

In Defense of the Motor Theory

Ignatius G. Mattingly

University of Connecticut, Storrs, CT, USA
Haskins Laboratories, New Haven, CT, USA

Abstract

MacNeilage (1991) criticizes the Motor Theory of Speech Perception as untestable and in conflict with what is known about language acquisition. To this it is responded that the theory indeed needs a more specific definition of a phonetic gesture, and proposals for this definition are made. However, the theory is certainly testable, and has in fact been tested. A genetically determined capacity for phonetic gestures is not improbable if the gestures are seen as constituting a system. The fact that it is some time before phonetic gestures are properly executed by a child is not inconsistent with such a capacity, when the asymmetry between production and perception are taken account of. On the contrary, the evidence suggests that phonetic capacity unfolds according to a biologically predetermined schedule.

In 1985, Alvin Liberman and I published "The motor theory of speech perception revised," in which we claimed, first, that speech perception and production were the business of a biologically distinct system -- the language module -- and, second, that the units of of this system were phonetic gestures. Since then, a number of objections to both our proposals have been advanced. Here, however, I consider only some objections to the second of our proposals, raised by Peter MacNeilage (1991)¹.

The criticism from MacNeilage that I take most seriously is his assertion that "the concept of gesture is not yet adequately defined." (p. 61). He continues:

The lack of explication by the authors of the claim that there are invariant underlying gestural control signals makes the theory untestable . . . This is particularly serious, because Liberman and Mattingly themselves point out that they see no straightforward relation between the underlying level and the level of programmed contextual variation. (p. 63).

This is surely very hard. One might have thought that the various experiments that we have adduced to show that the objects of perception were not acoustic but articulatory were tests of the Motor Theory. In each of these experiments there was, to borrow MacNeilage's words, "a choice between gestural theories of perception and more orthodox theories based on conceptions of auditory function." (p. 66).

Nevertheless, it would certainly strengthen the Motor Theory to be more specific about gestures. In our 1985 paper, we said that a gesture was a distinctive deformation imposed on the current vocal tract configuration, that instances of the same gesture have invariant topological properties, and that these properties correspond to the speaker's intentions. I still think these

¹My responses to the objections raised to our first proposal by Kluender, Diehl, & Killeen (1987), Lindblom (1991), and Fowler & Rosenblum (1991), were presented in the oral version of this paper, but are not included here for lack of space, and will be published elsewhere.

Mattingly

statements are correct, but some elaboration is called for.

The gestures must be necessarily be defined in a linguistic context. To say this does not dispose of the problem of specifying the set of gestures; there are numerous phonological issues. But I don't see why one would hope, as MacNeilage does, to define the gestures "independently of what we already know from linguistic analysis." (p.64). Surely, our object should be to account for the intuitions of speakers and listeners about their productions and perceptions that linguistic analysis formalizes. If we can manage this, we shall be doing very well.

A gesture, as we Motor Theorists use the term, is an event that happens to the vocal tract. Cavities of the tract are shortened or lengthened, widened or narrowed, closed and then opened. The aerodynamic state as well as the configuration of the tract may be affected. The gesture is a "coordinative structure," in Turvey's (1977) sense, in general accomplished by the movements of more than one articulator. These movements are not ends in themselves, but rather the means of bringing about the vocal tract event (For an elaboration of this point, see Mattingly, 1990).

Instances of a particular gesture may differ in two ways. First, the contribution of a particular articulator to the gesture may be variable. This may come about because gestures overlap, and a particular articulator may be required to contribute to some other gesture at the same time. But the effect on the vocal tract of the gesture in question remains the same, because some other articulator compensates for the first. Thus, the jaw and the lower lip may make varying and complementary contributions to the achievement of the same lip height.

The more interesting sort of variation, however, is when instances of a gesture differ spatially or temporally in their effect on the vocal tract. Thus, bilabial stops can be made with considerable variation in lip height and degree of protrusion at closure. But they have in common the property of closure at the lips, wherever the lips happen to be. It is examples of this sort that led us to use the term "topological" to refer to the properties of gestures (Lieberman & Mattingly, 1985). Of course this is somewhat metaphorical; what we were trying to say was that despite obvious within-category variation, categories such as bilabial stop and alveolar stop have invariant phonetic properties that distinguish them from one another, just as spheres and toruses have invariant geometrical properties that distinguish them.

We are not saying that the listener recovers actual articulator trajectories; to what extent this may be true is an open question. If we were, the lack of a theory accounting for the relation between the intended gestures and the observed trajectories would indeed be a problem. We are saying, rather, that it is these topological properties of the effects of a gesture on the vocal tract that are "intended" by the speaker and recovered by the listener. These properties, or features, we take to be the basic invariants of speech perception (Cf. Halle & Stevens, 1990). They may not necessarily be patent in limited physiological data. But they are hardly abstract and do not render the Motor Theory untestable.

Mattingly

It has also been argued that the Motor Theory conflicts with what is known about infant language acquisition. In the 1985 paper, we had said:

Perhaps, then, the sensitivity of infants to the acoustic consequences of linguistic gestures includes all those gestures that could be phonetically significant in any language, acquisition of one's native language being a process of losing sensitivity to gestures it does not use. (Lieberman & Mattingly, 1985, p. 24).

This proposal has been criticized on several grounds by MacNeilage. He says, first:

The problem is that the neonate is given a considerable capacity to do not only things that any one speaker is never called on to do, but also things that the ancestors of any one speaker were presumably never called on to do. That is, it includes a capacity to perceive and produce a large family of specific nonuniversal sounds. The question is, therefore, how could such capacities have been selected for, when no single line of descent ever exhibited them all? (p. 63).

But, contrary to what this comment implies, we do not propose that each of the various gestures found in the world's languages has a separate evolutionary history. What we propose is a genetically founded capacity to perceive and produce a particular class of gestures -- phonetic ones. These gestures are far from being a miscellany of different articulatory events; they form a system, as is obvious from the IPA chart. Just how this system evolved is unknown; but clearly, the sort of question we should ask of evolution is not: "How did the capacity to perceive and produce, e.g., [b], evolve?" but rather, "how did the capacity to perceive and produce stops evolve?" or "How did the capacity to produce bilabials evolve?"

MacNeilage continues:

If we indeed possess such an efficient set of presumably equipotential capacities, why does it take the typical child 3.5 years to successfully produce the sounds of the language, and why are most . . . universal sounds mastered before nonuniversal ones? In addition, why does perceptual ability seem to develop so much more readily than production ability in infants? (. p 64).

These questions are well-taken, and suggest that our account of the psychological status of phonetic gestures needs to be amplified in certain respects. It is necessary to distinguish the role of the linguistic module from the roles of the motor subsystems that control the various moveable organs, each of which has one or more nonphonetic functions in addition to its phonetic function, and to call attention to an asymmetry between speech production and speech perception. As I have just argued, each phonetic gesture in the universal inventory is defined by certain anatomical, dynamic and aerodynamic properties. In the case of the listener, the language module detects these properties, specifying the gesture perceptually. In the case of the speaker, the module requires that the articulators move so as to instantiate these properties, and so implement the gesture. But what the module probably does *not* do is to specify the actual muscle commands that will meet these criteria. It plans the utterance, and issues a series of demands that result in the execution of the gestures. The execution itself, however, is accomplished by the organ-specific motor subsystems, which do the best they can to satisfy the demands of the module, as well as demands from other quarters that may be even more imperative.

Mattingly

Various reasons for this arrangement suggest themselves. On the one hand, only the language module is in a position to organize the activities of the several articulators; there is no other single system available that can coordinate the behavior of organs as disparate as the lips, the larynx and the velum. On the other hand, while the module knows about the architecture and mechanics of vocal tracts in general, it probably does not know about the specific, changing dimensionalities of the articulators and fixed structures of the particular speaker's tract. Moreover, it would be maladaptive for the module to have complete control of the vocal tract during speech, to the exclusion of other reflex systems, such as breathing, swallowing and gagging. Thus, there is competition for the services of the articulators that can be resolved only by the organ-specific control systems, and the language module does not have the highest priority.

In the case of the listener, on the other hand, no comparable problem arises. There are no organs to be moved. There is competition between the language module and other auditory modules for auditory information, but the language module always wins the competition. As a consequence, there is an asymmetry between production and perception.

On this view, it is not surprising that while speech perception begins very early, it takes time to become a fluent speaker. The inventory of gestures may be genetically specified, but the actual execution, by the speaker's own articulators, of the gestures used in the language being acquired must still be practiced to become automatic. (Something like this must be true even in the seemingly much simpler case of walking, which though patently genetically determined but, in contrast to speech, requiring no input from other walkers, does not emerge instantaneously).

In this respect, speech production resembles other skilled behavior. In fact, considering the complex articulatory patterns observed in speech production, one might well ask: Why does it not take a lot longer to learn to talk? and: Why do all normal children do a pretty good job of it?

We can arrive at a possible answer by considering an implication of MacNeilage's question, quoted earlier: "Why are most . . . universal sounds mastered before nonuniversal ones?" The implication is that the all infants acquire the sounds of their native language in more or less the same order. Yet one might have expected, on the basis of what has been said so far (and even more so on an ordinary behaviorist account), that the order in which gestures were acquired by a particular infant would be heavily influenced not only by factors that might indeed favor a constant order, such as the amount of effort required to produce the gesture or its relative perceptual salience, but also by factors that would be likely to result in variability in the order, such as the relative frequency of the gestures in the particular infant's linguistic input. It would also be probable, given the variability of this input, that the infant would make unwise decisions about which sounds to work on first, wasting a lot of time on difficult sounds because it had overestimated their frequency. The infant might even fail to practice other sounds at all. If what we actually observe, on the other hand, is a definite regularity in order of

Mattingly

acquisition of sounds, independent of particular languages and particular speakers, this would seem to be very strong evidence that the language module is firmly guiding the infant's phonetic development, just as it is assumed to guide other aspects of its linguistic development. What gets worked on when is not determined by the whims of the child, but is scheduled by the module in a way that has proved to be adaptive for languages generally. Evidently, language acquisition is too important to be left to children. As for MacNeilage's question itself, universal sounds can be worked on before non-universal ones because their presence in the phonetic inventory can be assumed; no need to wait for a phonetic analysis of the linguistic input.

This view of the relation between the module and the motor systems that control the articulators also suggests an interpretation of the patient described by MacNeilage, Rootes and Chase (1967). This patient had severe congenital impairment of somesthetic perception and articulatory control. She could not organize the movements of her tongue or lips and had severe deficits in speech production. Yet she was able to understand speech and perceived it categorically. Contrary to MacNeilage's assertion, the perceptual abilities of this patient pose no problem for the Revised Motor Theory. The Motor Theorist would say that although her somesthetic system was disabled, and her motor-control systems therefore functioned poorly, her language module, and so her ability to perceive phonetic gestures, were intact. Of course, MacNeilage's interpretation of this case is that speech perception does not deal in gestures at all.

I believe that those who find it implausible that the units of speech perception should be motoric, fail to appreciate fully the significance of the fact that speech is part of a biological communication system in which the same individuals are senders and receivers. Speech production and speech perception evolved together, as the output and input sides of the system. It makes little sense to consider them apart from each other, as if they were simply special cases of motoric behavior and perceptual behavior. Certainly, we need to understand what speech production owes to more general motoric processes, and what speech perception owes to more general perceptual processes. But the central question has to be, how do speech perception and production fit together?

REFERENCES

- Fowler, C. A. & Rosenblum, L. D. The perception of phonetic gestures. In Mattingly, I. G. & Studdert-Kennedy, M., (eds.), *Modularity and the Motor Theory of Speech Perception* (pp. 33-59). Hillsdale, NJ, USA: Lawrence Erlbaum.
- Halle, M. & Stevens, K. N. (1990). Knowledge of language and the sounds of speech. Paper presented at the Symposium on Music, Language, Speech, and Brain, Stockholm, September, 1990.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237, 1195-1197.
- Liberman, A. M., & Mattingly, I. G. (1985). *The Motor Theory of Speech Perception revised*.

Mattingly

Cognition, 21, 1-36.

- Lindblom, B. (1991). The status of phonetic gestures. In Mattingly, I. G. & Studdert-Kennedy, M., (eds.), *Modularity and the Motor Theory of Speech Perception* (pp. 7-24). Hillsdale, NJ, USA: Lawrence Erlbaum.
- MacNeilage, P. F. (1991). Comment: The gesture as a unit in speech perception theories. In Mattingly, I. G. & Studdert-Kennedy, M., (eds.), *Modularity and the Motor Theory of Speech Perception* (pp. 61-67). Hillsdale, NJ, USA: Lawrence Erlbaum.
- MacNeilage, P. F., Rootes, T. P., & Chase, R.A. (1967). Speech production and perception in a patient with severe impairment of somasthetic perception and motor control. *Journal of Speech and Hearing Research*, 7, 209-232.
- Mattingly, I. G. (1990). The global character of phonetic gestures. *Journal of Phonetics*, 18, 445-452.
- Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In Shaw, R. & Bransford, J. (eds.) *Perceiving, acting, and knowing* (pp. 211-265). Hillsdale, NJ, USA: Lawrence Erlbaum.