

CHAPTER 9

The Relation of Speech to Reading and Writing

Alvin M. Liberman
Haskins Laboratories

Theories of reading/writing and theories of speech typically have in common that neither takes proper account of an obvious fact about language that must, in any reckoning, be critically relevant to both: there is a vast difference in naturalness (hence ease of use) between its spoken and written forms. In my view, a theory of reading should begin with this fact, but only after a theory of speech has explained it.

My aim, then, is to say how well the difference in naturalness is illuminated by each of two theories of speech—one conventional, the other less so—and then, in that light, to weigh the contribution that each of these can make to an understanding of reading and writing and the difficulties that attend them. More broadly, I aim to promote the notion that a theory of speech and a theory of reading/writing are inseparable, and that the validity of the one is measured, in no small part, by its fit to the other.

What does it mean to say that speech is more natural?

The difference in naturalness between the spoken and written forms of language is patent, so I run the risk of being tedious if I elaborate it here. Still, it is important for the argument I mean to make that we have explicitly in mind how variously the difference manifests itself. Let me, therefore, count the ways.

(1) Speech is universal. Every community of human beings has a fully developed spoken language. Reading and writing, on the other hand, are relatively rare. Many, perhaps most, languages do not even have a written form, and when, as in modern times, a writing system is devised—usually by missionaries—it does not readily come into common use.

(2) Speech is older in the history of our species. Indeed, it is presumably as old as we are, having emerged with us as perhaps the most important of our species-typical characteristics. Writing systems, on the other hand, are developments of the last few thousand years.

(3) Speech is earlier in the history of the individual; reading/writing come later, if at all.

(4) Speech must, of course, be learned, but it need not be taught. For learning to speak, the necessary and sufficient conditions are but two: membership in the human race and exposure to a mother tongue. Indeed, given that these two conditions are met, there is scarcely any way that the development of speech can be prevented. Thus, learning to speak is a precognitive process, much like learning to perceive visual depth and distance or the location of sound. In contrast, reading and writing require to be taught, though, given the right ability, motivation, and opportunity, some will infer the relation of script

to language and thus teach themselves. But, however learned, reading/writing is an intellectual achievement in a way that learning to speak is not.

(5) There are brain mechanisms that evolved with language and that are, accordingly, largely dedicated to its processes. Reading and writing presumably engage at least some of these mechanisms, but they must also exploit others that evolved to serve nonlinguistic functions. There is no specialization for reading/writing as such.

(6) Spoken language has the critically important property of 'openness': unlike nonhuman systems of communication, speech is capable of expressing and conveying an indefinitely numerous variety of messages. A script can share this property, but only to the extent that it somehow transcribes its spoken-language base. Having no independent existence, a proper (open) script is narrowly constrained by the nature of its spoken-language roots and by the mental resources on which they draw. Still, within these constraints, scripts are more variable than speech.

One dimension of variation is the level at which the message is represented, though the range of that variation is, in fact, much narrower than the variety of possible written forms would suggest. Thus, as DeFrancis (1989) convincingly argues, any script that communicates meanings or ideas directly, as in ideograms, for example, is doomed to arrive at a dead end. Ideographic scripts cannot be open—that is, they cannot generate novel messages—and the number of messages they can convey is never more than the inventory of one-to-one associations between (holistically different) signals and distinctly different meanings that human beings can master. Indeed, it is a distinguishing characteristic of language, and a necessary condition of its openness, that it communicates meanings indirectly, via specifically linguistic structures and processes, including, nontrivially, those of the phonological component. Not surprisingly, scripts must follow suit; in the matter of language, as with so many other natural processes, it is hard to improve on nature.

Constraints of a different kind apply at the lower levels. Thus, the acoustic signal, as represented visually by a spectrogram, for example, cannot serve as a basis for a script; while spectrograms can be puzzled out by experts, they, along with other visual representations, cannot be read fluently. The reason is not primarily that the relevant parts of the signal are insufficiently visible; it is, rather, that, owing to the nature of speech, and especially to the coarticulation that is central to it, the relation between acoustic signal and message is complex in ways that defeat whatever cognitive processes the 'reader' brings to bear. Narrow phonetic transcriptions are easier to read, but there is still more context-, rate-, and speaker-conditioned variation than the eye is comfortable with. In any case, no extant script offers language at a narrow phonetic level. To be usable, scripts must, apparently, be pitched at the more abstract phonological and morphophonological levels. That being so, and given that reading-writing require conscious awareness of the units represented by the script, we can infer that people can become conscious of phonemes and morphophonemes. We can also infer about these units that, standing above so much of the acoustic and phonetic variability, they correspond approximately to the invariant forms in which words are presumably stored in the speaker's lexicon. A script that captures this invariance is surely off to a good start. At all events, some scripts (e.g., Finnish, Serbo-Croatian) do approximate to purely phonological renditions of the language, while others depart from a phonological base in the direction of morphology. Thus, English script is rather highly morphophonological, Chinese even more so. But, as DeFrancis (1989; see also Wang, 1981) makes abundantly clear, all these scripts, including even the Chinese,

are significantly phonological, and, in his view, they would fail if they were not; the variation is simply in the degree to which some of the morphology is also represented.

Scripts also vary somewhat, as speech does not, in the size of the linguistic segments they take as their elements, but here, too, the choice is quite constrained. Surely, it would not do to make a unit of the script equal to a phoneme and a half, a third of a syllable, or some arbitrary stretch—say 100 milliseconds—of the speech stream. Still, scripts can and do take as their irreducible units either phonemes or syllables, so in this respect, too, they are more diverse than speech.

(7) All of the foregoing differences are, of course, merely reflections of one underlying circumstance—namely, that speech is a product of biological evolution, while writing systems are artifacts. Indeed, an alphabet—the writing system that is of most immediate concern to us—is a triumph of applied biology, part discovery, part invention. The discovery—surely one of the most momentous of all time—was that words do not differ from each other holistically, but rather by the particular arrangement of a small inventory of the meaningless units they comprise. The invention was simply the notion that if each of these units were to be represented by a distinctive optical shape, then everyone could read and write, 'provided he knew the language and was conscious of the internal phonological structure of its words.

How is the difference in naturalness to be understood?

Having seen in how far speech is more natural than reading/writing, we should look first for a simple explanation, one that is to be seen in the surface appearance of the two processes. But when we search there, we are led to conclude, in defiance of the most obvious facts, that the advantage must lie with reading/writing, not with speech. Thus, it is the eye, not the ear, that is the better receptor; the hand, not the tongue, that is the more versatile effector; the print, not the sound, that offers the better signal-to-noise ratio; and the discrete alphabetic characters, not the nearly continuous and elaborately context-conditioned acoustic signal, that offers the more straightforward relation to the language. To resolve this seeming paradox and understand the issue more clearly, we shall have to look more deeply into the biology of speech. To that end, I turn to two views of speech to see what each has to offer.

The conventional view of speech as a basis for understanding the difference in naturalness. The first assumption of the conventional view is so much taken for granted that it is rarely made explicit. It is, very simply, that the phonetic elements are defined as sounds. This is not merely to say the obvious, which is that speech is conveyed by an acoustic medium, but rather to suppose, in a phrase made famous by Marshall McLuhan, that the medium *is* the message.

The second assumption, which concerns the production of these sounds, is also usually unspoken, not just because it is taken for granted, though it surely is, but also because it is apparently not thought by conventional theorists to be even relevant. But, whatever the reason, one finds among the conventional claims none which implies the existence of a phonetic mode of action—that is, a mode adapted to phonetic purposes and no other. One therefore infers that the conventional view must hold (by default, as it were) that no such mode exists. Put affirmatively, the conventional assumption is that speech is produced by motor processes and movements that are independent of language.

The third assumption concerns the perception of speech sounds, and, unlike the first two, is made explicitly and at great length (Cole & Scott, 1974; Crowder & Morton, 1969;

Diehl & Kluender, 1989; Fujisaki & Kawashima, 1970; Kuhl, 1981; Miller, 1977; Oden & Massaro, 1978; Stevens, 1975). In its simplest form, it is that perception of speech is not different from perception of other sounds; all are governed by the same general processes of the auditory system. Thus, language simply accepts representations made available to it by perceptual processes that are generally auditory, not specifically linguistic. So, just as language presumably recruits ordinary motor processes for its own purposes, so, too, does it recruit the ordinary processes of auditory perception; at the level of perception, as well as action, there is, on the conventional view, no specialization for language.

The fourth assumption is required by the second and third. For if the acts and percepts of speech are not, by their nature, specifically phonetic, they must necessarily be made so, and that can be done only by a process of cognitive translation. Presumably, that is why conventional theorists say about speech perception that after the listener has apprehended the auditory representation he must elevate it to linguistic status by attaching a phonetic label (Crowder & Morton, 1969; Fujisaki & Kawashima, 1970; Pisoni, 1973), fitting it to a phonetic prototype (Massaro, 1987; Oden & Massaro, 1978), or associating it with some other linguistically significant entity, such as a 'distinctive feature' (Stevens, 1975).

I note, parenthetically, that this conventional way of thinking about speech is heir to two related traditions in the psychology of perception. One, which traces its origins to Aristotle's enumeration of the five senses, requires of a perceptual mode that it have an end organ specifically devoted to its interests. Thus, ears yield an auditory mode; eyes, a visual mode; the nose, an olfactory mode; and so on. Lacking an end organ of its very own, speech cannot, therefore, be a mode. In that case, phonetic percepts cannot be the immediate objects of perception; they can only be perceived secondarily, as the result of a cognitive association between a primary auditory representation appropriate to the acoustic stimulus that excites the ear (and hence the auditory mode) and, on the other hand, some cognitive form of a linguistic unit. Such an assumption is, of course, perfectly consistent with another tradition in psychology, one that goes back at least to the beginning of the 18th century, where it is claimed in Berkeley's "New Theory of Vision" (1709) that depth (which cannot be projected directly onto a two-dimensional retina) is perceived by associating sensations of muscular strain (caused by the convergence of the eyes as they fixate objects at various distances) with the experience of distance. In the conventional view of speech, as in Berkeley's assumption about visual depth, apprehending the event or property is a matter of perceiving one thing and calling it something else.

Some of my colleagues and I have long argued that the conventional assumptions fail to account for the important facts about speech. Here, however, my concern is only with the extent to which they enlighten us about the relation of spoken language to its written derivative. That the conventional view enlightens us not at all becomes apparent when one sees that, in contradiction of all the differences I earlier enumerated, it leads to the conclusion that speech and reading/writing must be equally natural. To see how comfortably the conventional view sits with an (erroneous) assumption that speech and reading/writing are psychologically equivalent, one need only reconsider the four assumptions of that view, substituting, where appropriate, 'optical' for 'acoustic' or 'visual' for 'auditory.'

One sees then, that, just as the phonetic elements of speech are, by the first of the conventional assumptions, defined as sounds, the elements of a writing system can only be defined as optical shapes. As for the second assumption—viz., that speech production is managed by motor processes of the most general sort—we must suppose that this is ex-

actually true for writing; by no stretch of the imagination can it be supposed that the writer's movements are the output of an action mode that is specifically linguistic. The third assumption of the conventional view of speech also finds its parallel in reading/writing, for, surely, the percepts evoked by the optical characters are ordinarily visual in the same way that the percepts evoked by the sounds of speech are supposed to be ordinarily auditory. Thus, at the level of action and perception, there is in reading/writing, as there is assumed to be in speech, no specifically linguistic mode. For speech, that is only an assumption—and, as I think, a very wrong one—but for reading/writing it is an incontrovertible fact; the acts and percepts of reading/writing did not evolve as part of the specialization for language, hence they cannot belong to a natural linguistic mode.

The consequence of all this is that the fourth of the conventional assumptions about speech is, in fact, necessary for reading/writing and applies perfectly to it: like the ordinary, nonlinguistic auditory and motor representations according to conventional view of speech, the correspondingly ordinary visual and motor representations of reading/writing must somehow be made relevant to language, and that can only be done by a cognitive process; the reader/writer simply has to learn that certain shapes refer to units of the language and that others do not.

It is this last assumption that most clearly reveals the flaw that makes the conventional view useless as a basis for understanding the most important difference between speech and reading/writing—namely, that the evolution of the one is biological, the other cultural. To appreciate the nature of this shortcoming, we must first consider how either mode of language transmission meets a requirement that is imposed on every communication system, whatever its nature and the course of its development. This requirement, which is commonly ignored in arguments about the nature of speech, is that the parties to the message exchange must be bound by a common understanding about which signals, or which aspects of which signals, have communicative significance; only then can communication succeed. Mattingly and I have called this the requirement for 'parity' (Lieberman & Mattingly, 1985; Lieberman & Mattingly, 1989; Mattingly & Lieberman, 1988). One asks, then, what is entailed by parity as the system develops in the species and as it is realized in the normal communicative act.

In the development of writing systems, the answer is simple and beyond dispute: parity was established by agreement. Thus, all who use an alphabet are parties to a compact that prescribes just which optical shapes are to be taken as symbols for which phonological units, the association of the one with the other having been determined arbitrarily. Indeed, this is what it means to say that writing systems are artifacts, and that the child's learning the linguistic significance of the characters of the script is a cognitive activity.

Unfortunately for the validity of the conventional assumptions, they require that the same story be told about the development of parity in speech. For if the acts and percepts of speech are, as the conventional assumption would have it, ordinarily motor and ordinarily auditory, one must ask how, why, when, and by whom they were invested with linguistic significance. Where is it written that the gesture and percept we know as [b] should count for language, but that a clapping of the hands should not? Is there somewhere a commandment that says, Thou shalt not commit [b] except when it is thy clear intention to communicate? Or are we to assume, just as absurdly, that [b] was incorporated into the language by agreement? It is hard to see how the conventional view of speech can be made to provide a basis for understanding the all-important difference in evolutionary status between speech and reading/writing.

The problem is the worse confounded when we take account of both sides of the normal communicative act. For, on the conventional view the speaker deals in representations of a generally motor sort and the listener in representations of a generally auditory sort. What is it, then, that these two representations have in common, except that neither has anything to do with language? One must thus suppose for speech, as for writing and reading, that there is something like a phonetic idea—a cognitive representation of some kind—to connect these representations to each other and to language, and so to make communication possible.

Thus it is that at every biological or psychological turn the conventional view of speech make reading and writing the equivalents of speech perception and production. Since these processes are plainly not equivalent, the conventional view of speech can hardly be the starting point for an account of reading and writing.

The unconventional view of speech as a basis for understanding the difference in naturalness. The first assumption of the unconventional view is that the units of speech are defined as gestures, not as the sounds that those gestures produce. (For recent accounts of the unconventional view, see: Liberman & Mattingly, 1985; Liberman & Mattingly, 1989; Mattingly & Liberman, 1988; Mattingly & Liberman, 1990). The rationale for this assumption is to be understood by taking account of the function of the phonological component of the grammar and of the requirements it imposes. As for the function of phonology, it is, of course, to form words by combining and permuting a few dozen meaningless segments, and so to make possible a lexicon tens of thousands of times larger than could ever have been achieved if, as in all natural but nonhuman communication systems, each 'word' were conveyed by a signal that was holistically different from all others. But phonology can serve this critically important function only if its elements are commutable; and if they are to be commutable, they must be discrete and invariant.

A related requirement has to do with rate, for if all utterances are to be formed by variously stringing together an exiguous set of signal elements, then, inevitably, the strings must run to great lengths. It is essential, therefore, if these strings are to be organized into words and sentences, that they be produced and perceived at reasonable speed. But if the auditory percepts of the conventional view are to be discrete and invariant, the sounds and gestures must be discrete and invariant, too. Such sounds and gestures are possible, of course, but only at the expense of rate. Thus one could not, on the conventional view, say 'bag,' but only [b] [a] [g], and to say [b] [a] [g] is not to speak but to spell. Of course, if speech were like that, then everyone who could speak or perceive a word would know exactly how to write and read it, provided only that he had managed the trivial task of memorizing the letter-to-sound correspondences. The problem is that there would be no language worth writing or reading.

There seems, indeed, no way to solve the rate problem and still somehow preserve the acoustic-auditory strategy of the conventional view. It would not have helped, for example, if Nature had abandoned the vocal tract and equipped her human creatures with acoustic devices adapted to producing a rapid sequence of sounds—a drumfire or tattoo—for that strategy would have defeated the ear. The point is that speech proceeds at rates that transmit up to 15 or even 20 phonemes per second, but if each phoneme were represented by a discrete sound, then rates that high would seriously strain and sometimes overreach the ability of the ear to resolve the individual sounds and to divine their order.

According to the unconventional view, Nature solved the problem by avoiding the acoustic-auditory strategy that would have created it. The alternative she chose was to de-

fine the phonetic elements as gestures, as the first assumption of the unconventional view proposes. Thus, [b] is a closing at the lips, [h] an opening at the glottis, [p] a combination of lip closing and glottis opening, and so forth. In fact, the gestures are far more complex than this, for a gesture usually comprises movements of several articulators, and these movements are exquisitely context-conditioned. Given such complications, I must wait on others to discover how best to characterize these gestures and how to derive the articulatory movements from them. But while I'm waiting, I can be reasonably sure that the unconventional view heads the theoretical enterprise in the right direction, for it permits coarticulation. That is, it permits the speaker to overlap gestures that are produced by different organs—for example, the lips and the tongue in [ba]—and to merge gestures that are produced by different parts of the same organ—for example, the tip and body of the tongue, as in [da]—and so to achieve the high rates that are common.

But the gestures that are coarticulated, and the means for controlling them, were not lying conveniently to hand, just waiting to be appropriated by language, which brings us to the second assumption of the unconventional view: the gestures of speech and their controls are specifically phonetic, having been adapted for language and for nothing else. As for the gestures themselves, they are distinct as a class from those movements of the same organs that are used for such nonlinguistic purposes as swallowing, moving food around in the mouth, licking the lips, and so on. Presumably, they were selected in the evolution of speech in large part because of the ease with which they lent themselves to being coarticulated. But the control and coordination of these gestures is specific to speech, too. For coarticulation must walk a fine line, being constrained on either side by the special demands of phonological communication. Thus, coarticulation must produce enough overlap and merging to permit the high rates of phonetic segment production that do, in fact, occur, while yet preserving the details of phonetic structure.

The third assumption of the unconventional view is that, just as there is a specialization for the production of phonetic structures, so, too, is there a specialization for their perception. Indeed, the two are but complementary aspects of the same specialization, one for deriving the articulatory movements from the (abstract) specification of the gestures, the other for processing the acoustic signals so as to recover the coarticulated gestures that are its distal cause. The rationale for this assumption about perception arises out of the consequences of the fact that coarticulation folds information about several gestures into a single piece of sound, thereby conveying the information in parallel. This is of critical importance for language because it relaxes by a large factor the constraint on rate of phonetic-segment perception that is set by the temporal resolving power of the ear. But this gain has a price, for coarticulation produces a complex and singularly linguistic relation between acoustic signal and the phonetic message it conveys. As is well known, the signal for each particular phonetic element is vastly different in different contexts, and there is no direct correspondence in segmentation between signal and phonetic structure. It is to manage this language-specific relation between signal and appropriate percept that the specialization for speech perception is adapted. Support for the hypothesis that there is such a specialized speech mode of perception is to be found elsewhere. (See references given at the beginning of this section.) What is important for our present purposes is only that, according to this hypothesis, the percepts evoked by the sounds of speech are immediately and specifically phonetic. There is no need, as there is on the conventional view, for a cognitive translation from an initial auditory representation, simply because there is no initial auditory representation.

Now one can see plainly the difference between speech and reading/writing. In reading, to take the one case, the primary perceptual representations are, as we have seen, inherently visual, not linguistic. Thus, these representations are, at best, arbitrary symbols for the natural units of language, hence unsuited to any natural language process until and unless they have been translated into linguistic form. On the other hand, the representations that are evoked by the sounds of speech are immediately linguistic in kind, having been made so by the automatic processes of the phonetic module. Accordingly, they are, by their very nature, perfectly suited for the further automatic and natural processing that the larger specialization for language provides.

As for parity and its development in evolution and in the child, it is, on the unconventional view, built into the very bones of the system. For what evolved, on this view, was a specifically phonetic process, together with representations that were thus categorically set apart from all others and reserved for language. The unconventional view also allows us to see, as the link between sender and receiver, the specifically phonetic gestures that serve as the common coin for the conduct of their linguistic business. There is no need to establish parity by means of (innate) phonetic ideas—e.g., labels, prototypes, distinctive features—to which the several nonlinguistic representations must be cognitively associated.

How can reading/writing be made to exploit the more natural processes of speech?

The conventional view of speech provides no basis for asking this question, since there exists, on this view, no difference in naturalness. It is perhaps for this reason that the (probably) most widely held theory of reading in the United States explicitly takes as its premise that reading and writing are, or at least can be, as natural and easy as speech (Goodman & Goodman, 1979). According to this theory, called 'whole language,' reading and writing prove to be difficult only because teachers burden children with what the theorists call "bite-size abstract chunks of language such as words, syllables, and phonemes" (Goodman, 1986). If teachers were to teach children to read and write the way they were (presumably) taught to speak, then there would be no problem. Other theorists simply ignore the primacy of speech as they describe a reading process in which purely visual representations are sufficient to take the reader from print to meaning, thus implying a 'visual' language that is somehow parallel to a language best described as 'auditory' (see, for example, Massaro & Schmuller, 1975; F. Smith, 1971).

On the unconventional view, however, language is neither auditory nor visual. If it seems to be auditory, that is only because the appropriate stimulus is commonly acoustic (*pace* Aristotle). But optical stimuli will, under some conditions, evoke equally convincing phonetic percepts, provided (and this is a critical proviso) they specify the same articulatory movements (hence, phonetic gestures) that the sounds of speech evoke. This so-called 'McGurk effect' works powerfully when the stimuli are the natural movements of the articulatory apparatus, but not when they are the arbitrary letters of the alphabet. Thus, language is a mode, largely independent of end organs, that comprises structures and processes specifically adapted to language, hence easy to use for linguistic purposes. Therefore, the seemingly sensible strategy for the reader is to get into that mode, for once there, he is home free; everything else that needs to be done by way of linguistic processing is done for him automatically by virtue of his natural language capacity. As for where the reader should enter the language mode, one supposes that

earlier is better, and that the phonological component of the mode is early enough. Certainly, making contact with the phonology has several important advantages: it makes available to the reader a generative scheme that comprehends all the words of the language, those that died yesterday, those that live today, and those that will be born tomorrow; it also establishes clear and stable representations in a semantic world full of vague and labile meanings; and, not least, it provides the natural grist for the syntactic mill—that is, the phonological representations that are used by the working memory as it organizes words into sentences.

The thoroughly visual way to read, described earlier, is the obvious alternative, doing everything that natural language does without ever touching its structures and processes. But surely that must be a hard way to read, if, indeed, it is even possible, since it requires the reader to invent new and cognitively taxing processes just in order to deal with representations that are not specialized for language and for which he has no natural bent.

What obstacle blocks the natural path?

As we have seen, the conventional view allows two equivalent representations of language—one auditory, the other visual—hence two equally natural paths that language processes might follow. In that case, such obstacles as there might be could be no greater for the visual mode; indeed, accepting the considerations I mentioned earlier, we should have to suppose that visual representations would offer the easier route.

The unconventional view, on the other hand, permits one to see just what it is that the would-be reader and writer (but not the speaker/listener) must learn, and why the learning might be at least a little difficult. The point is that, given the specialization for speech, anyone who wants to speak a word is not required to know how it is spelled; indeed, he does not even have to know that it has a spelling. He has only to think of the word; the speech specialization spells it for him, automatically selecting and coordinating the appropriate gestures. In an analogous way, the listener need not consciously parse the sound so as to identify its constituent phonological elements. Again, he relies on the phonetic specialization to do all the hard work; he has only to listen. Because the speech specialization is a module, its processes are automatic and insulated from consciousness. There are, therefore, no cognitively formed associations that would make one aware of the units being associated. Of course, the phonological representations, as distinguished from the processes, are not so insulated; they are available to consciousness—indeed, if they were not, alphabetic scripts would not work—but there is nothing in the ordinary use of language that requires the speaker/listener to put his attention on them. The consequence is that experience with speech is normally not sufficient to make one consciously aware of the phonological structure of its words, yet it is exactly this awareness that is required of all who would enjoy the advantages of an alphabetic scheme for reading and writing.

Developing an awareness of phonological structure, and hence an understanding of the alphabetic principle, is made the more difficult by the coarticulation that is central to the function of the phonetic specialization. Though such coarticulation has the crucial advantage of allowing speech production and perception to proceed at reasonable rates, it has the disadvantage from the would-be reader/writer's point of view that it destroys any simple correspondence between the acoustic segments and the phonological segments they convey. Thus, in a word like 'bag,' coarticulation folds three phonological segments into one seamless stretch of sound in which information about the several phonological segments is thoroughly overlapped. Accordingly, it avails the reader little to be able to

identify the letters, or even to know their sounds. What he must know, if the script is to make sense, is that a word like 'bag' has three pieces of phonology even though it has only one piece of sound. There is now much evidence (1) that preliterate and illiterate people (large and small) lack such phonological awareness; (2) that the amount of awareness they do have predicts their success in learning to read, and (3) that teaching phonological awareness makes success in reading more likely. (For a summary, see, for example, I. Y. Liberman & A. M. Liberman, 1990).

Why should the obstacle loom especially large for some?

Taking the conventional view of speech seriously makes it hard to avoid the assumption that the trouble with the dyslexic must be in the visual system. It is, therefore, not in the least surprising to find that by far the largest number of theories about dyslexia do, in fact, put the problem there. Thus, some believe that the trouble with dyslexics is that they cannot control their eye movements (Pavrides, 1981), or that they have problems with vergence (Stein, Riddell, & Fowler, 1989) or that they see letters upside down or wrong side to (Orton, 1937), or that their peripheral vision is better than it should be (Geiger & Lettvin, 1989), and so on.

The unconventional view of speech directs one's attention, not to the visual system and the various problems that might afflict it, but rather to the specialization for language and the reasons why the alphabetic principle is not self-evident. As we have seen, this view suggests that phonological awareness, which is necessary for application of the alphabetic principle, does not come for free with mastery of the language. As for dyslexics—that is, those who find it particularly hard to achieve that awareness—the unconventional view of speech suggests that the problem might well arise out of a malfunction of the phonological specialization, a malfunction sufficient to cause the phonological representations to be less robust than normal. Such representations would presumably be just that much harder to become aware of. While it is difficult to test that hypothesis directly, it is possible to look for support in the other consequences that a weak phonological faculty should have. Thus, one would expect that dyslexics would show such other symptoms as greater-than-normal difficulty in holding and manipulating verbal (but not nonverbal) materials in working memory, in naming objects (that is, in finding the proper phonological representation), in perceiving speech (but not nonspeech) in noise, and in managing difficult articulations. There is some evidence that dyslexics do show such symptoms. (For a summary, see: I. Liberman, Shankweiler, & A. Liberman, 1985).

What are the implications for a theory of speech?

Those who investigate the perception and production of speech have been little concerned to explain how these processes differ so fundamentally in naturalness from those of reading and writing. Perhaps this is because the difference is so obvious as to be taken for granted and so to escape scientific examination. Or perhaps the speech researchers believe that explaining the difference is the business of those who study reading and writing. In any case, neglect of the difference might be justifiable if it were possible for a theory of speech to have no relevant implications. But a theory of speech does inevitably have such implications, and, as has been shown, the implications of the conventional theory run counter to the obvious facts. My concern in this paper has been to show that, as a consequence, the conventional theory is of little help to those who would understand reading and writing. Now I would suggest that, for exactly the same reason, the theory

offers little help to those who would understand speech, for if the theory fails to offer a reasonable account of a most fundamental fact about language, then we should conclude that there is something profoundly wrong with it.

The unconventional theory of speech described in this paper was developed to account for speech, not for the difference between its processes and those of reading and writing. That it nevertheless shows promise of also serving the latter purpose may well be taken as one more reason for believing it.

Summary

The difference in naturalness between speech and reading/writing is an important fact for the psychology of language and the obvious point of departure for understanding the processes of literacy, yet it cannot be accounted for by the conventional theory of speech. Because this theory allows no linguistic specialization at the level of perception and action, it necessarily implies that the primary representations of speech are just like those of reading/writing: neither is specifically linguistic, hence both must first be translated into linguistic form if they are to serve a linguistic function. Thus, the effect of the conventional theory is to put speech and reading/writing at the same cognitive remove from language and so make them equally unnatural.

A less conventional view shows the primary motor and perceptual representations of speech to be specifically phonetic, the automatic results of a precognitive specialization for phonological communication. Accordingly, these representations are naturally appropriate for language, requiring no cognitive translation to make them so; in this important respect they differ from the representations of reading/writing. Understanding the source of this difference helps us to see what must be done if readers and writers are to exploit their natural language faculty; why reading and writing should be at least a little difficult for all; and why they might be very difficult for some.

Acknowledgment

This work was supported by NIH grant HD-01994 to Haskins Laboratories. This paper was presented at a conference on "Language and Literacy: Comparative Approaches" in Bellagio, Como, Italy in March, 1991. The conference was organized by Beatrice de Gelder and José Morais.

References

- Berkeley, G. (1709). *An essay towards a new theory of vision*. Dublin: Printed by Aaron Rhames for Jeremy Pepyal.
- Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review*, 81, 348-374.
- Crowder, R. G., & Morton J. (1969) Pre-categorical acoustic storage (PAS). *Perception & Psychophysics*, 5, 365-373.
- DeFrancis, J. (1991). *Visible speech: The diverse oneness of writing systems*. Honolulu: University of Hawaii Press.
- Diehl, R., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology*, 1, 121-144.
- Fujisaki, M., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute* (Faculty of Engineering, University of Tokyo), 29, 207-214.

- Geiger, G., & Lettvin, J. Y. (1988). Dyslexia and reading as examples of alternative visual strategies. In C. von Euler, I. Lundberg, & G. Lennerstrand (Eds.), *Wenner-Gren Symposium Series 54, Brain and Reading*. London: The Macmillan Press Ltd.
- Goodman, K. S. (1976). Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading*. Newark, DE: International Reading Association.
- Goodman, K. S., & Goodman, Y. M. (1979). Learning to read is natural. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading*, Vol. 1. Hillsdale, NJ: Erlbaum.
- Goodman, K. S. (1986). *What's whole in whole language: A parent-teacher guide*. Portsmouth, NH: Heinemann.
- 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.
- Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Lieberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, 243, 489-494.
- Lieberman, I. Y., & Liberman, A. M. (1990). Whole language vs. code emphasis: Underlying assumptions and their implications for reading instruction. *Annals of Dyslexia*, 40.
- Lieberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle*. Research Monograph Series. Ann Arbor: University of Michigan Press.
- Mattingly, I. G., & A. M. Liberman. (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.
- Mattingly, I. G. & A. M. Liberman. (1990). Speech and other auditory modules. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Signal and sense: Local and global order in perceptual maps*. New York: Wiley.
- Miller, J. D. (1977). Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory mechanisms. In T. H. Bullock (Ed.), *Recognition of complex acoustic signals*. (Life Sciences Research Report 5). Berlin: Dahlem Konferenzen.
- Oden, G. C., & D. W. Massaro (1978). Integration of featural information in speech perception. *Psychological Review*, 85, 172-191.
- Orton, S. J. (1937). *Reading, writing and speech problems in children*. New York: W. W. Norton & Co.
- Pavlidis, G. T. (1985). Eye movement differences between dyslexics, normal, and retarded readers while sequentially fixating digits. *American Journal of Opto and Physiological Optics*, 62, 820-832.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 13, 253-260.
- Stein, J., Riddell, P., & Fowler, S. (1988). Disordered right hemisphere function in developmental dyslexia. In C. von Euler, I. Lundberg, & G. Lennerstrand (Eds.), *Brain and Reading*. England: The Macmillan Press Ltd.
- Stevens, K. N. (1975) The potential role of property detectors in the perception of consonants. In G. Fant & M. A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press.
- Wang, W. S-Y. (1981). Language structure and optimal orthography. In O. J. L. Tzeng & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.