4

Modularity and the Effects of Experience

Alvin M. Liberman
Haskins Laboratories

Ignatius G. Mattingly
Haskins Laboratories and University of Connecticut

Experience is essential to the development of speech in the child, so we should ask of any psychological theory of speech how the effects of experience are to be accommodated. What does the theory have to say about how the child learns which phonetic distinctions are relevant in his native language and about how he adjusts to the phonetic capabilities of his own particular vocal tract as it changes in size and shape? As we show in this chapter, the more and less conventional views of speech account for such effects of experience in categorically different ways. Our aim is not to weigh these different accounts against the evidence, but only to use them to illuminate an important difference between the theories from which they follow.

SOUND LOCALIZATION AND EXPERIENCE

Before taking up the theories about speech, we should be as explicit as we can about two different ways in which experience might affect percepts and their relations to everything else. For that purpose, we begin with the example of sound localization. We do this not because sound localization is simple, for it is, in fact, marvelously complex, but because it is well understood and lends itself readily to displaying the theoretical choices with which we are concerned. Consider, then, two conceivable accounts of sound localization, together with their different implications for the ways in which experience might affect development.

By any account, the information for localization takes the form of interaural disparities of time of arrival and intensity (Hafter, 1984). Complicating the processes by which such information is used is the fact that the disparities vary, not only as a function of source location, but also of frequency, aural acuity, distance between the ears, and orientation of the head. Further complications arise, of course, when several sources are present at once.

On one theory, the currently prevailing one, sound localization is managed by a neural mechanism narrowly and exclusively adapted to cope with the attendant complications and derive location of a source of sound from the information about disparity. We will refer to this mechanism as a module, using the term in the sense of Fodor (1983). That such a sound-localizing module exists has been shown most plainly in experiments on the barn owl (Knudsen, 1984; Konishi, Takahashi, Wagner, Sullivan, & Carr, 1988). In these experiments, investigators have found cells in the inferior colliculus that respond selectively to sounds according to their location in space, and thus form a map. Moreover, it has been possible to observe some of the processes by which this map is derived. But what is of particular importance for our purposes is that, though the information critical for location takes the form of disparities of time and intensity, with appropriate corrections for frequency, source coherence, and phase ambiguities, the positions of the neural responses are fixed by coordinates that are purely spatial; the proximal stimulus dimension of disparity is not represented. Information about this dimension is therefore not available outside the localization module, and the owl presumably does not perceive it. All this being so, we assume that the spatially organized neural responses correspond to the perceptual primitives.

But it is conceivable that, contrary to the neurobiological data and the most obvious facts about the perception of sound location by human beings, someone might nevertheless maintain that the disparities *were* perceived, and that the organism's knowledge of sound-source location was the result of a higher-level or cognitive computation based on these perceived disparities. Certain disparities would be associated with certain locations, or perhaps there might be some heuristic computation for getting locations from disparities. Let us now consider how experience might be expected to have an effect, depending on whether the modular or cognitive account is the more nearly correct one.

The experience of interest is the change in the relation between the location of the source and the disparity cues as a consequence of changes, either natural or deliberately produced, in the acuity of the ears or the distance between them. The young barn owl's ear may be plugged by the curious experimenter; the child's ears get farther apart as his head grows. In either case, there is a change in the amount of disparity for a given deviation in source location from the midline or a given amount of head rotation. Yet the owl somehow makes the appropriate adjustments in its sound-localizing behavior, and the child continues to localize sounds correctly as he grows (Knudsen, 1988).

On the first and generally accepted view of sound localization, the effects of experience must take place within the module. Indeed, exactly this has been found to be the case in the asymmetrically deafened barn owl, for Knudsen (1988) has shown that the neural map is itself recalibrated so as to maintain the veridical relation, obtained before the owl was deafened. Thus, it is the perceptual representation itself that is "corrected," not the cognitive connection between this representation and others. Apparently, the module adapts to the new environment at a precognitive level. In the case of the child, one can plausibly suppose that something of the same sort occurs when the disparities change as the head grows bigger.

On the other view of sound localization, the organism would have first to learn—by trial and error, logical inference, or instruction—that a certain seemingly arbitrary range of perceived disparities was, in fact, relevant to sound localization, and then, more specifically, how each disparity was to be interpreted as location. Such learning would, of course, have to take into account the complication that, for a fixed location, the disparities are different as a function of frequency and, even worse, that at each frequency, the disparities change as the head gets bigger. Altogether, a formidable cognitive task. Of course, the task is no less formidable as it is done by the sound-localizing module. The difference is simply that the module is specifically and superbly adapted to its complex task, and carries it out without taxing in the least such cognitive capacities as the child might have.

Thus, the effects of experience that are cognitive contrast with the precognitive kind most obviously in the locus of the effects. In the precognitive kind, the effects are, as we saw, on the internal workings of the relevant module and thus on the perceptual representations themselves; in the cognitive variety, on the other hand, they would have to be in the connection between those representations and others. A further difference is that, while the precognitive calibrations of the module are highly selective in regard to the environmental conditions they respond to, cognitive learning is obviously quite promiscuous, being capable of forming connections to a wide variety of representations. Thus, an animal can be taught to make any of an indefinitely large number of responses whenever it perceives sound at a particular location. But this would in no way affect the localization module or the perceptual representations it produces; those would have changed only in response to environmental conditions that alter the relation between interaural disparities and the location of the sound source.

SPEECH AND EXPERIENCE

Turning now to speech, we see that the conventional view is analogous to the view of sound localization that is incorrect. For it is most commonly supposed about speech that underlying its perception are processes and primitives no different from those of nonspeech (Crowder & Morton, 1969; Kuhl, 1981; Miller, 1977; Oden & Massaro, 1978; Stevens, 1975). All sounds, whether they convey phonetic information or not, are supposed to excite the same specializations of the auditory system and evoke such standard auditory primitives as pitch, loudness, and timbre. The perceived difference between a stop consonant and a Morse code signal are only in the particular values that are assigned to each component of a common set of perceptual primitives. There are no specifically phonetic primitives.

Since, on this view, phonetic structures are not marked as a distinct class, the child must learn, obviously by some cognitive process, which of the indefinitely many percepts that belong to a common auditory mode are relevant to phonetic communication and which are not, and then, more specifically, which percepts are to be assigned to which phonetic categories. In this respect, learning to perceive speech would be, in principle, something like learning Morse code. In practice, it would be very much harder, of course, because, unlike the dots and dashes of Morse code, the sounds of speech bear a peculiarly complex relation to the phonetic structures they convey. One might suppose that, in trying to come to grips with this relation, the child would be aided by the results of experimenting with the acoustic consequences of his own articulatory gestures. But here again the conventional view imposes a considerable cognitive burden, for it assumes that the primitives of the speech production system are not specific to speech, but are rather common to a general action mode. Therefore, the child would have to discover about the phonetically unmarked movements of his articulators, just as he would about the unmarked auditory percepts, which ones were relevant to phonetic communication and what the more specific nature of their relevance might be. And, since the motor primitives would have nothing in common with the perceptual primitives, they would have to be linked, and establishing those links would necessarily be a highly cognitive process, depending, for the most part, on unrestrained trial and error.

So, on the conventional view of speech perception, development would have to take place at a stage beyond the primitives that any module produces. Also, of course, it would be relatively unconstrained in regard to the nature of the signals, processes, or events, that become connected; so, in this respect, too, learning to communicate with speech would be like learning Morse code.

But there is another view of speech, one that has implications more in accord with the most obvious facts of language development (Liberman & Mattingly, 1985, 1989). On this view, there is a phonetic module, a biologically coherent system specialized for the production and perception of phonetic structures. The primitives of this module, common to production and perception, are the articulatory gestures that serve as the building blocks of the phonological system. Thus, the phonetic module produces primitive representations that are specifically phonetic, hence categorically set apart from all others. There is, then, no need for a cognitive process that enables the child to learn to attribute communicative significance to some arbitrarily defined class of otherwise undistinguished representations. Moreover, as in the case of the sound localization module, experience calibrates and recalibrates the perceptual representations by processes that are entirely internal to the module. It is by means of this calibration that the child adjusts to the subset of phonetic gestures appropriate to his own language and to the changing anatomy of his vocal tract. Such precognitive calibration acts on specifically phonetic primitives; the effect of experience is to guide that calibration, not to teach the child how to translate nonphonetic primitives into phonetic categories. A consequence is that the only experiences that count for the module are those that are relevant to the phonetic environment.

It is particularly appropriate that we consider this matter in a book that honors James Jenkins, because there is an experiment by Jenkins and his colleagues that provides relevant data (Miyawaki, Strange, Verbrugge, Liberman, & Jenkins, 1975). This experiment was designed to assess the effects of linguistic experience on phonetic perception, and to find the locus of the effect. Using synthetic approximations to the syllables [ra] and [la] that differed only in the extent and direction of the third-formant transition, Jenkins and his colleagues found, first, that native speakers of English reliably sorted the syllables properly and showed a pronounced peak in discrimination at a point on the acoustic continuum of third-formant transitions that corresponded to the English phonetic boundary. Speakers of Japanese, on the other hand, discriminated the syllables very poorly, and their discrimination functions showed no signs of a peak at the point that corresponded to the English boundary. It is important that the two groups differed, not just in their ability to attach phonetic labels appropriately, but in the functions that were generated when they tried simply to discriminate one stimulus from another on any basis whatsoever. This indicates that the American listeners did not perceive these stimuli as the Japanese did, and then, by some cognitive process, apply the phonetic labels their language had taught them. Rather, the difference was in the precognitive, purely perceptual aspects of the process. And, obviously, the difference was a result of the differing linguistic experience of the groups, for, as is well known, the [r]–[l] distinction is not functional in Japanese. But Jenkins and his colleagues also undertook to find out just what it was that linguistic experience had affected. For that purpose they tested the ability of the American and Japanese listeners to discriminate the critical third-formant transition cue when, in isolation from the rest of the syllable, it did not sound like speech, but rather like a nonspeech 'bleat'. The result was that the two language groups discriminated the critical acoustic cue equally well. Thus, effect of linguistic experience was specifically on the phonetic system, not more generally on auditory perception.

The results of the experiment by Jenkins and his colleagues accord well with the view advanced in this chapter. Relevant linguistic experience acts on the internal workings of a phonetic module, with the result that the effect is on the representation itself, not on the way it becomes cognitively attached to phonetic labels or prototypes that exist at some further stage.

REFERENCES

Crowder, R. G. & Morton, J. (1969). Pre-categorical acoustic storage (PAS). *Perception & Psychophysics, 5,* 365–373.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Hafter, E. R. (1984). Spatial hearing and the duplex theory: How viable is the model? In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Dynamic aspects of neocortical function* (pp. 425–448). New York: Wiley.

Knudsen, E. I. (1984). Synthesis of a neural map of auditory space in the owl. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Dynamic aspects of neocortical function* (pp. 375–396). New York: Wiley.

Konishi, M., Takahashi, T. T., Wagner, H., Sullivan, W. E., & Carr, C. E. (1988). Neurophysiological and anatomical substrates of sound localization in the owl. In G. M. Edelman, W. E. Gall, & M. W. Cowan (Eds.), *Auditory function: Neurobiological Bases of Hearing,* (pp. 137–149). New York: Wiley.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *J. Acoust. Soc. Am., 70,* 340–349.

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

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243,* 489–494.

Miller, J. D. (1977). Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory mechanisms. In T. H. Bullock (ed.), *Recognition of complex acoustic signals* (Life Sciences Research Report 5, pp. 49–58). Berlin, Dahlem Konferenzen.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., & J. J. Jenkins. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18*(5), 331–340.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85,* 172–191.

Stevens, K. N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant & M. A. Tatham, (Eds.), *Auditory analysis and perception of speech,* New York: Academic Press.

COGNITION AND THE SYMBOLIC PROCESSES:

Applied and Ecological Perspectives

Edited by

Robert R. Hoffman
Adelphi University

David S. Palermo
The Pennsylvania State University

LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1991 Hillsdale, New Jersey Hove and London