Journal of Phonetics (1983) **11**, 303–322

Realism and unrealism: a reply

Carol A. Fowler

Dartmouth College, Hanover, New Hampshire 03755, and Haskins Laboratories, New Haven, Connecticut 06510, U.S.A.

Received 15th November 1982

Abstract: A recent article by Hammarberg (1982) criticizes my theoretical perspective on coarticulation as both "internally" and "externally incoherent". A major source of incoherence, in Hammarberg's view, is the alleged claim that phonetic segments are physical entities, and hence, not mental entities. The present paper ascribes the incoherence largely to Hammarberg's understanding of the perspective, pointing out that entities can be at once psychological ("mental") and physical. More usefully, perhaps, the article attempts to establish the philosophical, theoretical and practical grounds on which Hammarberg's own viewpoint and mine disagree most importantly. It is argued that the disagreements are substantial (mirroring, in part, those recently expressed between Fodor & Pylyshyn (1981) and Turvey *et al.* (1981)), and that their analysis may lead to more careful development of both theoretical perspectives; this should, in turn, promote a deeper understanding of speech itself.

Introduction

I have two goals in writing a reply to Hammarberg's recent manuscript, "On redefining coarticulation" (1982). One is to suggest that the disagreements between the "mentalist" view of speech production espoused by Hammarberg, ostensibly in the Chomskian tradition, and the view that I have proposed (Fowler, 1977, 1980*a*; Fowler, Rubin, Remez & Turvey, 1980) in the Gibsonian or "ecological" tradition (e.g. Gibson, 1966, 1979; Turvey, Shaw, Reed & Mace, 1981; Shaw, Turvey & Mace, 1982) are in fact interesting and substantial. With development, I believe, they may be exploited to promote more careful expression and development of both theoretical perspectives, and therefore, perhaps, also to promote understanding and explanation of speech itself.

This approach is guided by the recommendation of Feyerabend (1978) that researchers take an "anarchistic" approach to scientific study. He argues that individual theories develop optimally in a climate of several mutually inconsistent, but (apparently) factually adequate theories. In such a competitive climate, facts crucial to tests of individual theories can most readily be identified as such.

In the present instance, this would not be a reasonable approach if the disagreements between the theoretical perspectives were those outlined by Hammarberg; the physicalist, naive realist, radical empiricist view ascribed by him to me is clearly factually inadequate. My second, regretted, goal is to correct inaccuracies in Hammarberg's exposition of my proposals. This latter would better have been accomplished off-the-record before his paper was published, but the opportunity did not arise.

0095-4470/83/040303 + 20$03.00/0

In other publications (e.g. Fowler *et al.*, 1980), I have argued for the point of view proposed in Fowler (1980*a*). I will not repeat those arguments here except where I find it impossible to resist. Instead I will try to make clear the theoretical perspective itself, to defend it as a coherent one, and to contrast it with the "mentalist" view that Hammarberg espouses.

I plan to contrast the views in two domains, one, the domain of philosophical commitment on which Hammarberg focuses. This will require clarification and correction of the commitments ascribed to me by Hammarberg. The second domain is the theoretical and practical one of scholarly or scientific investigation. These contrasts will be addressed under "Contrasts in the domains of theory and practice". Finally, under "A tutorial", I will consider some specific challenges to the ecological view raised by Hammarberg.

Philosophical contrasts

Hammarberg classifies my approach as physicalist (as opposed to interactionist or mentalist), naive realist and empiricist ("of the empty mind"). I will consider each classification in turn.

Physicalism versus mentalism

Physicalism is the view that "all mental terms can be defined in a schema whose undefined descriptive predicates refer to characters exemplified by physical objects" (Bergmann, 1967, p. 48). That is, statements about mental events can be reduced to primitive (undefined) terms with physical characteristics as referents (e.g. Dennett, 1978, p. 94). Mentalism, then, is the view that theoretical terms referring to mental events cannot be so reduced.

Described this way, these two positions are mutually exclusive. Indeed, to some of its opponents at least (e.g. Skinner according to Dennett, 1978), mentalism is equated with dualism—the view that mind is spiritual and not physical. However, Chomsky, along with many philosophers—including Dennett and Fodor, who ally themselves with the Chomskian tradition, and others who do not (e.g. Putnam, 1973; Pattee, 1970; Medawar, 1974)—has argued that versions of physicalism and mentalism are compatible.

All of the cited authors seem to agree that mental events are physical as well. For example:

> The mentalist, in this traditional sense, need make no assumptions about the possible physiological basis for the mental reality that he studies. In particular, he need not deny that there is such a basis. (Chomsky, 1965, p. 193)

> To begin with, let us assume that it makes sense to say, as we normally do, that each person knows his or her language, that you and I know English for example, that this knowledge is in part shared among us and represented somehow in our minds, in structures and we can hope to characterize abstractly, and in principle quite concretely, in terms of physical mechanisms. (Chomsky, 1980*a*, p. 5)

> I think it is very likely that all of the organismic causes of behavior are physiological, hence that mental events have true descriptions in the vocabulary of an ideally completed physiology. (Fodor, 1975, p. 9)

> [T]he coexistence of physicalist doctrine with intentional or mentalist vocabulary, while perhaps not having received the justification it ought to have, is a typically undefended and unattacked feature of current discussions. (Dennett, 1978, p. 94)

In these undefended and unattacked views, mental events are seen as special kinds of physical events. Because they are special, statements about them couched in physical terms, though in a limited sense descriptively exhaustive (e.g. Pattee, 1970; Putnam, 1973), do not capture their distinct nature.

This is the view of psychological concepts that I have explicitly adopted (cf. Fowler, 1977; Fowler & Turvey, 1982). Unless Hammarberg intends to champion the idea that mind is spiritual–in apparent contrast to the philosophical view with which he identifies himself –there is no disagreement between us here.

In any case, where physicalist and mentalist views are seen to contrast, *when* they are seen to contrast, it is on the issue of how mental concepts are *embodied* in neural structures if at all. In fact Hammarberg's objections to my view of speech production do not seem to lie there in the relationship between mind and brain, but rather in the relationship between "mental concepts" and their real-world counterparts, if any. This suggests a disagreement on the issue of realism, not physicalism and mentalism–a notion bolstered also by Hammarberg's criticism of my view as "naive realist".

Naive realism

I found this ascription depressing because it joins at least one other published, but also erroneous, ascription of naive realism to theoretical developments in the Gibsonian tradition (see Henle, 1974). It was the more disheartening because it was, I thought, clearly refuted in excerpts from my paper that Hammarberg quoted at length.

"Realism" is the view that a real world exists; naive realism is the view that "what is known about the world is both unexpurgated and unadulterated" (Shaw & Bransford, 1977, p. 18); hence the perceived world is the real world. (See Shaw *et al.*, 1982, for a discussion of various forms of the realism view as they relate to theories of perception.) The theoretical perspective I have advocated is a realist perspective; but it is not naive realism. My example of the concept of "footwear", however amusing, refutes this ascription. I reproduce here the passages that Hammarberg excerpts:

> For example, . . . it probably takes a human mind to recognize something as being an example of "footwear". The essential properties of footwear–e.g. that it be foot-sized and shaped, that it have a means of attachment to the foot and so on–clearly are given in the optical signal to an eye. However, that collection of properties is only a significant collection for an animal that wears things on its feet. Other animals will not recognize that aggregate of properties as a significant collection and hence will not detect that a shoe, for example, is footwear. (Fowler, 1980*a*, p. 115)

And further:

> Clearly the aggregate of properties–foot-shaped and -sized, protection for the foot, etc.–is available to the light to an eye when that eye focuses on an instance of footwear. Thus the essential properties of the concept are "objectively" manifest. But it takes a wearer of shoes to detect that particular aggregate of properties as a significant collective. (p. 121)

Hammarberg has a number of difficulties with these passages, including scepticism that the relevant properties of footwear are given in the stimulation to an eye. This latter difficulty probably can be ascribed to an unclarity in my expression of the case that is not relevant in the present context. For the present, I want to point out that a naive realist would not write

such a passage. From the perspectives of naive realism, the perceived world is the real world. Hence it could not be *selective* of real-world properties, nor could different perceptual systems be selective in different ways.

A more appropriate ascription would be "direct realism" (Blackmore, 1979; Shaw *et al.*, 1982). It is the view, associated most closely in the psychological study of perception with Gibson (1966, 1979), that "people with normal sensory organs can directly notice *part* of the external physical world . . . [W]hat is perceived is in no way determined by the process of knowing or by the presence of consciousness" (Blackmore, 1979, p. 128). In this view, what is perceived is real, but less than all that is real is perceived.

As for Hammarberg's own view on the issue of realism, he is clear in asserting that phonological segments are not articulated and are not directly given in stimulation. However, two sources of evidence suggest that in his view, subjective existence of phonological segments does not imply that they are not real. First, he objects to my suggestion that phonological segments seen as subjective categories are thereby "fictitious" (Hammarberg, 1982, p. 131). Presumably this is because workable phonological rules can be written which do depend on the concept of segment. Second, he argues that "science aims for a theory of the real" (1976, p. 355).

In Hammarberg's view, then, phonetic and phonological segments are real, but exist only in the minds of talkers and listeners. This is an unusual sense of real. In general, a view that perceived events have no existence outside the mind would thereby be excluded from the domain of realist theories. However, I can think of a rationale for such a claim that may be what Hammarberg has in mind. Perception succeeds (or *occurs*, for a direct realist) when a perceiver recovers properties of the distal source of stimulation. In the case of a linguistic utterance, the most distal source is the message that the talker wishes to convey. If the listener recovers the message, then perception has succeeded (or occurred).

If this is how Hammarberg sees linguistic communication, then our disagreement is not, perhaps, on the issue of realism any more than it is on the issue of physicalism or mentalism. Instead, largely, it is whether or not it is correct to say that linguistic segments are articulated and that they are given in acoustic signals.

The two questions—whether segments are articulated and whether they are given in acoustic speech signals—are separate, although a negative answer to the first obviates asking the second. Hammarberg answers "no" to the first question on two grounds. He argues that components of language are entities of linguistic competence and that consequently they are "purely mental":

> Both phonemes and allophones[1] are entities of linguistic competence, something on the order of categories which we impose on sensory experience, and which enables us to grasp such experiences in the first place. (1982, p. 124)

Second, somewhat unnecessarily given his first argument, he argues that no one has ever found articulatory correspondents of phonological segments. In addition, and again unnecessarily, Hammarberg answers "no" to the question whether phonological segments are given in acoustic speech signals, also on the grounds that they have never been observed. In respect to this conclusion, he adopts an indirect, as opposed to a direct realist, stance:

[1] Although Hammarberg identifies his view as a "consistent development of Chomskian theory" (p. 124), he refers in the next sentence and elsewhere to "allophones" and phonemes"—entities which Chomsky (1964) and Halle (1959) explicitly reject as components of linguistic competence.

> There are no invariant physical cues of segmentalness. There is no extra-human, i.e. non-subjective, way of analyzing the speech continuum into discrete parts corresponding to the notion of segment . . . . Nor can segments be inferred from the speech signal or the articulatory movements without resorting to a host of phonological presuppositions. What all this adds up to is that the concept of segment is brought to bear *a priori* on the study of the physical-physiological aspects of language. (1976, p. 355)

In contrast to Hammarberg's view, my working hypotheses are that phonological and phonetic segments are not only psychological ("mental") entities, they are also inherently articulatory. Consequently, I answer "yes" to the question whether phonological and phonetic segments are articulated. Moreover, I claim, descriptions of phonological and phonetic segments in articulatory terms is not a *reduction* as, say, a description in physiological terms would be. Rather, segments are categories definable in articulatory terms that are psychological by virtue of the function they serve in linguistic communication. To argue that they must be either mental *or* physical–here, either psychological or articulatory –is to make Ryle's category-mistake of treating mental concepts as if they were of the same logical type as physical things and therefore of having to suppose that something must be either the one or the other (as in one of Ryle's examples, having to suppose that a fielder may be either catching the ball or displaying *esprit de corps*, but cannot be doing both at once (Ryle, 1949)).

As for the second question, whether phonological segments are realized in the acoustic signal, I adopt the direct realist hypothesis that essential properties of phonological and phonetic segments are given in acoustic speech signals.[2]

Thus, whereas Hammarberg answers both questions–whether phonological and phonetic segments are uttered and whether they are fully specified acoustically–in the negative, my working hypotheses are that they can be answered affirmatively. Below I will present Hammarberg's defense of his position and my reasons for finding alternatives viable.

Phonemes as entities of linguistic competence

For Hammarberg and others (e.g. Repp, 1981; Parker, 1977), there is a "category error" (Repp, 1982; but not Ryle's) involved in any attempt to find phonological or phonetic segments in articulation or in acoustic signals. In Hammarberg's view, phonemes and

[2] Both Hammarberg and I have written as if the issue with regard to the second question were whether or not phonological segments are *in* acoustic signals. Although I do propose that phonological segments are articulated, I do not in fact hypothesize that they are in acoustic signals, but rather that acoustic signals specify them. Air, like light, is a *medium* that can be constructed by events of various sorts. Whereas light is structured by environmental objects and events, expired air during speech is structured as it passes through a moving vocal tract. In either case, by virtue of its acquired structure, a medium provides information about its structuring source (the "distal stimulus") to a sensitive perceiver.

The relevant structures in reflected light and air are not inherent in the medium; they are acquired. Consequently even were invariant correlates for phonetic segments, say, to be found in the acoustic signal, it would not be any more correct to say that phonetic segments are *in* the acoustic signal than it would be to say that the distal source of reflected light, a table say, is *in* the reflected light. Instead, what is in the medium is *information for* the properties of the distal source (the articulated segments or the table), but not the properties themselves.

phonetic segments are components of a language user's linguistic competence and therefore never can be produced or heard. Again:

> Both phonemes and allophones are entities of linguistic competence, something on the order of categories which we impose on sensory experience and which enable us to grasp such experiences in the first place. (1982, p. 124)

Repp agrees, but for a different reason:

> [S]pecial perceptual and cognitive processes *intervene* between the acoustic signal and the phonetic percept. Therefore phonetic categories—consonants, vowels, and even syllables—cannot be said to be in the acoustic signal. *They have no physical properties*—such as duration, spectrum, and amplitude—and therefore *cannot be measured.* (p. 1463)

Both of the claims being made here—Hammarberg's that elements of linguistic competence are perforce incapable of realization in articulation, and Repp's that abstract concepts mediated in perception by (special) mental processes cannot be identified with real-world classifications—are open to debate. I will consider them in reverse order.

First, the occurrence *per se* of perceptual-cognitive processes in the achievement of a percept would not preclude the correspondence of percepts to real-world events, although it would not promote it either. The context in which Repp's conclusion is drawn, however, suggests that the mental processes he has in mind transform the signal; in particular, they integrate components of the signal that are physically separate. Moreover, Repp agrees with Hammarberg that there are no acoustic invariants corresponding to phonetic segments. Consequently, mental processes impose a phonetic interpretation on a signal that does not "objectively" support that interpretation.

As I will argue in the next section, drawing this conclusion depends on accepting conventional segmentations or descriptions of the acoustic signal (and of articulation) as *the* objective ones. No one has shown that these segmentations and descriptions are in fact privileged, however, or that there can be no equally objective ones compatible with the hearer's perceptions.

Hammarberg argues against the physical existence of phonetic segments in part on grounds that they could have no physical existence because they are elements of linguistic competence and hence are purely mental.

The distinction between competence and performance intended by Hammarberg is Chomsky's. In Chomsky's theory of generative grammar, the core of a speaker/hearer's ability to produce and understand language is his or her knowledge of the language constituted as a generative grammar. A generative grammar is a set of rules that recursively enumerates, by free iteration, all and only grammatical sentences of the language. As such, it cannot by itself be the means by which talkers utter sentences and listeners understand them. Instead, it is, at best, neutral with respect to the requirements of talking and listening, and it can, at most, serve as a component of a performance capability.

I find this proposed separation of linguistic knowledge from its use sufficiently unattractive to warrant seeking an alternative (see also, Derwing, 1973; Steinberg, 1975; Linell, 1979). However, I do not find Chomsky's competence-performance distinction unattractive because it entails the idea that the mental rules and representations constituting linguistic competence can have no realizations in behavior; in fact, I do not believe that the distinction entails this idea.

Engaged by an appropriate set of "use rules" (Steinberg, 1975), when a talker wishes to produce a sentence, the set of rules and representations constituting competence will specify a sequence of *phonetic* segments organized by their participation in words and phrases. In Chomsky & Halle's (1968) system, the phonetic segments each are bundles of distinctive features where "each feature is a physical scale" (p. 299) with articulatory values. What, then, prevents the idea that a phonetic segment so specified is in fact uttered?

I can conceive of two arguments construable as refuting the view that linguistic units are realized in articulation and in acoustic signals. First it may be acknowledged (but it is not by Hammarberg; see his reference to allophones in the excerpt above), that phonetic segments are produced, but denied that any more abstract units can be realized. The rationale for this is that any abstract classifications (say of phones into phonological segments) are language-specific conventions. Thus, they are not real-world categories, but instead must be imposed by the mind. This reasoning is fallacious, however. The talkers' "competence" requires them to use phones in particular ways–contrastively, in free variation or in complementary distribution, to use the descriptivists' classifications. When they do so, they are using the conventional classifications of their language *as* classifications. Therefore, they are producing utterances composed at once of phonological segments and of phones. Indeed, their doing so, allows the linguist to discover the classifications.

A second argument that components of linguistic competence are not physically realized is that listeners unfamiliar with the language have great difficulty segmenting an utterance into its words and phonological segments. Knowing a language facilitates extraction of its units from an acoustic signal. This observation has at least two interpretations, however, not just one. On the one hand, it may signify that the units are not in the signal; instead the signal offers no more than "cues" which can be interpreted only by a knowledgeable listener. On the other hand, linguistic competence may be seen to direct the listeners' attention to the segments, which are fully specified in the signal. If this is the case, then in learning a language, one also learns what to listen for in the signal, more or less as an expert has learned what to look for in an X-ray or as an experienced skier has learned what to look for on a snow-covered hillside.

Perception and articulation: direct or indirect?

As I have said, the questions whether phonological or phonetic segments are articulated and whether an acoustic speech signal specifies them are separate. However, some of Hammarberg's reasons for answering "no" in respect to articulation are the same as those which justify his answering "no" in respect to the acoustic signal. These reasons will be addressed here jointly for the two modes.

In a number of contexts, Hammarberg asserts that phonological segments are not articulated or specified in an acoustic signal:

> Segments cannot be objectively observed to exist in the speech signal nor in the flow of articulatory movements. (1976, p. 355)
>
> The segment has no physical existence of the kind demanded by those who would base science on pure objective observation. (1976, p. 355)
>
> [N]o discrete entities correlatable with such [phonemic or allophonic] segments *can* be found in actual physical articulation (nor in the acoustic signal, for that matter). Segmentation . . . is a purely mental phenomenon. (1982), p. 125, underlining added)

> We cannot perceive or cognize a token, except in terms of a type. In other words, perception-cognition is categorical, and without categories there would be no perception-cognition. Since there exist no physical criteria, in terms of physiology, neurology, or acoustics, which suffice to define a token . . . the types shall have to be regarded as primary, and the tokens as derived. (1982, p. 135)

Hammarberg does not defend these claims, and I can think of no defense for them save the rather weak one that "segments have never been observed" (Hammarberg, 1982, p. 129) –knowingly, at least–in visible speech displays. However, on several grounds, this is not a sufficient reason to conclude that "segments do not exist outside the human mind" (1976, p. 355). First:

> We cannot refute a theory merely by failing to observe something which it treats as possible–perhaps we just have not looked far enough; but we refute it at once if we observe something which it treats as impossible. (Sampson, 1975, p. 47)

Indeed, we would not continue to study the acoustic signal if we thought we knew all of the information it had to offer. Discoveries have been made recently and remain to be made. For example, in their recent work, Stevens & Blumstein (1978) have proposed candidate invariants for segments that are quite different in kind from acoustic "cues" as usually conceived. They have characterized the spectral cross-section of a stop at release as a spectral shape that remains invariant over different following vowels. Kewley-Port (1981) has found that time-varying spectra provide better fits to the various manifestations of stops in syllable-initial position than Stevens and Blumstein's static spectra. Even if both these approaches fail in the long run, they serve as important indicators that the search is still open; no one has shown that invariant correspondents to perceived segments cannot exist.

A second consideration derives from the literature on concepts and categories (e.g. Rosch & Lloyd, 1976; Smith & Medin, 1982). Investigation has not uncovered invariant attributes used by observers to recognize individual tokens of natural categories as such. For example, it is not known how observers recognize instances of birds as birds. It is possible that there exist observable properties necessary and sufficient to classify all and only instances of birds that have not yet been identified by investigators. Alternatively, classification may not depend on identification of essential properties–there may be none to be found. Rather, it may, for example, rest on detection of "family resemblances" among class members. Whichever is the case, the class of birds remains (common-sensically anyway) a natural category *with a real-world, not a "purely mental", existence.* That is, the real-world existence of superordinate categories may not depend on the existence of observable invariants necessary and sufficient to define the classification.

Phonological segments may appear special cases of abstract categories because classifications of phonetic into phonological segments differ across languages. This does not raise any special problems, however, as long as the classifications are systematic and capable of articulatory definition. Hammarberg holds a stronger view. He argues that for "physicalism to be taken seriously it must be able to show that whatever is perceived as a sound segment corresponds to something that constitutes a natural class of events or states in physiology or acoustics" (1982, p. 134). Three points need making here. One is that in a view of physicalism that is compatible with mentalism, as outlined above, no correspondence need be presumed between psychological and physiological types (*cf.* Fodor, 1975). Second, as I pointed out earlier, the acoustic signal is an informational medium; therefore, although in a view that phonological segments have a physical (articulatory) realization, acoustic

structure must provide information about phonological types definable in articulatory terms, the categories it specifies do not have to be *acoustic* types. Third, here as before, I do not really think that physicalism is at issue. The issue is whether it can be said that abstract linguist entities, and rules, for that matter, can be articulated and can be specified in an acoustic signal. There is no compelling reason, known to me, why the answer to that should be no.

A third reason to discount Hammarberg's rejection of the articulatory reality of phonological segments and their reflections in the acoustic signal on grounds that they have not been observed is that his alternative is no better supported by observation. The "perceptual-cognitive" processes that are alleged to impose classifications on the impoverished acoustic "cues" have never been observed either. Nor will they ever, for they take place, if at all, in the privacy of the mind.

As Sampson points out (see the excerpt above), following Popper (1959) and others, theoretical claims *are* refuted if something is observed that is impossible according to the theory. A candidate, in Hammarberg's view, might be the characters of articulation and of the acoustic signal in relation to that of phonetic segments.

Elsewhere (Fowler *et al.*, 1980), my colleagues and I have discussed three apparent barriers to identification of linguistic segments with segments as produced and specified acoustically. The barriers we considered are that abstract segments are said to be *discrete*, *static* and *context-free* while segments are produced and as reflected in the acoustic signal are *overlapping*, *dynamic* and *context-sensitive*. We suggest that these barriers may be only apparent, and perhaps can be removed if abstract segments and realized segments are characterized slightly differently. I refer the reader to Fowler *et al.* (1980) for our efforts to characterize both abstract and realized segments in a way that does no violence to the lingusitic nature of abstract segments but allows the possibility that they can be articulated and specified acoustically. Here, by way of example, I will suggest why the problems of articulatory and acoustic overlap may not preclude the idea that separate and serially-ordered *linguistic* segments are specified directly in acoustic signals.

A major barrier to discovery of acoustic support for the perception of segments has been the "segmentation" problem. The acoustic signal appears not to offer discrete segments that correspond with phonetic segments. I think that such a conclusion is premature, however. Some insights from research in visual perception perhaps can be helpful in suggesting alternative perspectives on the acoustic signal that may make it appear more compatible with the character of phonological segments as perceived and as they participate in lingustic sequences.

Johansson (1950; see also, 1974) has presented the point-light array in Fig. 1, among others, to viewers. In the display, lights A and C move horizontally in phase. Light B moves diagonally, also in phase with lights A and C. Viewers see horizontal movement of the display with point B moving *vertically* between A and C.

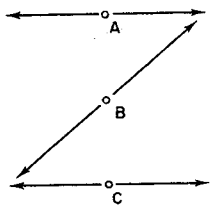

Figure 1

This display can be parsed in several ways. Two segmentations are salient. On the one hand, the lights, being spatially separate, can be considered independent and their independent paths can be tracked. If this is done, then point light B is seen to move diagonally. On the other hand, groupings can be based on common notion of the point lights. All of them move horizontally; in addition, one moves vertically in relation to the horizontal movement. Regarded without reference to the nature of the visual system or of the environment, neither parsing is demanded by the display. Yet viewers have no choice in this instance; they select the second parsing.

The viewers' selection can be explained in either of two ways. One could deny that the selected parsing is, in fact, in the display, and argue, therefore, that it is imposed by the mind on the display–that is, it is the result of a perceptual-cognitive operation performed on the display. An alternative, following Gibson and others, is that the perceived parsing is in the display although it is not the only one the display offers. The viewer's role is one of selective registration of one parsing rather than others.

However achieved, the perceived parsing is realistic. Naturally, movements by parts of a visible scene that are synchronized and occur in the same direction are coupled. When a car moves, its hood, windows, doors and occupants all move in the same direction at the same rate. When an occupant of a car turns his head to talk to another occupant, the movement of his head is a complex vector, just as the movement of point-light B is complex, that combines the forward movement of the car and the sideways rotation of the head relative to the car. Viewers on the street do not see the complex motion vector. Instead, realistically, they see common forward movement of car and occupants and sideways movement of a head relative to the moving frame. The scene is parsed as if by vector analysis, just as Johansson's point light display is parsed.

In the example of the car, the perceived partitioning is in the *physical* event itself. According to direct-realist theory, the partitioning is also available in the stimulation (here, the reflected light).

I have used Johansson's point light display as an example, because it offers a simple but revealing analogue to the acoustic speech signal. Like the point-light displays, the acoustic signal, regarded without reference to the nature of perceptual systems or of the vocal tract, allows many different parsings. A common one for investigators making measurements is one related to Fant's partitioning of the signal in acoustic segments. Acoustic segments are discrete stretches of a signal bounded by qualitatively distinct acoustic events. This is a possible partitioning, and it is useful for some purposes. However, in some ways, it is not a realistic reflection of segment production. First, it masks the well-established fact that articulatory gestures associated with different segments overlap in time. There is very clear evidence, for example, that articulatory gestures associated with the vowels in VCVs extend throughout the interval in which gestures associated with the consonant are produced (e.g. Ohman, 1966; Butcher & Weiher, 1976; Barry & Kuenzel, 1975; Carney & Moll, 1971). This leads to stretches of an acoustic signal in which characteristics of more than one phonetic segment contribute to the character of the signal. This is analogous to the diagonal motion of point light B in Fig. 1 and to the complex motion vector of a car occupant who turns his head while the car is in motion. A realistic parsing of a speech signal, like a realistic parsing of the visible event, cannot use coincidence (temporal in speech, spatial in vision) as information for coherence. It is not reliable information for that.

The parsing of a speech signal into acoustic segments is unrealistic in a second way as well. In Johansson's display, spatial separation is no more evidence for functional separation

than spatial coincidence is evidence for coherence. In the display, points A, B and C are seen jointly to constitute a moving frame of reference even though they are spatially discrete. Similarly, in visible displays of speech, acoustic discreteness (that is, separation by acoustic-segment boundaries) is not evidence for discreteness in production (or perception). Gradual release of consonant closure, for example, may create a burst followed by frication and then by aspiration (Fant, 1973, Chapter 7). These discrete acoustic segments–and more–jointly specify a coherent articulatory event.

It is important to realize, then, that there can be more than one objective way to segment the acoustic signal. Segmentation into more-or-less discrete acoustic segments is one of many ways. The segmentation characteristic of the perceptual system for speech, however, most probably is sensitive to parsings, if any, that reflect true phonetic coherence and separation just as viewers of Johansson's displays select the parsing reflecting functional coherence and separation. I interpret some of my research on perception of vowels coarticulating with other vowels (Fowler, 1981*a*) as suggesting this kind of parsing by perceivers (see "A tutorial" below).

The foregoing discussion suggested that a different perspective on the *acoustic signal* may help to dissolve the barrier to discovery of acoustic correspondents of linguistic segments that the segmentation problem provides. A complementary step in this direction may be to reconsider the nature of *abstract linguistic units* as they participate in a phonological sequence (Fowler *et al.*, 1980). It is generally supposed that linguistic segments are discrete. But if "discrete" means "nonoverlapping", the modifier may be too strong. Whether or not phonological segments are *discrete* seems to me irrelevant to their participation in a lingusitic sequence; what is essential is that they be *separate* one from the other and *serially-ordered.* Writing a phonetic sequence as a sequence of discrete feature columns signals separation and serial ordering in the way that *orthographic* notations signal those properties. But articulation does not preserve separation and serial ordering in this way. All that is required for articulation and its acoustic product to provide *directly* the sequencing of phonological segments is that articulated segments and their acoustic products be separate and serially-ordered; they need not be discrete in the sense of nonoverlapping.

Despite their overlap, acoustic correlates of phonetic segments *are* serially ordered. No segment reaches its maximum dominance of the signal before a segment produced or heard as earlier than it reaches its own maximum dominance. As for what in the acoustic signal, if anything, specifies the separateness of overlapping components of the signal that are heard as separate, and complementarily the coherence of noncoincident components of the signal that are heard to cohere, that is as yet unknown; however, the literature provides sources of guidance. Johansson addresses similar problems in vision, Bregman (e.g. 1977) addresses them in the domain of nonlingustic auditory perception, and more recently Darwin (Darwin & Bethell-Fox, 1977; Darwin, 1982) has begun to study the problems in fluent speech.

In short, I see no barrier, either empirical or logical to the view that phonetic and phonological segments are realized in articulation (at least in "careful" articulation; see Linell, 1979) and are specified in an acoustic speech signal.

Radical empiricism "of the empty mind"

Empiricism is usually opposed to nativism or rationalism. Historically, the views disagreed on the origin of ideas. For rationalists, ideas originate in innate principles of mental functioning, while for empiricists, they originate in experience (e.g. Katz, 1966).

With the possible exception of some behaviorists, there are few avowed radical empiricists. Nor, with the exception of Chomsky and his followers, are there many avowed nativists.

Individuals may disagree on the *extents* to which nature and nurture underlie various capabilities, and on the kinds of things that can be inherited, but few hold a radical version of either position.

The view I have expressed in my published papers is neither radically empiricist nor nativist. Although I am skeptical that representational knowledge–in particular, anything so specific, for example, as the "specified subject condition" (Chomsky, 1980*b*)–will turn out to be the kind of thing that is inherited, I do not disagree that evolution prediposes organisms to engage in certain kinds of activities or to seek certain kinds of information from stimulation. This, too, should have been clear enough from my beleaguered footwear example.

Contrasts in the domains of theory and practice

Theory

Hammarberg wishes to elevate coarticulation to the status of phonological assimilation. His argument is that both allophones and phonemes are categories, and categories are mental impositions on sensory experience. Consequently, both allophones, to which descriptions of coarticulatory effects would refer, and phonemes, to which (according to Hammarberg) descriptions of phonological processes would refer, are mental entities. Descriptions of coarticulatory effects would describe contexts in which different allophones of a phoneme will occur. These descriptions would reveal that coarticulation is largely assimilative. Many phonological rules also describe assimilative processes–vowel harmony, for example. Because both sets of rules are "mental", there is no principled reason to treat them distinctively. Hence, coarticulation might as well be identified with phonological assimilation.

There is, however, a reason to keep coarticulation separate from phonological assimilation. Collapsing the two processes obscures a critical difference between rules such as vowel harmony and coarticulatory assimilative effects. In a recent paper, Anderson (1981) argues that phonology is just what is arbitrary about systematic processes involving phonological segments. That is, phonological processes are just what cannot be exhaustively explained by invoking, for example, articulatory dispositions. (Notice that this is not to deny that phonological processes are achieved in articulation. Rather, it is to say that the reason why the processes are the ones they are cannot be explained fully on articulatory grounds.) Anderson is criticising attempts, more-or-less opposite to Hammarberg's, to reduce phonological processes to phonetic explanations. But his criticism extends to Hammarberg's attempts to elevate coarticulation just as well:

> On this view, it is still very much part of the business of phonologists to look for "phonetic explanations" of phonological phenomena, but not in order to justify the traditional hope that all phenomena of interest can be exhaustively reduced in this way. Rather, . . . the reason is to determine what sorts of facts the linguistic system proper is *not* responsible for: to isolate the core of features whose arbitrariness from other points of view makes them a secure basis for assessing the properties of a language faculty itself. (Anderson, 1981, p. 497)

Most coarticulatory effects are very clearly the sorts of facts the linguistic system proper is not responsible for, because they are not arbitrary from other points of view; in large

part, they reflect either requirements or dispositions of the implementing articulatory system.[3]

The difference between phonological assimilation and assimilative coarticulatory processes becomes clear when vowel harmony is contrasted with the tendency for talkers to co-produce vowels. Vowel harmony is not general to languages. Although it is not *un*natural in the sense that it seems to be an exaggeration of a coarticulatory tendency general to languages, it is a linguistic convention, used for the purposes of conveying linguistic messages. In contrast, many languages, perhaps all, exhibit vowel-to-vowel assimilative coarticulation (e.g. Ohman, 1966, for Swedish; Butcher & Weiher, 1976, for German; Fowler, 1980*b*, for English) even if they do not they have vowel harmony. Whereas vowel harmony is a linguistic process, vowel-to-vowel coarticulation is a process introduced by the special properties of the vocal tract. Thus, presumably, whereas appeal to articulatory dispositions can explain vowel-to-vowel coarticulation (*cf.* Fowler, 1977), no appeal to articulatory dispositions can explain why Hungarian has a rule of vowel harmony while English does not; moreover, if such an appeal were to succeed, vowel harmony would thereby be unmasked as a nonlinguistic coarticulatory process.

This distinction, though important for maintaining a clear understanding of the tiered structure of speech, seems to me independent of the issue addressed earlier of whether phones and phonological segments can be said actually to be uttered. In languages with vowel harmony, utterances *are* vowel-harmonized where appropriate. That vowel harmony is a convention of the language and not an articulatory disposition does not change the fact that harmony is achieved in speech.

Practice

The approach to studying speech production that I have adopted contrasts with the approach of "mentalists" (or "cognitivists" in psychology) in focusing attention not only on the constraints that linguistic messages place on the vowel tract, but also on the contributions to performance of the implementing articulatory system itself. I have motivated this approach elsewhere (Fowler, 1977; Fowler, 1980*a*; Fowler *et al.*, 1980), but Anderson's motivation cited above will do as well. It is to remove from the domain of things requiring explanation in terms of mental computations performed on mental representations whatever deserves explanation instead in terms of dispositions or properties of the articulatory system itself. Here I will contrast this approach with the approach characteristic of a mentalist in the practical domain by contrasting attempts at explanations developed in the two theoretical camps to two unexplained phenomena in speech production.

Declination

In fluent speech, the fundamental frequency of a talker's voice drifts downward over the course of a phrase or sentence. The downdrift, known as declination, exhibits the following, more-or-less well-established properties. The starting fundamental frequency of the talker's voice is higher the more he or she has to say, but the ending fundamental frequency is nearly invariant. The fundamental-frequency peaks in an utterance decline in such a way that one-half of the downdrift is accomplished in one-fourth of the temporal extent of the utterance.

[3] It is true that, in detail, coarticulatory effects differ across languages. This does not mean that they are conventions, however, used to convey linguistic messages. Rather, by hypothesis, an articulatory disposition will manifest itself in a language to a lesser or greater degree to the extent that it conflicts or fails to conflict with requirements to preserve language-specific contrasts.

Sorenson & Cooper (1980; Cooper & Sorenson, 1981) offer the following "psychological" model of declination. A talker calculates when, in seconds, relative to beginning to talk his or her utterance will end. Based on that calculation, a starting peak value is selected; the ending peak is invariant. Intermediate peaks are computed according to the following formula for the slope of a line connecting all of the peaks except the first:

$$(2/3)(P_4 - P_1/t_4 - t_1)$$

(P is a peak value in Hz; t is time from utterance onset in seconds.) This model does a fair job of reproducing peak-to-peak declination in read sentences.

However, as others have pointed out, it is not yet warranted by the data (Cohen, Collier & t'Hart, 1983) and it is a rather extravagant first hypothesis (Simon, 1980). According to Simon:

> If it could be shown that the observed invariants could only be reproduced by a process like the one postulated by the authors, then we would have to live with that fact. But until a proof is forthcoming, I think we should look hard for alternatives . . . . It would be nice if we could show that the invariant was a by-product of a simple physiological process—for example, a relation between the physiological mechanisms for producing speech and the expected duration of that speech. (1980, pp. 542–543)

That such an explanation-type may be forthcoming is suggested by several facts. Nearly all languages exhibit downdrift in fundamental frequency. Moreover, downdrifting tones are more common than rising ones in tone languages. Furthermore, untrained singers produce pitch falls with a shorter latency than pitch rises (Ohala & Ewan, 1973; Sundberg, 1979). All these facts suggest that it is easier for vocal tracts to produce falling than rising fundamental-frequency contours. It is true that the reasons for this have not yet been established. However, despite attempts to argue that no physiological/articulatory account will succeed (Breckinridge, 1977), accounts in these terms are still viable (Cohen *et al.*, 1983).

A sample account might go as follows. Talkers who know that they will be producing a long sentence or clause (either because they see the to-be-uttered sentence before them on a sheet of paper or because of intuition) take a deep breath. This has various effects on subglottal pressure and perhaps on laryngeal tension, say, which raise the starting fundamental frequency over its level if a more shallow breath were taken. This gives a starting fundamental frequency roughly correlated with utterance length. The invariant final value of fundamental frequency may be the lowest frequency at which the talker can readily maintain voicing. Finally, it has not been ruled out that the form of the declination line, in which half of the frequency drop is accomplished in one-quarter of the temporal duration of the utterance, is not simply the consequence of effects of air expenditure on subglottal pressure and laryngeal tension (*cf.* Cohen *et al.*, 1983).

Context-sensitivity of vowel production

Talkers produce normal or near-normal vowels when they clench a bite-block between their teeth which prevents normal participation by the jaw in production of vowels (e.g. Lindblom, Lubker & Gay, 1980; Fowler and Turvey, 1980). It is not remarkable that talkers can produce context-sensitive vowels; the variety of segmental contexts in which vowels are produced leads to adoption of differing jaw–tongue contributions to the production of a given vowel in different contexts. What is remarkable is that, presumably, the context supplied by the bite-block is somewhat novel; yet vowels are near-normal from the first pitch pulse.

Two explanations have been proposed recently to account for bite-block findings. Lindblom *et al.* (1979) suppose that segments are produced by "predicative simulation". A talker stores motor commands for each segment. When a segment is to be produced, rather than implementing its appropriate motor commands immediately in the vocal tract, a talker first implements them mentally. This is necessary to attune them to the particular context in which they will be implemented physically. In the simulation, the effects of realizing the segments' motor commands in the vocal tract given its present state are computed and the discrepancy between the projected effects and the intended effects are used to adjust the stored commands to fit current requirements. The stored commands for vowels generally will specify some jaw movement. In the case of speech produced with a bite block, the simulation will reveal that these commands will not succeed in generating the intended vowel because, despite motor commands to move the jaw, it will not move. Hence appropriate commands for compensatory tongue movement will have to replace those for jaw movement.

An alternative model, no less compatible with the bite-block data, is that the intention to produce a vowel is realized as a physiological linkage between jaw and tongue capable of bringing about movement towards a target tongue-palate relationship. The linkage ensures that the target tongue and jaw positions are negatively correlated. To the extent that a bite-block keeps the jaw in an unusually open position for a given vowel, then, by virtue of the linkage, the tongue is positioned higher than usual relative to the jaw to preserve the vowel-characteristic tongue–palate relationship.

Elsewhere (Fowler & Turvey, 1980) I have argued that the first, computational model is far more complicated than it needs to be to handle the data for which it was devised. A streamlined computational model apparently could do as well. In any case, the computational model cannot handle the observation that talkers' generative capacities extend to unexpected perturbations (Folkins & Abbs, 1975; Hughes & Abbs, 1976; Kelso, Tuller & Fowler, 1982). Once a mental simulation has taken place and the motor commands have been adjusted to the context present during the simulation, segment production should be at the mercy of subsequent unforseen perturbations. But the research of Abbs and his colleagues, in which the jaw is unexpectedly perturbed, reveals immediate compensation by the upper lip in bilabial consonant production. Unless the jaw–tongue system for vowel production is, for some reason, different from the jaw–lip system for bilabial-consonant production, vowel production too should be generative even in confronting unexpected perturbations.

For present purposes, it is less important to show that one model-type succeeds where the other fails. Indeed the physiological-linkage proposal has no direct empirical support. The more important point is that the two approaches contrast in what they take as plausible starting points for explanation. The mentalist or cognitive researcher is inclined to ascribe any systematicity in language behavior to mental computations performed on mental representations. The explanation specifically discounts (by neglect) any contribution to the form of the activity provided by special characteristics of the implementing system. In contrast, for the nonmentalist ecological researcher, a starting hypothesis ascribes as little as possible to computational-representational control and as much as possible to special characteristics of the implementing system. It seems to me that concurrent development of both explanatory strategies cannot help but be salutary to both.

A tutorial

Hammarberg brings out a number of points that he takes either to falsify my view of speech production or to demonstrate its incoherence. I will address a sample few. Others speak for themselves.

1.

> [H]ow can the coproduction model show . . . that in English "shoe" there is no coproduction of rounding (since š is always rounded in English) whereas in East Swedish "sju" (= seven) there would be coproduction (since Swedish š is not always rounded). (p. 128)

The coproduction view has the advantage of access to data (for example, Mann & Repp, 1980) showing that there *is* coarticulation of lip rounding in "shoe". This is revealed by acoustic analysis of /š/ produced in the environment of rounded and unrounded vowels (in Mann & Repp, /u/ and /a/). This finding does not contradict the observation that /š/ is normally rounded in English; it signifies only that owing to the vector-summation-like effects of coarticulation of lip rounding, some /š/s are more rounded than others. As for the more general point that coarticulatory effects may differ across languages, see footnote 3.

2.

> [I]t is obvious that a necessary condition on a coproduction model will be that segments that occur in the same time slot must not be incompatible with each other . . . . Now look at the coproductionist representation of "can", above, and consider what happens in the case of aspiration. Here the voicelessness of the consonant contradicts the voicing of the vowel. (p. 128)

As I suggested earlier, the serial ordering of segments is preserved in the order in which they predominate in articulation and in their effects on the acoustic signal. Production of an individual segment begins gradually, reaches a maximum during which time the quality of the acoustic signal is dominated by the properties of the segment over any others, and finally trails off, yielding dominance of the vocal tract and acoustic signal to a following segment. The onset of production of /æ/ in "can" begins during the closing phase of the /k/. Its own early phase consists largely of forward movement of the tongue body (which "vector sums" with tongue-body movement for the /k/). There is a crossover phase following release of the /k/ when format structure for the vowel begins to be apparent in the acoustic signal, but voicing has not yet begun. In this phase, the influence of /k/ in the signal is waning and that of the vowel is increasing. Seen in this way (*cf.* Fowler, 1983), there is nothing to prevent coproduction of segments having contradictory articulatory requirements. At any point in time, only some of the gestural correlates of a segment are being realized—more and more as the segment becomes predominant, fewer at its edges.

3. In the same example of "can":

> Next take the initial consonant in "can". This is a case where a feature spreading model would claim that the point of articulation of the consonant has been affected (in this case fronted) by the nature of the point of articulation of the vowel. But how can the coproduction model handle this? (p. 128)

As follows. The movement of the tongue body effecting closure for the consonant /k/ is, to a first approximation, a vector summation of two gestures. It is a weighted sum (weighted by the differences in predominance of the /k/ and /ae/) of the closing gesture for the /k/ and the fronting movement for the vowel. Neither gesture has been changed by the other (again, to a first approximation); they simply co-occur for a period of time.

This claim makes testable predictions, one of which I have tested in respect not to consonant-vowel coproduction, but to vowel-schwa coproduction (Fowler, 1981*a*). One

prediction is that listeners will ascribe the coarticulatory influence of one segment on another to the *influencing* segment, not the influenced segment. That is, they will extract a "vector-analyzed" parsing of the acoustic signal analogous to the one they extract from Johansson's point light display depicted in Fig. 1. During production of schwa, in a V́b__bV́ frame, the acoustic signal is influenced by properties of the reduced vowel itself, and by those of both flanking stressed vowels. Listeners' perceptual judgments indicate that they have extracted a parsing of the signal in which these influences are separated. Thus, acoustically distinct versions of schwa, which are distinct because they were produced in the conext of different flanking vowels, sound alike to listeners presented in their original vocalic frames, but different presented in isolation. More importantly, a given token of schwa spliced into different frames sounds different in the two contexts (for example, schwa produced and heard in ib__bi sounds like schwa; the same vowel heard in ab__ba sounds like /ɪ/), and it sounds more different from itself in these contexts than do acoustically different schwas heard in their original frames (Fowler, 1981*a*). The listeners treat the coarticulatory influences of the flanking vowels as if they belong to the flanking vowels and not to the schwa.

4. Hammarberg finds utterly absurd my efforts to incorporate the time dimension into the concept of segment. The reason he finds them absurd is that he considers only interpretations of the concept of time that are absurd in the present context. He considers two interpretations of the concept, neither corresponding to my intent (nor those of Lisker (e.g. 1972), who has been making this argument for many years). In one of Hammarberg's interpretations, time seems to refer to Heraclitus' view, leading to his claim that "You can't step twice into the same river". If time in this sense were an essential property of a segment, Hammarberg points out, one could never produce the same segment twice. His next idea is that, by time, I might have meant "duration", and somehow from that premise, arrives at the same conclusion as before. In any case, I didn't mean anything abstruse, but only that in view of the fact that speech essentially involves movement, it is unrealistic to suppose that the essential properties of phonetic segments are *states* of the vocal tract ("high", "rounded", for example). Lisker (1972) has discussed some segments that cannot be naturally defined in these terms.

5. "Fowler seems to believe that awareness of segments and the ability to report on the segmental composition of utterances is something completely unproblematic". (p. 130)

I don't actually:

> Skilled readers differ strikingly both from illiterate adults (Morais, Cary, Alegria & Bertelson, 1979) and from prereading children (e.g. Liberman, Shankweiler, Fischer & Carter, 1974) along a dimension that Mattingly (1972, 1979) has called linguistic awareness. Linguistic awareness refers to a language user's explicit (or accessible) knowledge of the structure or his or her language . . . . In contrast to skilled readers, illiterate adults and prereading children perform poorly on tasks that require access to the more atomic units of the language. Thus, for example, they find it difficult to count the syllables in a word, and many are unable to count the phonemes in a monosyllable—despite their proficiency in producing and perceiving these units under ordinary conditions. (Fowler, 1981*b*, p. 175–176)

Awareness of segments is required in order to understand the alphabetic principle and to exploit the properties of alphabets. It is in no way required for production of speech, however. The literature on errors in speech production shows convincingly that phonological

(not only phonetic) segments are, in some sense units of production; they shift as coherent units in errors, and adjust phonetically to their new context (see for example, Garrett, 1981). This, I presume, is no less true of illiterate than of literate talkers. If so, it must be possible to produce sequences of phonological segments without being conscious of doing so. If explicit awareness of segments is not required for production, presumably it is not required for perception either.

6. Hammarberg provides quotations from Fowler (1980*a*) in which the word "intent" is used to refer both to phonetic sequence that a talker utters and to a listener's perception of the (intended) phonetic sequence. He says:

> I was rather startled to read that communicative intent is a phonetic matter. Conventional wisdom could claim that what one intends to communicate is a semantic message. (p. 129)

First, if Hammarberg was startled to *read* of phonetic intents in my paper, one can only imagine how startled he was to *write* the following passages a few years ago:

> The rounding on the /s/ and /p/ in the word *spoon* is obviously intentional. It is not the result of accident or sloppy, inefficient articulation, and it would be absurd here to invoke mechano-inertial factors. Intentionality here would mean that the instructions to the articulators in this case would be to round the /s/ and the /p/. (1976, p. 360)

And later:

> Furthermore, if coarticulatory events are to be attributed to the physiological phonetic process, it must be the case that the input, the intent of the phonological process, differs from what actually is carried out by the phonetic process. Thus the intents of the phonological component would be defeated by the workings of the phonetic component. (p. 361)

Second I have no difficulty seeing intentions as tiered. Finally, Hammarberg and I are in good company:

> What is required, if the time-varying transitions are to be perceived (appropriately) as unitary segments, is that the percept reflect neither the proximal sound nor the more distal movements it betokens, but, rather, the still more distal, and presumably more nearly unitary, neural command structure that occasioned the movements. A less timid writer might call that the talker's *phonetic intent.* (Liberman, 1982, p. 152)

Preparation of the paper was supported by NSF Grant BNS 8111470 and NICHD Grant HD16591-01 to Haskins Laboratories.

References

Anderson, S. (1981). Why phonology isn't "natural". *Linguistic Inquiry*, **12**, 493–539.

Barry, W. & Kuenzel, H. (1975). Coarticulatory airflow characteristics of intervocalic voiceless plosives. *Journal of Phonetics*, **3**, 263–282.

Bergmann, G. (1967). *The Metaphysics of Logical Positivism.* Madison: University of Wisconsin Press.

Blackmore, J. (1979). On the inverted use of the terms "realism" and "idealism" among scientists and historians of science. *British Journal of the Philosophy of Science*, **30**, 125–134.

Breckinridge, J. (1977). *Declination as a Phonological Process.* Bell Laboratories Technical Memorandum. Murray Hill.

Bergman, A. (1977). *Asking the "What for" Question in Auditory Perception.* Paper presented at the Conference on Perceptual Organization. November 8–9, 1977, Washington, D.C.
Butcher, A. & Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics*, 4, 59–74.
Carney, P. & Moll, K. (1971). A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica*, 23, 193–202.
Chomsky, N. (1964). *Current Issues in Linguistic Theory.* The Hague: Mouton.
Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, Massachusetts: MIT Press.
Chomsky, N. (1980*a*). *Rules and Representations.* New York: Columbia University Press.
Chomsky, N. (1980*b*). On cognitive structures and their development: A reply to Piaget. In: *Language and Learning.* (M. Piatelli-Palmarini, ed.). Cambridge, Massachusetts: Harvard University Press.
Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English.* New York: Harper and Row.
Cohen, A., Collier, R. & t'Hart, J. (1983). Declination: construct or intrinsic feature of speech pitch? *Phonetica.*
Cooper, W. & Sorensen, J. (1981). *Fundamental Frequency in Sentence Production.* New York: Springer Verlag.
Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Quarterly Journal of Experimental Psychology*, **33A**, 185–207.
Darwin, C. J. & Bethell-Fox, C. E. (1977). Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 665–672.
Dennett, D. C. (1978). A cure for the common code? In: *Brainstorms.* (D. C. Dennett, ed.). Bradford Books.
Derwing, B. (1973). *Transformational Grammar as a Theory of Language Acquisition.* Cambridge: Cambridge University Press.
Fant, G. (1973). *Speech Sounds and Features.* Cambridge, Massachusetts: MIT Press.
Feyerabend, P. (1978). *Against Method.* London: Verso.
Fodor, J. (1975). *The Language of Thought.* New York: Thomas Y. Crowell.
Fodor, J. & Pylshyn, Z. (1981). How direct is visual perception? Some reflections on Gibson's 'ecological approach', *Cognition*, 9, 139–196.
Folkins, J. & Abbs, J. (1975). Lip and jaw motor control during speech: responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, 18, 207–220.
Fowler, C. (1977). *Timing Control in Speech Production.* Bloomington: Indiana University Linguistics Club.
Fowler, C. (1980*a*). Coarticulation and theories of extrinsic timing control. *Journal of Phonetics*, 8, 113–133.
Fowler, C. (1980*b*). A relationship between coarticulation and compensatory shortening. *Phonetica*, 38, 35–50.
Fowler, C. (1981*a*). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, 46, 127–139.
Fowler, C. (1981*b*). Some aspects of language perception by eye. In: *Perception of Print: Reading Research in Experimental Psychology.* (O. Tzeng & H. Singer, eds). Hillsdale, N.J.: Lawrence Erlbaum Associates.
Fowler, C. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllable stress feet. *Journal of Experimental Psychology: General.*
Fowler, C., Rubin, P., Remez, R. & Turvey, M. (1980). Implications for speech production of a general theory of action. In: *Language Production.* (B. Butterworth ed.). London: Academic Press.
Fowler, C. & Turvey, M. (1980). Immediate compensation for bite-block speech. *Phonetica*, 37, 306–326.
Fowler, C. & Turvey, M. (1982). Observational perspective and descriptive level in perceiving and acting. In: *Cognition and the Symbolic Processes,* II. (W. Weimer & D. Palermo, eds). Hillsdale, N.J.: Lawrence Erlbaum Associates.
Garrett, M. (1981). The limits of accommodation: Arguments for independent processing levels in sentence production. In: *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand.* (V. Fromkin, ed.). New York: Academic Press.
Gibson, J. (1966). *The Senses Considered as Perceptual Systems.* Boston: Houghton-Mifflin.
Gibson, J. (1979). *The Ecological Approach to Visual Perception.* Boston: Houghton-Mifflin.
Halle, M. (1959). *The Sound Pattern of Russian.* The Hague: Mouton.
Hammarberg, R. (1976). The metaphysics of coarticulation. *Journal of Phonetics*, 4, 353–363.
Hammarberg, R. (1982). On redefining coarticulation. *Journal of Phonetics*, 10, 123–137.
Henle, M. (1974). On naive realism. In: *Perception: Essays in honor of James J. Gibson.* (R. MacLeod & H. Pick, eds). Ithaca: Cornell University Press.
Hughes, O. & Abbs, J. (1976). Labial-mandibular coordination in the production of speech: Implications

for the operation of motor equivalence, *Phonetica*, 33, 199–221.
Johansson, G. (1950). *Configurations in Event Perception.* Uppsala: Almquist and Wiksell.
Johansson, G. (1974). Projective transformations as determining visual space perception. In: *Perception: Essays in honor of James J. Gibson.* (R. MacLeod & H. Pick, eds). Ithaca: Cornell University Press.
Katz, J. (1966). *The Philosophy of Language.* New York: Harper and Row.
Kelso, J. A. S., Tuller, B. & Fowler, C. (1982). The functional specificity of articulatory control and coordination. Paper presented at the meeting of the Acoustical Society of America, November 1982, Orlando Florida.
Kewley-Port, D. (1982). *Representations of Spectral Change as Cues to Place of Articulation in Stop Consonants.* PhD Dissertation, City University of New York.
Liberman, A. (1982). On finding that speech is special. *American Psychologist*, 37, 148–167.
Lindblom, B., Lubker, J. & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech-motor programming by predictive simulation. *Journal of Phonetics*, 7, 147–161.
Linell, P. (1979). *Psychology Reality in Phonology.* Cambridge: Cambridge University Press.
Lisker, L. (1972). On time and timing in speech. In: *Current Trends in Linguistics.* (T. Sebeok, ed.). The Hague: Mouton.
Mann, V. & Repp, B. (1980). Influence of vocalic context on perception of the /s/–/ʃ/ distinction. *Perception and Psychophysics*, 28, 213–229.
Medawar, P. (1974). A geometric model of reduction and emergence. In: *Readings in the Philosophy of Biology.* (F. Ayala & T. Dobzhansky, eds). Los Angeles: University of California Press.
Ohala, J. & Ewan, W. (1973). Speed of pitch change. *Journal of the Acoustical Society of America*, 53, 345A.
Ohman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 41, 310–320.
Parker, F. (1977). Distinctive features and acoustic cues. *Journal of the Acoustical Society of America*, 62, 1051–1054.
Pattee, H. (1970). The problem of biological hierarchy. In: *Towards a Theoretical Biology*, Vol. 3. (C. Waddington, ed.). Chicago: Aldine.
Popper, K. (1959). *The Logic of Scientific Discovery.* New York: Harper.
Putnam, H. (1973). Reductionism and the nature of psychology. *Cognition*, 2, 131-146.
Repp, B. (1981). On levels of discription in speech research. *Journal of the Acoustical Society of America*, 69, 1462–1464.
Rosch, E. & Lloyd, B. (1978). *Cognition and Categorization.* Hillsdale, N.J.: Lawrence Erlbaum Associates.
Ryle, G. (1949). *The Concept of Mind.* New York: Barnes and Noble.
Sampson, G. (1975). *Forms of Language.* London: Weidenfeld and Nicolson.
Shaw, R. & Bransford, J. (1977). Introduction: Psychological approaches to the problem of knowledge. In: *Perceiving, Acting and Knowing: Toward an Ecological Psychology.* (R. Shaw & J. Bransford, eds). Hillsdale, N.J.: Lawrence Erlbaum Associates.
Shaw, R., Turvey, M. & Mace, W. (1982). Ecological psychology: The consequence of a commitment to realism. In: *Cognition and the Symbolic Processes*, Vol. II. (W. Weimer & D. Palermo, eds). Hillsdale, N.J.: Lawrence Erlbaum Associates.
Simon, H. (1980). How to win at twenty questions with nature. In: *Perception and Production of Fluent Speech.* (R. Cole, ed). Hillsdale, N.J.: Lawrence Erlbaum Associates.
Smith, E. & Medin, D. (1981). *Categories and Concepts.* Cambridge, Massachusetts: Harvard University Press.
Sorenson, J. & Cooper, W. (1980). Syntactic coding of fundamental frequency in speech production. In: *Perception and Production of Fluent Speech.* (R. Cole, ed.). Hillsdale, N.J.: Lawrence Erlbaum Associates.
Steinberg, D. (1975). Chomsky: From formalism to mentalism to psychological invalidity. *Glossa*, 9, 218–252.
Stevens, K. & Blumstein, S. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 64, 1358–1368.
Sundberg, J. (1979). Maximum speech of pitch changes in singers and untrained subjects. *Journal of Phonetics*, 7, 71–79.
Turvey, M., Shaw, R., Reed, E. & Mace, W. (1981). Ecological laws for perceiving and acting: A reply to Fodor and Pylyshyn. *Cognition*, 10, 237–304.