What a perception–production link does for language

Alvin M. Liberman 426
Haskins Laboratories, New Haven, Conn. 06510

Among the results obtained by Ojemann and his colleagues are some that imply a link, especially in the phonetic domain, between perception and production. It may be appropriate, therefore, to comment on the function that such a link might serve. I will speak of two. One pertains to the "immediacy" of phonetic perception, the other to the plasticity that underlies the development of dialect.

It is characteristic of specialized perceptual modules that they yield the distal object without conscious mediation. Thus, in perceiving depth through binocular disparity, the viewer is aware of depth as such, not of the displacements in the proximal stimulus that provide the relevant information. Similarly, in perceiving where a sound comes from, the listener is immediately aware of the distal location, not of the proximal diotic differences in time of arrival and intensity. Both perceptual results are accomplished, presumably, by modules specialized to take account of the relation between distal object and proximal stimulus. I have argued elsewhere that it is just so with phonetic perception, but with an important difference (Liberman 1982). For in phonetic perception the distal object is not, as in the otherwise analogous cases, an external thing; it is rather a process in the talker's brain that represents the phonetic structure he means to communicate. Hence, the peculiarities of the relation between distal phonetic object and proximal sound are governed by just those neuromuscular activities, including especially coarticulation, that occur as the speaker converts the one to the other. If perceiving is to take account of that relation, and since the listener is potentially a speaker, one might suppose (details of the mechanism aside) that a link between perception and production is the way it is managed. The important consequence is that the distal phonetic object can be perceived, as it normally is, without conscious translation from the auditory percepts that the proximal sounds would otherwise produce. Thus, perception of the [d] in [da], for example, is not mediated by awareness of the numerous pitch glides or glissandi that characterize the formant-transition cues when they are perceived, outside the phonetic mode, as acoustic objects (Mann & Liberman, in press). Nor is there in the perceived contrast between [jus] and [juz], for example, awareness that the relevant information for the final segments is, in fact, spread through the entire syllable and overlapped thoroughly with information for the segments that occupy the first and second positions (Soli 1982).

The point of the foregoing argument is that, given a link between perception and production, the listener's awareness of speech is appropriately phonetic, not auditory. Perhaps the most direct way to appreciate what that means, and at the same time to see how it might depend on a link between perception and production, is to experience a phenomenon recently discovered by McGurk and MacDonald (1976), a phenomenon that is, in its way, as striking as the facts developed by Ojemann. An example, taken from my own modification of their procedure, is as follows. I watch a televised face as it articulates [ði], but I am presented with a clearly spoken [ba]. What I perceive is [ða]. The effect is quite compelling, so compelling indeed as to render me incapable of discovering by introspection that the information for the perceived consonant was primarily optical and for the vowel acoustic. Presumably it is the link between perception and production that permits the visual information about articulation and the auditory information about the sound to converge on a unitary phonetic percept appropriate to their common source.

As for the role of a perception–production link in the development of dialect, we should note first that all languages do, in fact, change, though not in the same way. They are not falling away from some ancient ideal (Hebrew, Sanskrit, Greek?), nor are they converging, as if by natural selection, toward some still more efficient future form. They are simply changing. The important result is that, though all human beings talk, they do not all talk in the same way. Language serves thus as a shibboleth – indeed, as *the* shibboleth in the biblical account – marking each of us as a member of this or that group. The underlying tendency of language to change is not just a result of accumulated error or conscious invention, I think, but must rather rest on special provisions that are fundamental to language and its biology. One of those provisions, and the one most appropriate to this discussion, is the ability (and propensity) of young human beings to acquire speech by imitation and thus to preserve such language changes as occur. One might argue, I suppose, that such imitation requires no special equipment, only a general ability to learn by trial and error. But given what is known about phonetics, and what can be inferred about the afferent–efferent translation that must occur in phonetic mimicry, one might rather suppose that a biologically given link between perception and production is a necessary condition. In that connection, it is relevant that vocal learning is not widespread among animals, but occurs only in us and in certain birds (Marler 1976). It is even more relevant that in just those birds there is a specialized piece of neuroanatomy known to be associated with vocal learning, and that one of its characteristics is a rich set of interconnections of auditory afferent and vocal-controlling efferent fibers – that is, a perception–production link (Nottebohm, Stokes & Leonard 1976; McCasland & Konishi 1981).