The goal of a biological psychology is to undermine the autonomy of whatever it studies. For language, the goal is to derive its properties from other, presumably prior, properties of the human organism and its natural environment (cf. Lindblom, 1980). This does not mean that we should expect to reduce language to a mere collection of non-linguistic capacities in the individual, but it does mean that we should try to specify the perceptual and motor capacities out of which language has emerged in the species. The likelihood that this endeavor will go far with syntax in the near future is low, because we still know very little about the perceptuomotor principles that might underlie syntactic capacity—but that is why current study of syntax is, from a biological point of view, descriptive rather than explanatory. But the prospects are better for phonology, because phonology is necessarily couched in terms that invite us to reflect on the perceptual and motor capacities that support it.

As we come to understand the extralinguistic origins of the sound pattern of language, we may also come upon hypotheses as to its perceptuomotor mechanisms. Those hypotheses must be compatible with (and may even derive from) our hypothesis as to phylogenetic origin. If we forget this, we risk offering tautology as explanation, because we are tempted to attribute descriptive properties of language to the organism rather than functional properties of the organism to language (cf. Turvey, 1980). I believe that this happens at several points in the otherwise excellent discussions of infant and adult speech perception by Eimas (in press) and of hemispheric specialization by Morais (in press). Both authors, at some point, take a descriptive property of language, its featural structure, and attribute a matching mechanism of featural analysis to the language perceiver. This, of course, is mere tautology. Plausible hypotheses as to the nature of the perceptual mechanism must await a deeper understanding of the functions and extralinguistic origins of linguistic structure.

*This article is a revised version of a paper given at the Centre National de la Recherche Scientifique (C.N.R.S.) Conference on Cognition, held at the Abbaye de Royaumont, France, June 15-18, 1980, and will be published in the proceedings of that conference.
+Also at Queens College and the Graduate Center, City University of New York. Acknowledgment: Preparation of this chapter was supported in part by NICHD Grant HD-01994 to Haskins Laboratories. I thank Ignatius Mattingly for his careful reading and for his instructive comments.

[HASKINS LABORATORIES: Status Report on Speech Research SR-65 (1981)]
Consider, in this light, the data and inference that have led to current interest in features and the perceptual mechanisms that supposedly extract them from the signal. The story begins with early studies intended to define the acoustic boundaries of phonetic categories (e.g., Cooper, Liberman, Delattre, & Gerstman, 1952). The experimental paradigm entailed synthesizing a consonant-vowel syllable, varying some property, or set of properties, along an acoustic continuum from one phonetic category to another, and then calling on listeners to identify or to discriminate between the syllables. Since the end-point syllables typically differed from each other by a single phonetic feature, such as manner or place of consonant articulation, the procedure served to specify an acoustic correlate of that feature.

As is well known, listeners typically divide such a continuum into sharply defined categories and, when asked to discriminate between syllables, do well if the syllables belong to different categories, badly if they belong to the same category, so that a peak appears in the discrimination function at the boundary between categories. This phenomenon, termed "categorical perception," was of interest for several reasons. First, it was believed to be peculiar to speech; second, it was assumed to be the laboratory counterpart of the process by which listeners categorize the acoustic variants of natural speech; third, the sharp categories and poor within-category discrimination hinted at some specialized mechanism (such as analysis-by-synthesis or a feature detecting device) for transforming a physical continuum of sound into the abstract, opponent categories that are the stuff of phonetic and phonological systems.

In due course, the experiments of Eimas and his colleagues, using "high amplitude sucking" with infants and selective adaptation with adults, led to an explicit model of categorical perception, in particular, and of phonetic perception, in general. This work has already stimulated almost a decade of invaluable research from which there has emerged a preliminary taxonomy of the infant's perceptual capacities for speech. However, the model that the research has inspired is weak on several counts. In its early versions, the model invoked devices for extracting abstract, phonetic features; later versions, faced with accumulating evidence of contextual dependencies in selective adaptation (e.g., Bailey, 1973), not to mention the unexpected skills of the chinchilla (Kuhl & Miller, 1978), substituted acoustic for phonetic feature detectors (Eimas & Miller, 1978).

But consider the difficulties. First, we now know that categorical perception is not peculiar to speech, nor even to audition (e.g., Pastore, Ahroon, Baffuto, Friedman, Paleo, & Fink, 1977), so that students of speech perception are excused from postulating a specialized mechanism to account for it. Second, we have no grounds for supposing that the laboratory phenomenon of categorical perception has anything more important in common with the categorizing processes of normal listening than that they both involve classifying variants. The acoustic variations within categories of natural speech are either prosodic variants associated with a particular phone in a particular segmental context (e.g., [d] before [a]), spoken at different rates, with different stress and so on, or segmental variants, intrinsic to the production of a particular phone in different contexts (e.g., [d] before [a] or [i]). These are the types of variant that the listener has to categorize in natural speech, and neither of them is known to be mimicked by
the continua of synthetic speech. Indeed, acoustic variants that surround a phonetic boundary on a synthetic continuum (where all the interesting experimental effects appear, such as discrimination peaks and adaptive shifts in identification) may not only never occur in natural speech, but may even be literally unpronounceable (as in a synthetic series from [b] to [d], for example). They can hardly therefore operate as psychologically effective barriers to ensure a "quantal" percept (Stevens, 1972).

The third and most serious weakness is with the presumed role of acoustic feature-detecting devices in speech perception. As we have noted, the categorical perception paradigm typically manipulates a single dimension of the signal at a time to assess its contribution to a particular phonetic contrast. However, virtually every phonetic contrast so far studied can be cued along several distinct dimensions, and the various cues then enter into trading relations. The precise position of the boundary along a synthetic continuum for a given cue varies with the values assigned to other contributing cues. The most familiar instance comes from trading relations among cues to the voicing of syllable-initial stop consonants (e.g., Lisker & Abramson, 1964; Summerfield & Haggard, 1977), to which burst energy, aspiration energy, first formant onset frequency, fundamental frequency contour and the timing of laryngeal action all contribute. Other instances are provided by cues to the fricative-affricate distinction (Repp, Liberman, Eccardt, & Pesetsky, 1978), to stops in English fricative-stop-liquid clusters (Fitch, Halwes, Erickson, & Liberman, 1980) and in fricative-stop clusters (Bailey & Summerfield, 1980), and so on (for a preliminary review, see Liberman & Studdert-Kennedy, 1978). Are we to assign a new pair of opponent feature detectors (with contextually dependent, "tuneable" boundaries) to each new dimension that we discover? This may be difficult since, as several authors have remarked (e.g., Lisker, 1978; Bailey & Summerfield, 1980; Remez, Cutting, & Studdert-Kennedy, 1980), the number of isolable dimensions, relevant to any particular perceptual distinction, may have no limit.

We cannot escape from this reductio ad absurdum by positing fewer and higher order detectors, because the absurdity lies in the detectors, not in their proliferation. For example, the goal of Stevens' work (e.g., Stevens, 1975; Stevens & Blumstein, 1978) is to arrive at an integrated, summary description of the cue complex associated with each phonetic feature contrast. Thus, in his work on stops, Stevens describes various general properties of the whole spectrum, using the terminology of distinctive feature theory (e.g., grave-acute, diffuse-compact), and posits a matching set of acoustic "property detectors." This ensures that the number of supposed detectors will be no more than exactly twice the number of distinctive feature contrasts. However, by adopting the terminology of phonological theory, it also makes plain that we are dealing with tautology, not explanation.

The error in postulating detectors does not lie therefore in the claim that the signal undergoes analysis along several channels—that might even be true. Rather, the error lies in offering to explain phonetic capacity by making a substantive physiological mechanism out of a descriptive property of language. The error is attractive, because the feature or property detector has a veneer of biological plausibility: it promises to link language with ethology, on the one hand, through the trigger features of Tinbergen (1951; Mattingly, 1972) and the bird-song templates of Marler (1970), and with
physiology, on the other, through the selectively responsive cells of the bullfrog (Capranica, 1965), the cat (Whitfield & Evans, 1965), and the squirrel monkey (Wollberg & Newman, 1972). Yet, whatever the importance of this single-cell work to physiology, its psychological import is nil, since it merely supports the truism that some isolable and distinctive physiological event corresponds to every isolable and distinctive property of the physical world to which an organism is sensitive. The notion of innate song or call templates has even less to offer for an understanding of human language ontogeny. Such devices may ensure species recognition and successful reproduction among organisms, such as the chaffinch and the bullfrog, which have brief or non-existent periods of parental care, and therefore, little or no opportunity to discover the marks of their species. But this is not the human condition. And, given the varied solutions to the problem of learning a species-specific song, even among closely related species of songbird (Kroodsma, 1981), it is implausible to suppose that we can explain language ontogeny by invoking mechanisms proper to animals with a different ecology and for which we have no evidence in the human (for elaboration, see Studdert-Kennedy, 1981). What we should be asking instead is: What function does the capacity for perceptual analysis fulfill? Or, a little differently, what properties of the human organism force language into a featural structure?

Before I suggest an approach to this question, let me comment on another area of research where we run into a dead end, if we do not raise the question of biological function: hemispheric specialization. Morais (in press) brings together an impressive body of experimental findings from laterality studies, and shows conclusively that we simplify and gloss over discrepancies, when we characterize the left hemisphere as linguistic, the right as non-linguistic. He proposes to resolve the discrepancies by superordinate classification of the tasks at which the hemispheres excel,terming the left hemisphere "analytic," the right "holistic."

These descriptions certainly provide a fair partition of the reported data. But there are two objections to the proposal. First, it is too narrow, because it confines itself to the supposed perceptual modes of the hemispheres. Yet we act no less than we perceive: perception is controlled by, and controls, action. Therefore, it is the joint perceptuomotor processes that we should try to capture in a description of a hemispheric mode. Second, the proposal is too broad, because it does not consider the question of phylogenetic origin. Presumably, a behavioral mode (if there be such) does not evolve without a behavior to support. But Morais has no suggestions as to what that behavior might be. For my part, I am inclined to suppose that it might be language.

In any event, the linguistic capacities of the left hemisphere, in most individuals, are attested to by a mass of clinical and experimental data (e.g., Milner, 1974; Zaidel, 1978; Zurif & Blumstein, 1978). These capacities call for more than mere classification with supposedly kindred skills: they call for explanation. That is, they raise the question: What property of the left hemisphere predisposed it to language? Three items of evidence converge on a possible answer. First is the dominance of the left hemisphere in the motor control of speech for some 95% of the population. Second is the dominance of the left hemisphere in manual praxis for some 90% of the population. Third is the recent demonstration that American Sign Language
(ASL), the first language of some 100,000 deaf individuals in the United States, has a defining property of primary, natural languages: a dual pattern of formational structure ("phonology") and syntax (Klima & Bellugi, 1979). Presumably ASL uses the hands rather than, say, the feet, because the hand has the speed and precision to support a rapid, informationally dense signaling system of the kind that a language demands.

Taken together, these facts almost force the hypothesis that the primary specialization of the left hemisphere is motoric rather than perceptual. Language would then have been drawn to the left hemisphere because the left hemisphere already possessed the neural circuitry for control of fingers, wrists, arms and for unilateral coordination of the two hands in the making and use of tools—precisely the type of circuitry needed for control of larynx, tongue, velum, lips and of the bilaterally innervated vocal apparatus. (Perhaps it is worth remarking that the only other secure instance of cerebral lateralization is also for control of a complex bilaterally innervated vocal apparatus—in the canary [Nottebohm, 1977]).

The general hypothesis is not new. Semmes (1968), for example, proposed such an account of the cerebral link between speech and manual control. She argued from a study of the effects of gunshot lesions that the left hemisphere was focally organized for fine, sequential, sensorimotor control, while the right was diffusely organized for holistic perception and action. Recently, Kimura (e.g., Kimura & Archibald, 1974; Kimura, 1979) and Kinsbourne (e.g., Kinsbourne & Hicks, 1978) have carried the hypothesis further, looking for evidence of competition and facilitation between speaking and manual action. Current research is developing procedures and paradigms to increase the precision and rigor of such work (Kelso, personal communication).

What insight can this motoric view of language and hemispheric specialization lend into the origins of phonetic features? Note, first, that the signs of ASL, no less than the syllables and segments of spoken language, can be economically described in terms of features (Klima & Bellugi, 1979). Moreover, the articulators of both vocal tract and hands are relatively few: most are engaged, even if only passively, in the production of every sign or syllable. An ample repertoire of units therefore calls for repeated use of the same gesture by the same articulator in combination with different actions of other articulators. These recurrent gestures are, we may surmise, the instantiation, alone or in combination, of phonetic features (Studdert-Kennedy & Lane, 1980). However, the features are not detachable entities; rather, they are recurrent properties or attributes of the signs and segments (Fowler, Rubin, Remez, & Turvey, 1980; Turvey, 1980; Bladon & Lindblom, in press). This view sits comfortably with recent evidence that metathesis tends to involve unitary phonetic segments rather than features (Shattuck-Hufnagel & Klatt, 1979). And from this we may well infer that, just as they are not put in, features are not taken out. That is to say, the perceived feature is an attribute, not a constituent, of the percept, and we are absolved from positing specialized mechanisms for its extraction.

None of what I have said above should be taken to imply that speech is not the peculiar and peculiarly efficient acoustic carrier of language. On the contrary, speech is peculiar and distinctive precisely because its processes of production and perception must have evolved pari passu with
language itself. Just how speech gives the listener access to his language is still a puzzle, and not one that seems likely to be solved by bare psychoacoustic principle.

Let me illustrate with two recent experiments. First is a study by Fitch, Hailer, Erickson, and Liberman (1980), demonstrating the perceptual equivalence, in a speech context, of two distinct cues to a voiceless stop in a fricative-stop-liquid cluster: silence and rapid spectral change. These investigators constructed two synthetic syllables, [plt] and [lxt], the first differing from the second only in having initial transitions appropriate to a labial stop. If a brief bandpassed noise, sufficient to cue [s], was placed immediately before these syllables, both were heard as [slt], but if a small interval of silence (long enough to signal a stop closure) was introduced between [s] and the vocalic portion, both were heard as [split]. What is of interest is that the silent interval necessary to induce the stop percept was shorter when the vocalic portion carried transitions than when it did not. By systematically manipulating the duration of the silent interval before each of the two syllables, Fitch et al. titrated the effect of the initial transition and found it equivalent to roughly 25 msec of silence. Moreover, they demonstrated that these two diverse cues—silence and spectral shift—were additive (or multiplicative) in the sense that discrimination between [slt] and [split] was close to chance when the cues were in conflict (e.g., a short interval + [plt], or a long interval + [lxt]), but was facilitated when they worked together: a long interval + [plt] was usually perceived as [split], a short interval + [lxt], as [slt]. Presumably, the grounds of this spectral-temporal equivalence are simply that the duration of stop closure and the extent of a following formant transition covary in the articulation of a natural utterance. Certainly, there are no psychoacoustic grounds for expecting the equivalence, and we may therefore fairly conclude that it is peculiar to speech.

In fact, Best, Morrongiello, and Robson (in press) have demonstrated just this in an ingenious experiment using "sine-wave speech" (cf. Remez, Rubin, Pisoni, & Carrell, in press). Best and her colleagues constructed a sound from three sine waves modulated to follow the path of the center frequencies of the three formants of a naturally spoken syllable, [dx], in two forms: one form had a relatively long initial F1 transition ("strong" [dx]), one had a relatively short initial F1 transition ("weak" [dx]). Given a perceptual set for speech, some listeners identify these sounds as [dx] and [ex], while others hear them as different non-speech chords. If a suitable patch of noise is placed immediately before these sounds, they can be heard as [ex]; if a sufficient silent interval is introduced between noise and sine waves, a "speech" listener will hear [sdx], and he will hear it with a shorter interval before "strong" [dx] than before "weak" [dx].

On this basis, Best et al. constructed two continua, analogous to those of the earlier experiments, varying silent interval in combination with one or other of the [dx] "syllables." To obtain identification functions without an explicit request for identification, they used an A X B procedure. In this procedure A and B are endpoints of a synthetic continuum. The task of the listener on each trial is to judge X as "more like A" or "more like B." Thus, despite the bizarre quality of their stimuli, Best et al. were able to obtain identification functions and to assess the perceptual equivalence of silence.
and formant transitions in a manner analogous to that of the earlier /slit-slit/ studies. Their fifteen listeners divided themselves neatly into three groups of five. Two of these groups never heard the sounds as speech and demonstrated no perceptual equivalence between silence and spectral change: one group was sensitive to variations in silence, but not in frequency, the other to variations in frequency, but not in silence. Only the five listeners who heard the sounds as /ser/ or /ser/ demonstrated a trading relation between silence and spectral change.

The burden of this elegant study matches the conclusion drawn by Jusczyk (in press) from his review of infant research and by my colleague, Donald Shankweiler, and me some years ago from a dichotic study: "...the peculiarity of speech may lie not so much in its acoustic structure as in the phonological information that this structure conveys. There is therefore no reason to expect that specialization of the speech perceptual mechanisms should extend to the mechanisms by which the acoustic parameters of speech are extracted" (Studdert-Kennedy & Shankweiler, 1970, p. 590).

If this conclusion is correct, we may review the goals of those who hope to advance our understanding of the biological foundations of language by studying infants. Their proper task is not so much to establish psychoacoustic capacity as to track the process by which infants discover the communicative use and linguistic organization of the sounds they hear and the signs they see (cf. MacKain, Note 2). This is the species-specific, epigenetic process for which we shall find no counterpart in the chinchilla.

REFERENCE NOTE

1. Bailey, P. J. Perceptual adaptation for acoustical features in speech. Speech Perception (Department of Psychology. The Queens University of Belfast), 1973, Series 2, 29-34.
2. MacKain, K. S. On assessing the role of experience in infant speech perception. (Forthcoming)

REFERENCES

Best, C. T., Morrongiello, B., & Robson, R. The perceptual equivalence of two acoustic cues for a speech contrast is specific to phonetic perception. Perception & Psychophysics, in press.


