As we have noted, talkers and listeners must characteristically achieve parity to communicate successfully (Liberman, 1996), and parity achievement requires use of a common currency by talkers and listeners. That the speech percept has a motor character suggests that this common currency is defined in gestural task space: listeners *detect* the proximal acoustic signal, but they *perceive* the distal gestural activities of talkers. According to the motor theory, gesture perception is fostered by motor recruitment in perception.

This is not very different from what is required for successful nonlinguistic transactions with the environment, including those with other actors. Although proximal energy patterns stimulate the sense organs, animals must perceive the distal possibilities afforded for action (e.g., Gibson, 1979; see also GRASPING MOVEMENTS: VISUOMOTOR TRANSFORMATIONS). For actions to be felicitous, parity is required among perceived possibilities for action, real possibilities for action, and action itself. This is real-world, functional perception-action coupling. Plausibly, neural-motor recruitment in

perception fosters achievement of these parities as well as those of linguistic communication.

In much cognitive science research, perception and action are assumed to be sufficiently distinct and autonomous that they can be studied independently. However, consideration of the relations between animals and their environments uncovers no principled way to draw such a sharp distinction. Perception-action couplings are central to the design of animals. Understanding the real-world settings in which cognitive activity occurs reveals that it could not be otherwise.

Road Map: Linguistics and Speech Processing

Related Reading: Language Evolution: The Mirror System Hypothesis; Language Processing; Optimality Theory in Linguistics; Speech Production

References

- Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G., 1995, Motor facilitation during action observation: A magnetic stimulation study, *J. Neurophysiol.*, 7:2608–2611.
- Gibson, J. J., 1979, *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin. ♦
- Hauser, M., 1996, *The Evolution of Communication*, Cambridge, MA: MIT Press. ♦
- Hoy, R., Hahn, J., and Paul, R., 1977, Hybrid cricket auditory behavior: Evidence for genetic coupling in animal communication, *Science*, 195:82–84.
- Liberman, A. M., 1996, *Speech: A Special Code*, Cambridge, MA: MIT Press. ♦
- McGurk, H., and MacDonald, J., 1976, Hearing lips and seeing voices, *Nature*, 264:746–748.
- Ohala, J., 1996, Listeners hear sounds not tongues, *J. Acoust. Soc. Am.*, 99:1718–1728.
- Prinz, W., 1997, Perception and action planning, *Eur. J. Cogn. Psychol.*, 9:129–154. ♦
- Rizzolatti, G., and Arbib, M., 1998, Language in our grasp, *Trends Neurosci.*, 21:188–194. ♦
- Stürmer, B., Aschersleben, G., and Prinz, W., 2000, Correspondence effect with manual gestures and postures: A study of imitation, *J. Exp. Psychol. Hum. Percept. Perform.*, 26:1746–1759.
- Viviani, P., Baud-Bovy, G., and Redolfi, M., 1997, Perceiving and tracking kinesthetic stimuli: Further evidence of motor-perceptual interactions, *J. Exp. Psychol. Hum. Percept. Perform.*, 23:1232–1252.
- Viviani, P., and Stucchi, N., 1989, The effect of movement velocity on form perception: Geometrical illusions in dynamic display, *Percept. Psychophys.*, 46:266–274.
- Viviani, P. and Stucchi, N., 1992, Motor-perceptual interactions, in *Tutorials in Motor Behavior II* (G. Stelmach and J. Requin, Eds.), Amsterdam: North Holland.
- Williams, H., and Nottebohm, F., 1985, Auditory responses in avian vocal motor neurons: A motor theory for song perception in birds, *Science*, 229:279–282.