SCIENTIFIC STUDIES OF READING, 3(2), 95–111

The Reading Researcher and the Reading Teacher Need the Right Theory of Speech

Alvin M. Liberman
Haskins Laboratories
New Haven, Connecticut

Speaking for Isabelle, I thank you for what she would have thought a kind and touching recognition of her devotion to the scientific study of reading. As for me, I am grateful for having been made part of this event but also embarrassed because I am related to it only by marriage. Having been wedded to Isabelle by choice, I was bound to reading by necessity. For I could not but be infected by her commitment to understand how children learn to read and why some cannot. Nor could I avoid being persuaded by her and Donald Shankweiler that my interest in speech was related to their interest in reading. After all, they appreciated early on that the connection between speech and reading is a two-way street and that one is well advised to look in both directions before proceeding. Thus, looking first toward speech, they observed in the work of their colleagues at Haskins Laboratories that the alphabetic structure of words is not to be found at the surface of the acoustic signal but only at a deeper, less accessible level. Then, looking in the other direction, toward reading, they foresaw the fateful consequence—namely, that mastery of speech does not normally make a child aware that words do, in fact, have an alphabetic structure. It would seem appropriate to this occasion that I follow the path they blazed. To that end, I promote the notion that only the right theory of speech can provide insight into the process by which a child who speaks is converted to one who also reads and

Requests for reprints should be sent to Alvin M. Liberman, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511–6695.

In 1998, the Society for the Scientific Study of Reading (SSSR) gave its award for Outstanding Scientific Contribution to the Study of Reading to Alvin M. Liberman, the late Isabelle Y. Liberman, and Donald Shankweiler. This article is based on Dr. Liberman's talk at the awards presentation at the annual meeting of SSSR in San Diego, April 1998.

writes. Seeing that process, as it is thus illuminated, should help the teacher to understand the relation between what her would-be reader already does and what more the teacher must now teach her to do.

The key to the teacher's understanding of that relation is found, I think, in five questions: a nucleus and four satellites. To appreciate the nucleus, we must look beyond the superficial differences between ear and eye, or tongue and hand, to ask how speech differs biologically from reading and writing and why it is more user-friendly. The four satellite questions arise from subordinate facts that are not always well or widely understood: Given that phonemic awareness is necessary for proper reading and writing, why is it not important for speech? Why does mastery of speech not yield phonemic awareness for free, and why is the cost of acquiring it so very high for some children? Why might preliterate experience with speech not only fail to produce phonemic awareness but also fix habits that make it harder to acquire? Why, as recent evidence shows, is the route to the lexicon always phonological, not visual, even for the skilled reader?

So far as I can tell, those questions are rarely answered because they are rarely asked. The reason, I think, is that they are far removed from the issues that are brought to the surface by the conventional theory of speech that is held explicitly by most speech scientists, and tacitly, it seems, by just about everybody else, including those who do research in reading. My aim, therefore, is twofold: (a) to show that the conventional view is implausible and likely, therefore, to mislead the reading researcher or teacher who follows it; and (b) to offer as a surer guide an alternative theory that is, in my view, as much more plausible as it is less conventional.

CONVENTIONAL (HORIZONTAL) THEORY OF SPEECH

The theory that most accept takes as its guiding assumption that consonants and vowels are managed by processes of production and perception that are not specialized for language but rather underlie the widest possible variety of behaviors, linguistic and nonlinguistic alike (Crowder, 1983; Diehl & Kluender, 1986; Fujisaki & Kawashima, 1970; Kuhl, 1981; Lindblom, 1991; Massaro, 1987; Stevens & Blumstein, 1978; Sussman, 1989; Tallal, Miller, & Fitch, 1993; Warren, 1993). On that view, speech is seen, rather meanly, to be a mere vehicle for language, not an integral part of it. When properly elaborated, the theory comprises four assumptions:

1. The elements of speech are sounds.
2. Those sounds are produced by motor processes of the most general sort, processes no different in any important way from those that wiggle our toes or twiddle our thumbs.

3. Just as production of speech is like any other kind of action, so is its reception like any other kind of hearing. Accordingly, speech is supposed to be governed by the general processes of audition, processes that occupy a common domain, employ a common mode of signal analysis, and evoke in a common register a common set of auditory primitives. It follows that percepts corresponding to a stop consonant and, say, a squeaking door differ only in the weights assigned to the auditory primitives they share; they are made of the same perceptual stuff. This third assumption is of particular interest to students of the reading process and is a proper companion to the second.

4. A fourth assumption is necessary to bridge the gap to language that is created by the second and third. According to those assumptions, the primary representations of *speaker* and *listener* are merely motor and merely auditory; neither is linguistic, hence both require to be made so, and that can be done only at a further remove, a cognitive stage beyond action and perception, where the secular motor and auditory representations are invested with phonetic privileges and thus made appropriate for language.

Given such conventional assumptions, a listener arrives at phonetic structure via initial processes of a broadly auditory sort, followed by a correspondingly broad cognitive process that affixes appropriate phonetic labels and then, presumably, does the heavy lifting that is required when a body manages something really hard, like syntax or rocket science. For convenience, I call the conventional theory "horizontal" to reflect its nature rather than its currency and to contrast it with the "vertical" theory I mean, in due course, to put in its place.

Why Is the Horizontal Theory Wrong for Speech and for Reading?

There are two ways to know that the horizontal theory is wrong: the hard way and the easy way. The hard way is to review a long history of experimentation that is rife with technical complications and vexed interpretations (e.g., Liberman, 1996). The easy way is to measure the theory against three very obvious considerations of a biological sort. The latter way, which I choose to follow here, requires only that one look at facts that are in plain view and then spend a few minutes thinking about what is seen.

Speech is easy. The first and most readily observed fact is simply the difference in biological status between speech and reading or writing. The former is vastly more natural, less likely to cause trouble than the latter. It is a serious mistake, I think, simply to take that difference for granted, which is what reading re-

searchers commonly do, for if we are to understand why reading is at least a little bit hard, we have got to know why speech is so very easy.

As for the relative difficulty of reading, it arises from the fact that the alphabetic characters and the visual percepts they evoke are not themselves pieces of language but only their arbitrary stand-ins. It follows that the visual percepts are useless for reading, or, indeed, for almost anything else, until they have been translated into the phonologic units for which they stand. (The possibility that they take the reader direct to meaning is considered, and dismissed, in sections that follow.) Yet the horizontal theory tells us that speech perception is in similar case because the sounds of speech evoke purely auditory percepts, which—like the visual percepts of the reader—have then to be cognitively converted into appropriate phonologic form. Given that implication of the horizontal theory of speech and the aforementioned fact about reading, we conclude that the primary representations of *reader* and *listener* are equally remote from language, hence equally in need of the intellectual exertions required to fit them for linguistic use. Accordingly, the horizontal theory of speech can make no sense of the difference in difficulty between the speech the child already commands and the skill in reading she will acquire. That failure is ample justification, indeed, for a fundamental assumption made by the avatars of Whole Language, which is that there is no important difference, from which they conclude that the child should be taught to read just as she was presumably taught to speak. I am confident that the reading researchers who tacitly accept the horizontal theory of speech never meant it to justify whole language, for surely they understand that speech is natural in a way that reading is not. I would rather make the more charitable assumption that the boost to whole language is an unintended consequence of thoughtlessly rooting a theory of reading in the wrong theory of speech.

The horizontal view is the worse confounded for our purposes because it implies that the advantage of ease must rest with reading and writing, not with speech. Consider, first, that print is by far the better signal. The printed character typically presents a clear contrast to its background, with edges that are sharp and clear. The linguistically relevant information in the speech signal, on the other hand, is rather badly smeared across the spectrum, and the acoustic features that are most important for many consonants are, often enough, among the least energetic parts of the signal. As for effectors, the moving finger writes and, having written, goes on to play the piano or knit a sweater. The moving tongue, for its part, speaks and, having spoken, relapses into inactivity, except as it is occasionally called on to help in chewing, swallowing, or doing something equally gross. Imagine attaching a pencil to your tongue and trying then to write your name. Finally, in the matter of the receptors, there is simply no contest: By any measure, the eye is a more accommodating channel than the ear. The upshot is that the horizontal theory makes us wonder about speech, not why it works so well and so easily by comparison with reading and writing, but how it manages to work at all.

Speech meets the parity requirement. The second consideration that a theory of speech must take into account derives from a requirement that is imposed on all communication systems, whether linguistic or not. That requirement is seemingly as obvious as it is fundamental, yet in all the vast literature on language and reading it figures in a scant few articles. On the assumption that a thing is more likely to be noticed if it has a name, Mattingly and I called it the "requirement for parity" and invited our colleagues to take it into account (Liberman, 1996; Liberman & Mattingly, 1985, 1989; see Rizzolatti & Arbib, 1998, for an account of the relevant neurobiology and its provocative linguistic implications). Although the response from psycholinguists and reading researchers has been one of the most intense indifference, my estimate of the importance of the parity requirement is in no way diminished, nor is my conception of its three facets in any way changed.

The first facet applies even to one-way communication. There, speaker and listener must be bound by a common understanding of what counts—what counts for the speaker must count for the listener. Thus, both must be of one mind about, say, *ba* and a sniff, agreeing that the one counts but not the other. According to the horizontal theory, however, both *ba* and the sniff evoke auditory percepts; the difference is only that one of the percepts is connected to a phonetic unit, the other not. But the phonetic unit is neither act nor percept, having the status, apparently, of an idea. How, then, was that idea established, how did it acquire a function in language, and how was it decided which auditory percept would be connected to it? Was that settled by our ancestors, perhaps, as they sat around a campfire one evening after a satisfying dinner of roast mastodon? Such an absurd possibility is no more than an instance of the more generally absurd conclusion to which the horizontal view leads, which is that the units of speech are not products of nature but artifacts, just like the letters of the alphabet.

The second facet applies to two-way communication, where speaker and listener exchange roles. There, the phonetic representation in the brain of the one party must be replicated in the brain of the other, else they cannot communicate in the same code. But the horizontal theory tells us that the one primary representation is generally motor, the other generally auditory; so the two are alike only in that neither has anything to do with language. Where, then, do the speakers find the common ground on which they must stand, and how was that ground reserved exclusively for communicative events? The horizontal theory must find those questions very hard, leaving us no alternative but to suppose that the ability of speakers to communicate in a common code must have been established by calculated agreement and then maintained across generations by deliberate instruction, exactly as has been done for reading and writing.

The third facet of the parity requirement takes into account that speech is a product of biological evolution, and that its development must have been by processes of a coevolutionary sort. Production and perception had to have marched through evolution in lock step, else the system would have come to a grinding halt.

Yet the horizontal view provides no connection between production and perception, except for that which is established by learning in the life history of the speaker. But unless inheritance is Lamarckian, which it is not, such a learned connection is of no use to evolution. Some might say further in the matter of evolution that speech wants comparison with sex because sex, too, must have been the product of coevolution. In the case of sex, however, it was disparity rather than parity that had to be preserved. Still, the comparison is appropriate because the assumption about the role of learning in the preservation of linguistic parity is no more implausible than the corresponding assumption about sexual disparity as it might have been learned in the lives of men and women. There is, I think, simply no way that a horizontal kind of theory can explain how speech could ever have evolved. We must suppose then that the development of speech was entirely cultural, not biological, just like the development of writing systems.

In sum, then, we look in vain to the horizontal theory for some insight into the difference that must exist between speech and reading or writing in how the parity requirement was met. The unhappy result is that we cannot take even the first step toward understanding the biological difference between learning to speak and learning to read or write.

Speech is biologically unique. The third consideration that theory must take into account is that speech is not just a product of biological evolution, but also species specific. Therefore, to know what we are trying to understand—whether speech or the use of its alphabetic transcription—we must first know what evolved. Not when, where, how, or by what progression from earlier-existing states. The first question is simply: what? What makes speech unique among natural communication systems, and what correspondingly unique requirements must it meet? At this point I should take into account that most teachers and other educated people seem to assume that speech is not biologically unique. Many animals have "language," they hold, which is to say that those animals "speak." It is true, of course, that all God's creatures communicate, exploiting for that purpose every possible medium. Thus, some use acoustical signals, some optical, some chemical, some mechanical, and just in case that does not exhaust the possibilities, electric fish do it electrically. But whatever the medium, all nonhuman communication systems have in common that every signal is meaningful in its own right; signals are never rearranged to convey new meanings. So, short of calling a convention of the animals and getting them to agree that some previously unused signal will henceforth be understood to have communicative significance, there is no way a nonhuman animal can convey a new message. All nonhuman systems of communication are closed.

Language is different in a profoundly important way, for language is open, capable of encoding an infinite number of meanings. That power, which is the very basis of our humanity and of the civilization it created, is given to language because it ad-

heres to what Abler (1989) called the particulate principle of self-diversifying systems. In all particulate systems, as Studdert-Kennedy (1998, in press) characterized them, discrete units taken from a small set are variously combined and permuted to form larger units that stand higher in a hierarchy and have properties different from those of the underlying constituents. It is, indeed, the particulate principle that is at work in chemical compounding, genetics, and language, serving in all three domains to achieve infinite ends with finite means. At the lowest level of language, where our interest lies, an indefinitely large number of words is formed by variously combining and permuting a small number of particles. Those words are then combined and permuted in turn to form sentences in which the meaning is not taken as the sum of the meanings of the individual words but rather derived, in significant measure, from the structural properties of the sentence. Such particulate communication imposes four requirements that must be met by speech. These should, in my view, be taken into account by all who would understand the reading process and the critical role that phonology plays in it.

The first requirement is that the fundamental particles be meaningless, for the particulate strategy cannot be used to good effect by combining and permuting only the small and fixed set of words that would otherwise be available. (Think how few words there would be if each one had to be a different sound and how awkward it would be to try to create new ones.) Indeed, the very basis of the uniqueness of language and of its uniquely valuable properties is that evolution for the first time created signals that were distinctly communicative without also assigning meanings to them. That is, evolution met the first facet of the parity requirement (see the earlier discussion) by marking a select class of signals in a way that identified them for the language user as having a communicative function and no other, but abstained from specifying just what it was that was being communicated. In all other forms of natural communication, signals that belong in a communicative modality have meaning as an unavoidable concomitant. A consequence for those nonhuman systems is that evolution had not only to set the communicative signals apart, and so meet the parity requirement, but also to establish the connection between signal and meaning and to do that independently for each signal–meaning pair. Barring further evolutionary development, vocabularies that are formed in that way are firmly and hopelessly closed. Surely Studdert-Kennedy (1998) was right when he observed that a critical step in the evolution of language was that which "bleached the signals of meaning" (p. 205).

The aforementioned considerations tell us that a *phonology*—a foundation of meaningless particles—is essential to language, as it is to a writing system. Without it, neither would be able to communicate the limitless number and variety of meanings of which language is, in fact, capable. The moral for the reading researcher and teacher is that the phonological underpinning of language holds the key to the reading process, as it does to speech (cf. Frost, 1998; Van Orden, 1991; Van Orden, Pennington, & Stone, 1990; Van Orden et al., 1992). For whatever else

a word may be, whether spoken or read, it is always a phonologic structure, and every phonologically legal structure is a candidate word. Consider, for example, the new word *nerd,* a phonologic structure immediately recognizable by every speaker of English as a possible word, the meaning of which such a speaker might have inferred on first hearing the sentence, "All work and no play makes Jack a nerd."

If the basic particles must be meaningless, then the relation between the phonologic (i.e., particulate) structure of a word and its meaning cannot be other than wholly arbitrary. Hence, reading the word and knowing its meaning (or, more commonly, its family of possible meanings) are separate matters. Moreover, the dependence of the one on the other is obviously asymmetric for the reader. It is not possible to know the meaning of a text without getting the words—that is, the phonologic structures—but it is easy enough to get the words and yet miss the meaning. Indeed, I do that every time I read an article in a scientific field about which I know nothing. Given that I am able to read the words rapidly and easily enough that my attentional resources can be applied to those aspects of meaning that depend on sentence structure, I will come to understand the article, not by becoming a better reader, but by overcoming my ignorance of its subject. I here labor the obvious, only because the conflation of reading and comprehension, so common among reading experts, makes it hard to understand either process or to know how to improve it. Word meanings can be made available to readers (or speakers) only by attaching them appropriately to the phonologic structures that are dedicated to conveying them; there is no other way. Moreover, given that most words have a complex of meanings and that some meanings are purely grammatical—that is, without specifiable referents—the phonology is not only the necessary basis for openness, as argued previously, but also the only dependable and secure anchor for the speaker or reader who would avoid being tossed about in an uncertain semantic sea.

It does not follow from any of this that failure to engage the phonologic structure makes reading impossible, just that it makes possible only a little bit of reading. For, given no phonologic foundation, the irreducible elements of the process are meaningful, hence serviceable only as the basis for a tightly closed system. One might suppose, then, that a reader who fails to recover the phonology would read as a chimpanzee might. That follows on the assumption that creatures whose natural communication system does not have a phonology cannot make use of a system that does.

The second requirement derives from the first, for if meaning is not given immediately by the signal, as it is in all natural but nonhuman communication, then it must depend on structures of an intermediate sort. Those structures—phonemes, morphemes, words, phrases, and sentences—form the hierarchical edifice that goes beyond the indefinitely large number of words made possible by the phonology at the bottom, reaching to the indefinitely large number of propositions made possible by syntax at the top. As Studdert-Kennedy (1998, in press) pointed out, they also give

language freedom from control by stimuli that are immediately present, allowing language users to talk, write, or simply muse not only about who is doing what to whom but also who might have done it had he not lost his nerve, and to manage that in the absence of all three elements: the who, the what, and the whom.

There are in the foregoing arguments several implications for the reading researcher and teacher. One is that language understanding cannot be achieved by going direct from print to meaning, as some methods of reading instruction would have children do. Such a procedure is appropriate for the aforementioned chimpanzees, perhaps, or for preliterate children who find it useful to "read" the McDonald's™ sign, but not for anyone whose reading would tap the vast resources of language. Another, and related, implication arises from the fact that the critical intermediate structures are specifically, distinctly, and exclusively linguistic. Accordingly, there is no such thing as an "auditory" speech perceiver or a "visual" reader, for, as I have already said several times, the meaning of an utterance is not to be had directly from the auditory or visual representations of its words; rather, it is necessary to access the distinctly linguistic constituents, among which are the phonologic structures to which the various meanings of the words, together with their grammatical privileges, are connected. Given the ready availability and stability of the tens of thousands of those structures and their connections, it is hard to see why a reader would take pains to create wholly new ones just so he could make do with representations that are visual instead of phonologic. Some researchers have argued that readers do that, at least for familiar words, to gain faster access to meaning (Coltheart, 1978). That argument is apparently based on the assumption that first converting the print to phonology is time-consuming, the more so, perhaps, if there is, as the horizontal view would suggest, the need to hear the word in the mind's ear and so to recover the auditory representation to which the phonology is presumably connected. However, there is an abundance of recent and, I think, convincing evidence that the phonologic route is, in fact, the faster, occurring within 15 msec at the outside; a spelling check that would distinguish *damn* from *dam,* for example, comes later, if at all (Lukatela, Frost, & Turvey, in press-a, in press-b; Lukatela & Turvey, 1994a, 1994b; Perfetti & Bell, 1991; Perfetti, Bell, & Delaney, 1988; Perfetti, Zhang, & Berent, 1992).

Questions about the claimed speed advantage of a visual route aside, readers cannot in any case make do indefinitely with the initial visual percepts but must, as I have already pointed out, at some point convert to the specialized phonologic, morphologic, and syntactic structures that language processing requires. It seems eminently plausible, therefore, to suppose that if readers are to get all that language has to give, they had best convert sooner, into the phonology, rather than later, into its higher ranking cousins. The newfound evidence that skilled readers do exactly that helps to explain why they are skilled. I should think it a wise policy for teachers to bear that in mind and to design teaching methods accordingly.

The third requirement of particulate communication is that the elements be commutable, from which it follows that they must be discrete, invariant, and categorical. They must not be scalar, equivocal, or ambiguous, as meanings so commonly are. That is an absolute requirement, not to be compromised in any way.

The fourth and last requirement that is of interest to us has to do with the rate at which the particles must be produced and perceived. To appreciate the importance of rate, consider that, given the nature of the vocal tract and the ear, the particles must be laid down in strings and, given the limitations of the vocal tract, the number of particles must be small. A consequence is that the strings commonly run to considerable lengths. But if the particles are to be organized into the units that stand higher in the hierarchy, they must be held in a buffer, the working memory component of the syntactic processor that has little patience with information that comes in too slowly. There is ample evidence that the working memory is best served by phonetic structures, not by visual representations—still another reason why the primary visual responses of the reader must be converted to phonetic units.

Now, to see the problem the horizontal view has with those requirements, consider again its fundamental assumption, an assumption made by virtually every reading researcher and reading teacher: The elements of speech are sounds. If that were the correct assumption, however, the essentially particulate nature of speech would be manifest at the acoustic surface, in which case speech would be an acoustic alphabet. But as I said at the very beginning of this article, it was knowing that assumption to be wrong that led Isabelle and Donald to their pioneering work on phonemic awareness. They knew it was wrong because of experimental work on the speech signal and its perception that had been done by their colleagues at Haskins Laboratories. It is enough for our purposes, however, simply to appreciate that it is wrong if only because, given the need for invariant, discrete, and categorical particles, it would not be possible with an acoustic alphabet to say *bag,* but only something like *buh–a–guh,* and to say that is not to speak but to spell. To know what perception would be like in that case, imagine listening to a Victorian novel as it is spelled to you, letter by painful letter. Clearly, an acoustic alphabet could not meet the rate requirement that particulate communication imposes.

In fact, the problem that would be created by an acoustic alphabet is not easily solved. To see why, consider that as I speak I produce phonetic segments at a rate of about 12 per second on average with short stretches that may go to 20. Now if each phonetic segment were an acoustic segment, rates that high would overreach the temporal resolving power of the ear and also its ability to apprehend the order of the segments. The listener would hear an approximation to a buzz and could not, in any case, discover how to order the segments so as to spell *pat, tap,* and *apt.*

So to tell a child that the word *bag* consists of three sounds that follow each other in close temporal succession is to tell a lie. It is a white lie, I suppose, much like the one about babies and the stork, and it is a better lie than those that would have the child believe that the word is only a meaning or a picture, or that it does

not really matter exactly what it is so long as the reader guesses it from the illustrations or the context and manages every once in a while to get into the right semantic field—reading *bag* as *sack,* for example. Good readers commonly and properly do infer the meaning from the context but not the word, which is explicitly there, after all, in starkest black and white. Unfortunately for beginners, however, what is there in black and white are discrete letters, yet the word is not formed of correspondingly discrete sounds. Still, we are sure that words do come apart into the meaningless pieces that the discrete letters represent, for if they did not, alphabetic transcription would not be possible. It remains, then, for the right theory to resolve this troubling matter, and so to acquaint the teacher with the truth about the alphabetic principle.

So How Did We Get Into This Mess?

To explain the wide acceptance of the horizontal view without being invidious, I can offer only one hypothesis: The horizontal view is consistent with an ancient tradition, still honored by many students of speech, that equates primary perceptual representations with end organs. We have ears, therefore we have primary auditory percepts; and we have eyes, therefore primary visual percepts. But language has no end organs of its own, so it must presumably rely, in speech, on the primary auditory representations that are evoked by the processes that happen to own the necessary peripheral equipment. People are therefore attached to the horizontal view, I think, simply because they find it hard to imagine an alternative. Consider, for example, the history of research on visual perception of depth. Berkeley (1709) anticipated the horizontal view of speech when he pointed out, implicitly and correctly, that there is no end organ for depth, and so concluded, explicitly and incorrectly, that perception of depth must be secondary, a learned translation from the sensations of strain that are produced in the extraocular muscles as they differentially converge to see objects at different distances. Over a century later, Wheatstone (1838) invented the stereoscope and so put the matter in a very different light. The stereoscope is best seen as a device for synthesizing perceived depth, and like the speech synthesizers that came along after another 125 years, it made possible the experiments that identified the essential cue for the percept being investigated. That cue turned out to be, not sensations of strain from the extraocular muscles, but binocular disparity—that is, the different projections onto the retinas that arise by geometric necessity from the fact that the two eyes have, in the most literal sense, two points of view. What is interesting for us is that vision scientists did not then assume, with Berkeley and with the latter-day horizontalists in speech research, that just because depth has no end organ, its perception must necessarily be a two-stage process: a first stage in which viewers perceive the disparity as such and then a second in which they translate the disparity into depth. There has for a very long time

been agreement that the primary percept is depth and that it is made available to the viewer by a neural system specialized to compute information about binocular disparity and then to represent that information immediately—that is, without intermediating double vision—as depth. That is not an idle digression, because I would claim that phonetic percepts are primary in much the same way that depth is and for much the same kind of reason (Liberman & Mattingly, 1985; Mattingly & Liberman, 1988; Xu, Liberman, & Whalen, 1997)

AN UNCONVENTIONAL (VERTICAL) THEORY OF SPEECH

The guiding assumption of the vertical view is that perception of phonetic structure is managed by the phonetic component of a biologically coherent system, specialized from top to bottom for particulate communication. It follows that there is a distinct mode in which phonetic representations are primary percepts, not translations from initial representations of a common auditory sort. The vertical view rests on four assumptions (Liberman, 1996; Liberman & Mattingly, 1985; Mattingly & Liberman, 1988).

The first assumption is that the fundamental particles of speech are not sounds but gestures of the articulatory organs. A *gesture* is a change in the cavities of the vocal tract—an opening or closing, lengthening or shortening, widening or narrowing. Those gestures, which are controlled by neuromotor units like those of Turvey's (1990) coordinative structures, are the discrete, invariant, and categorical elements of speech, the primitives that must be exchanged between speakers if linguistic communication is to occur.

The second assumption is that the elemental gestures evolved with language, specifically in the service of a phonetic function. Hence, they are phonetic ab initio, requiring no cognitive translation to make them so. With the evolution of those gestures, there was also, presumably, the particulation of the vocal tract that underlies the ability, unique to human beings, to control the gestures independently (Studdert-Kennedy, 1998, in press). Selection pressure in evolution was for those gestures that lent themselves to being interleaved, overlapped, and merged—that is, coarticulated. The biologically critical payoff was fast phonetic action from slow-moving machinery (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

The third assumption is that the specifically phonetic gestures are controlled by a biological specialization—a component of the larger specialization for language—that Mattingly and I (1988) called the phonetic module. The particular task of that module is to manage the coarticulation so as to maximize interleaving and overlapping of the gestures while precluding those temporal relations among

them that would cause the acoustic consequences of the one or the other to be obscured.

Thus we see how, in production, the rate requirement of particulate communication is met. But the vertical view has happy consequences for perception, too, for coarticulation folds into a single piece of sound information about several successive phonetic segments, thus relaxing by a large factor the constraint on rate of perception that is imposed by the temporal resolving power of the ear. Moreover, coarticulation clearly marks the order of segments by the shape of the acoustic signal, thus making it unnecessary for the ear to do what it does badly, which is to judge the order of discrete sounds that follow each other in close temporal succession.

The aforementioned advantages are crucial, but they do not come without cost, for coarticulation destroys any correspondence in segmentation between signal and message, which is, of course, exactly why speech is a complex code, not a straightforward cipher or alphabet. So how does the listener cope with that complexity? Not to worry; the phonetic module has within it everything needed to constrain an analysis-by-synthesis that recovers from the acoustic signal the coarticulated gestures that caused it. It is just those gestures that constitute the phonetic segments. Thus, perception of speech is easy, not because the process is simple, but because the phonetic module is so well adapted to its complex task.

There is, finally, the fourth assumption of the vertical view, which is implied by the first three but important enough in its own right to deserve explicit statement: The listener perceives the phonetic gestures that the speaker produced.

Having seen how, according to the vertical theory, speech is able to meet the requirement that discrete, invariant, and categorical particles be produced and perceived at a high rate, we look to the theory for a plausible account of how speech meets the parity requirement. What we see is that parity is built into the very bones of the system. As I have already said and now say again, the elemental gestures evolved with language in the service of a specific linguistic function. Hence, they form a natural class, phonetic by its very nature. No agent need ever have decided what is or is not phonetic, or which motor and auditory representations were to be connected to which phonetic units. Speakers and listeners conduct their linguistic business in the common coin of phonetically significant gestures; there is no need for cognitive translation to get the communicating parties into the language system, or for arbitrary agreements to establish a common code between them. Moreover, coevolution of speech is guaranteed because production and perception are but opposite sides of the same coin; a change in one is automatically a change in the other.

BACK TO THE QUESTIONS ABOUT READING

Speech production is easy because the phonetic module does all the hard work, selecting and coordinating the relevant phonetic gestures. The speakers have only to

think of the word they want to say; they need not be able to spell the word unless they want to write it down, in which case the phonetic module is struck dumb, forcing them to fall back on their conscious awareness of phonologic structure. The listener also relies on the module for immediate representation of units that are phonetic by nature, hence perfectly suited for processing by the other components of the language system. Unfortunately for the reader, the module devotes itself exclusively to acoustic signals that contain phonetic information. It does not respond to letters of the alphabet, no matter what information they contain. When reading an alphabetic transcription, therefore, the reader is left to her own explicit knowledge of phonology to recover the language that was transcribed.

Phonemic awareness is not important for speech because there is never a need at any stage of language development to arrive at phonetic structure by consciously translating representations that are initially merely motor or auditory. Why would one be aware of what would be required only for that which does not happen?

Phonemic awareness does not result from learning to speak because the primary representations of the phonetic module are already perfectly suited for the other processes of the language specialization. Those primary phonetic representations do not get attention because they do not need it, and they do not need it because, unlike the visual representations of the letters or the auditory representations assumed by the horizontal theory, they need not be converted into something other than what they already are. It is also relevant, as Isabelle and Donald noted many years ago, that the nonalphabetic nature of the sounds makes it that much harder to become aware that they do, nevertheless, convey an alphabetically coded message.

The module that underlies phonetic perception may in some would-be readers work less efficiently than normal, in which case the phonetic representations it produces are less than ideally clear, hence harder to bring to awareness. Such representations also serve less well in working memory, so processing at the syntactic level is likely impaired. The result is difficulty in learning to read. Such disabled readers also have some difficulty in speech discrimination, not because their auditory systems cannot cope with the important acoustic cues for phonetic elements, as Tallal et al. (1993) claimed, but because an imperfect phonetic module is not properly engaged by those cues. Perception of the same acoustic cues in a nonphonetic context may be, and apparently is, perfectly normal (Mody, Studdert-Kennedy, & Brady, 1997).

Because preliterate children need pay no attention to the phonetic structure, they are free to devote all of it to getting the meaning, which is, after all, what the child is really after. Years of experience in doing that are calculated to reinforce the tendency to look for meaning and for nothing else. The result, as Byrne (1996) showed, is that preliterate children readily learn to read subword structures when they are morphemic, as in *run* versus *runner,* for example, but not when they have only a phonologic function, as in *corn* versus *corner.*

The reason readers follow the phonologic route to the lexicon takes us back to my earlier, rather long-winded consideration of the particulate principle and its application in language. Briefly stated, the point is that proceeding from a base of meaningless particles—that is, a phonology—is a necessary condition for the ability of language, whether spoken or written, to encode an indefinitely large number of meanings. (See DeFrancis, 1991, for support of the notion that true writing systems always have a significant place for such phonologic units as phonemes or syllables.) Going from print directly to meaning is therefore to enter a dead end. As for communicating the sentence, syntactic processes must apparently rely on units that are phonetic, not visual, for temporary storage in the buffer where the structural properties of the sentence are apprehended. Finally, we should take into account that the syntactic component of the larger language module must have evolved to work in harmony with phonologic structures; presumably, therefore, it would not know what to do with representations of a visual sort.

CODA

If Isabelle had read this article, she would have reacted with characteristic incisiveness to reduce my extended remarks to the seemingly obvious claim that speech is language, whereas the alphabet only refers to it. Put that way, she would then have added that my claim would not pass the "grandmother test." By that, she would have meant that Grandma could have been expected to respond, contemptuously, "So what else is new?" or, more testily, "You mean that boy went to college for 8 years just to learn that?" I might then have reminded Grandma that I met Isabelle at college, which would have brought forth the firm declaration that my attendance there was the best thing I ever did. It is possible that I might even have taken advantage of Grandma's now mellower mood to wring from her the concession that it is one thing to state the obvious, quite another to explain it. In that case, she and Isabelle might have agreed that explaining the obvious is what I have been about in this article and, further, that if my explanation is anywhere near as plausible as I think, it might help to head the reading researcher and teacher toward a better understanding of the alphabetic principle.

ACKNOWLEDGMENTS

Research on speech and its relation to reading has been supported at Haskins Laboratories for many years by the National Institute of Child Health and Human Development. I could not have written this article but for that research and that support.

For helpful comments about an earlier version of this article, I am indebted to Marilyn Adams, Carol Fowler, Margaret Hemphill, Mark Liberman, Ruth Millikan, Donald Shankweiler, Michael Studdert-Kennedy, and Michael Turvey.

REFERENCES

Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures, 12,* 1–13.

Berkeley, G. (1709). *An essay toward a new theory of vision.* Dublin: Aaron Rhames.

Byrne, B. (1996). The learnability of the alphabetic principle: Children's initial hypotheses about how print represents spoken language. *Applied Psycholinguistics, 17,* 401–426.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). London: Academic.

Crowder, R. G. (1983). The purity of auditory memory. *Philosophical Transactions of the Royal Society, B, 302,* 251–265.

DeFrancis, J. (1991). *Visible speech: The diverse oneness of writing systems.* Honolulu: University of Hawaii Press.

Diehl, R., & Kluender, K. (1986). On the objects of speech perception. *Ecological Psychology, 1,* 121–144.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123*(1), 71–95.

Fujisaki, H., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, 29,* 207–214.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech–sound categories. *Journal of the Acoustical Society of America, 70,* 340–349.

Liberman, A. M. (1996). *Speech: A special code.* Cambridge, MA: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461.

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

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

Lindblom, B. (1991). The status of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 7–24). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lukatela, G., Frost, S. J., & Turvey, M. T. (in press-a). Identity priming in English is compromised by phonological ambiguity. *Journal of Experimental Psychology: Human Perception and Performance.*

Lukatela, G., Frost, S. J., & Turvey, M. T. (in press-b). Phonological priming by masked nonword primes in the Lexical Decision Task. *Journal of Memory and Language.*

Lukatela, G., & Turvey, M. T. (1994a). Visual lexical access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 107–128.

Lukatela, G., & Turvey, M. T. (1994b). Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 331–353.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological enquiry.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Mattingly, I. G., & Liberman, A. M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Functions of the auditory system* (pp. 775–793). New York: Wiley.

Mody, M., Studdert-Kennedy, M., & Brady, S. A. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology, 64,* 199–231.

Perfetti, C. A., & Bell, L. C. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and masked priming. *Journal of Memory and Language, 30,* 473–485.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27,* 59–70.

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a universal phonological principle. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 61–75). Amsterdam: North-Holland.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience, 21,* 188–194.

Stevens, K. N., & Blumstein, S. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64,* 1358–1368.

Studdert-Kennedy, M. (1998). The particulate origin of language generativity: From syllable to gesture. In J. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language.* Cambridge, England: Cambridge University Press.

Studdert-Kennedy, M. (in press). Implications of the particulate principle for the evolution of language: The dissociation of phonetic from semantic function. In C. Knight, J. Hurford, & M. Studdert-Kennedy (Eds.), *Approaches to the evolution of language* (Vol. 2). Cambridge, England: Cambridge University Press.

Sussman, H. (1989). Neural coding of relational invariance in speech: Human language analogs to the barn owl. *Psychological Review, 96,* 631–642.

Tallal, P., Miller, F., & Fitch, R. H. (1993). Neurobiological basis of speech: A case for the preeminence of temporal processing. In P. Tallal, A. M. Gallaburda, R. R. Llinas, & C. von Euler (Eds.), *Annals of the New York Academy of Sciences. Vol. 82. Temporal information processing in the nervous system* (pp. 27–47). New York: New York Academy of Sciences.

Turvey, M. T. (1990). Coordination. *American Psychologist, 45,* 938–953.

Van Orden, G. (1991). Phonologic mediation is fundamental to reading. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 77–103). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Van Orden, G., Pennington, B. F., & Stone, G. O. (1990). Word identification and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488–522.

Van Orden, G., Stone, G. O., Garlington, K. L., Markson, L. R., Pinnt, G. S., Simonfly, C. M., & Brichetto, T. (1992). "Assembled" phonology and reading: A case study in how theoretical perspective shapes empirical investigation. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 40–59). Amsterdam: North-Holland.

Warren, R. M. (1993). Perception of acoustic sequences: Global integration versus temporal resolution. In S. McAdams & E. Bigand (Eds.), *Thinking in sound.* Oxford, England: Oxford University Press.

Wheatstone, C. (1838). On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions,* Part I.

Xu, Y., Liberman, A. M., & Whalen, D. H. (1997). On the immediacy of phonetic perception. *Psychological Science, 8,* 358–362.

Manuscript received July 20, 1998
Accepted July 23, 1998