14
Reading is Hard just because Listening is Easy

Alvin M. Liberman

In their work on reading and the difficulties that attend it, investigators have commonly omitted to ask the questions that are, in my view, prior to all others: why is it easier to perceive speech than to read, and why is it easier to speak a word than to spell it? My aim is to repair these omissions. To that end, I divide my talk into two parts. First, I say why we should consider that the greater ease of perceiving and producing speech is paradoxical, by which I mean to suggest that the reasons are not to be found among surface appearances. Then I propose how, by going beneath the surface, we can find the reasons, and so resolve the paradox.

THE PARADOX

Before developing the paradox, I should first remind you that perceiving and producing speech are easier than reading or writing, for this is the point from which I depart and to which I will, at the end, return. The relevant facts include the following. (1) All communities of human beings have a fully developed spoken language; in contrast, only a minority of these languages has a written form, and not all of these are in common use. (2) Speech is first in the history of the race, as it is in the child; reading and writing are later. (3) For the proper development of speech, a neurologically normal child need only be placed in an environment where language is used; reading and writing typically require formal tuition. And, (4), most people master speech well enough to find it useful; many fail utterly, or nearly so, to read or write.

Consider, now, in how far this is a paradox. It is paradoxical, first, because the eye, not the ear, is the better organ. By any standard one can think of, the eye is a broader channel than the ear, capable of transmitting more information about the environment and at a far higher rate. One supposes, indeed, that few human engineers would ever have tried to put so rapid a flow of information as language requires through so narrow a bottleneck as the ear affords. Which is, we might suppose, another demonstration of the validity of the observation by Francois Jacob that nature is not so much an engineer as she is a tinker.

Another aspect of the paradox is that, by comparison with the sounds of speech, the letters of print are by far the cleaner signal. Think of the precise edges that set letters apart from their backgrounds, and contrast this with the fuzzy patterns one sees in a spectrogram of speech, where information that is most important for linguistic purposes is often among the least prominent parts of the signal.

A further contribution to the paradox is that, unlike human beings, machines find it much easier to 'read' print than to 'perceive' speech. Thus, 'reading' machines (optical character readers, so called) can, without too much difficulty, be made to work according to the phonologic (hence, the alphabetic) principle. That is, given a machine designed to recognize the /d/ in 'hid', it will, without further ado, recognize the /d/ in 'do'. Not so for machines that are supposed to perceive speech. Not only do they have trouble generalizing from one occurrence of /d/ to another, but they also cannot readily discover about the word 'hid' that it has three phonological segments, while 'do' has only two. Though all words are combinations and permutations of a small set of such segments, they cannot readily be made to appear so to an automatic speech recognizer.

We come, then, to a fact that is, at once, the crux of the paradox and the key to its resolution: the relation between signal and language is relatively simple for an alphabetic representation but inordinately complex for speech. Because this is the key fact, I will develop it in greater detail as we go along. For now, it will suffice to say that it is true. An alphabetic orthography is, with occasional exceptions, a relatively straightforward reflection of the phonological or morphophonological structure of some language. Speech conveys the same structure, but in a far less straightforward way.

RESOLVING THE PARADOX

Taking into account that speech and writing systems must communicate words, we should first remind ourselves that all words are combinations and permutations of the specifically linguistic elements we know as consonants and vowels. These are governed, of course, by phonology, the component of every language that forms all its words--past, present, and future--by a systematically constrained application of the combinatorial principle. Thus, as several members of the symposium have said (I. Y. Liberman, 1988; Lindblom, 1988), the critically important function of phonology is to allow the inventory of words--i. e., of phonological structures--to be vastly larger and more flexible than it could be if, as in all nonlinguistic forms of natural communication, the signal for each meaningful element were holistically different from the signals for all others. It need not matter, not in principle at least, whether the phonological structures are conveyed acoustically, optically, or, for that matter, chemically. What does matter is that these structures should exist, however abstractly, and so provide the basis for a large lexicon. Thus, the unique characteristic of words, as opposed to all nonlinguistic vehicles of communication, is that their meanings are given, not by a direct link to the physical signals (sounds, sights, smells, etc.), but, indirectly, via the phonological structures these signals convey. We should suppose, therefore, that every word, whether spoken or written, refers to such a structure. If the language system is to find the meaning of a word, it must, perforce, know which word--i. e., which phonological structure--it is to find the meaning of.

It is, of course, nonetheless true that, just as nonhuman animals can respond more or less appropriately to some spoken words, treating them nonphonologically as if they were so many sounds, so, too, can would-be readers respond to printed words as if they were pictures or optical shapes of some nondescript sort. But such responses do not engage the language system; hence they stand apart from the kind of communication that language alone makes possible.

Granting, then, that words are systematic phonological structures, let us imagine a circumstance in which reading and writing would be, at worst, no harder than listening and speaking, and then see why, under that circumstance, language as we know it could not have developed. For that purpose, consider how the phonological structures might have been realized by the speaker as articulatory gestures and sounds. Perhaps the most reasonable-seeming possibility is that, since nature had elected to transmit the phonological structures acoustically, she would have defined its elements in acoustic terms. In that case, there would have been, for each consonant or vowel, a discrete and invariant sound, providing an acoustic 'alphabet' entirely analogous to one of the optical alphabets used for reading and writing. Speaking a word would then have been exactly equivalent to spelling it--that is, to setting out the sequence of consonants and vowels it comprised. Listening to such an explicitly segmented sound would, of course, have made its phonological spelling equally plain. To convert, then, from this acoustic alphabet to one conveyed optically, by letters, would have been trivially straightforward and trivially easy; anyone who could produce and perceive speech would have had only to learn a few perfectly transparent associations. Consequently, reading and writing would have been at least as easy as speaking and listening. But there would have been no language worth trying to read or write, for producing and perceiving an acoustic alphabet would have been so tedious and time-consuming as to have precluded the use of those larger syntactic structures that take advantage of essential groundwork by phonology to give language its truly productive character.

What makes speech, as we know it, possible is that the consonants and vowels are defined ultimately, not as sounds, but as gestures or, more properly, as the remote and abstract structures that control them (Liberman and Mattingly, 1985; cf. Browman and Goldstein, 1985; Fowler, 1986). The important consequence is that the elements of these controlling phonetic structures can be variously overlapped and merged--that is, coarticulated--and so produced at rates many times faster than is possible with an acoustic alphabet. Such coarticulation is carried out--presumably, it can only be carried out--by a biologically coherent module adapted specifically to this function. This module is special on at least two counts: its constituent gestures are a distinct set, used for producing phonetic structures and for nothing else; and their coarticulation is so coordinated as to gain high rates of transmission--about ten consonants and vowels per second, on average--and yet retain information about their underlying structure (Mattingly and Liberman, in press, a). But what is relevant to our concern is that, like all such specialized modules, this one operates at a precognitive level, below the threshold of awareness. As a result, speaking a word does not require knowing how it is spelled, or even that it has a spelling. Given that the speaker has thought of the word, the phonetic module takes care of the rest, automatically selecting and regulating the string of consonants and vowels the word comprises.

Coarticulation also speeds the perception of speech, since it folds into a single piece of sound information about several successive elements of the phonetic structure (Liberman et al., 1967). But this considerable gain in efficiency comes at a cost: there is now a complex and specifically linguistic relation between the sound and the phonetic structure it conveys. The nature of the complexity has often been described (see, for example, Liberman and Mattingly, 1985; Fant, 1962). I would speak only of the one aspect that is most relevant to our concern: the acoustic segments do not correspond directly to the segmentation of the phonological structure; accordingly, they correspond no better to the segments of the alphabetic representation. Thus, to take the example offered to this symposium by Isabelle Liberman (1988), the word 'bag' has three phonological segments, but only one segment of sound. Any attempt to divide the sound into the linguistic segments that underlie it will produce three syllables, 'buh ah guh', which are not, on their face, recognizable as the word 'bag'.

How, then, do listeners cope? The answer my colleagues and I have offered is that they rely on a phonetic module, specifically adapted for processing the speech signal so as to recover the coarticulated gestures that produced it (Liberman and Mattingly, 1985; Mattingly and Liberman, in press, a), Mattingly and Liberman, in press, b). This module is, of course, merely the other face of the one that governs the processes of coarticulation. Thus, there is but a single module with two complementary and specifically linguistic processes, one for producing phonetic gestures, the other for perceiving them. I will return to this property of the phonetic specialization later, for it is central to understanding why speech, but not reading or writing, could have evolved in the race, and why speech, but not reading or writing, can develop without formal instruction in the child. But, for the moment, the important point is that, just like the processes that underlie the production of phonetic structures, those that manage their perception take place below the level of awareness. Thus, the phonetic module automatically recovers the 'spelling', which it uses, just as automatically, to find the word in the listener's lexicon. So, to perceive a word, one need only listen to the sound. There is no need to puzzle out the underlying linguistic structure, or even to know that it exists; the phonetic module does all the hard, analytic work.

Of course, recovering the phonetic structure is not quite enough, for, as is well known, the invariant form of the word--that is, its phonological and morphophonological structure--undergoes systematic changes on its way to the phonetic surface. This occurs, both within and across word boundaries, as a function of context, stress, rate, and dialect. Thus, the voiceless fricative of 'bats' becomes the voiced fricative of 'balls'; the voiceless alveolar stop of 'bat' becomes (in American English) the voiced flap of 'batter' (as if 'batter' were the comparative form of 'bad'); all the vowels of 'telegraph' change in 'telegraphy'; and the initial semivowel of 'you', though reasonably preserved in 'Will you...?', usually becomes a voiced affricate in 'Did you...?' Surely, reading and writing would be even harder than they are if the orthography rendered words in all their myriad phonetic forms, which is presumably why all alphabetic systems, excluding only those used for technical linguistic purposes, are pitched at a reasonably abstract phonological level.

To speculate about how the speaker-listener manages the interconversion of phonetic and phonological structures, and about how much of it is attributable to the constraints of articulation and coarticulation, would take us far beyond the purpose of this essay. What is relevant for now is simply that the interconverting processes are, for the most part, just about as automatic as those of the phonetic module.

As for readers, the point is that the orthographic spelling is, by and large, closer to the invariant phonological form of the word than any of its phonetic realizations is likely to be. Therefore, readers are well advised to identify the word directly from the information about its internal structure that the orthography offers. Indeed, there is little else they can do, for orthographies, whether broadly phonological or narrowly phonetic, do not specify how words (or pseudowords) are to be pronounced -- that is, how the movements of the speech organs are to be governed. There is, then, no way that readers can get to a word via an articulatory or acoustic route, either explicitly or implicitly. The order must rather be the reverse: readers first apprehend the word; that done, they have the articulatory representation automatically available to them, ready to be used in working memory for coping with the syntax (Shankweiler and Crain, 1986; I. Y. Liberman, 1988; cf. Gathercole and Baddeley, 1988; Kean, 1988), and also for pronouncing the word should circumstances call for reading aloud.

So it is that a specialized, precognitive module, part of the larger specialization for language, automatically interconverts between spoken words and their internal structure. Consider, now, how this pertains to the four points I made at the outset about what it means to say that reading is hard and speech is easy.

(1) Speech is universal because, being managed by a module specifically adapted for the purpose, it evolved with our species. Writing systems are artifacts, hence comparatively rare.

(2) Speech could evolve, and thus come first in the history of the race, while reading and writing could not, because of the very different ways they meet a requirement that is imposed on all communication systems: what counts as structure for the sender must count for the receiver; otherwise, communication does not occur. Though rarely taken into account by theories about the processes of spoken and written language, this requirement, which Mattingly and I have called the requirement for parity, is real (Mattingly and Liberman, in press, b). To say how it is met in the development of the system is to test any theory of the biology and psychology of language or, indeed of any natural communication system. To tie it to our immediate concern is to go to the core of the difference between spoken and written language.

In this connection, consider again that there is, in speech, a phonetic module, specialized to conduct the producing and perceiving aspects of its business in a common coin of specifically linguistic gestures. This provides a built-in guarantee that, at every stage of evolutionary development, the perceptions of the listener will be in the same domain as the intentions of the speaker, a domain that is, for both, inherently and singularly linguistic: message sent and message received will always be made of the same stuff. Thus, parity is at the very heart of the phonetic module. No conventions, no cognitive interventions from outside the phonological system, no translation across domains were required in order to link, to each other and to language, what must otherwise have been wholly separate modalities of perception and action. There was but one modality, and it was essentially linguistic; hence, speech was free to evolve.

Obviously, the alphabet did not enjoy this biological advantage. It had no module adapted to its linguistic purposes, so its use had, necessarily, to rest on modalities of action and perception that evolved entirely apart from language. As a consequence, parity could only have been established by arbitrarily associating these modalities to each other and to the linguistic structures they were required, somewhat unnaturally, to convey. Accordingly, a writing system was a triumph of applied linguistics, half discovery, half invention. The discovery--surely one of the most important ever made--was that the words people had been speaking for so many thousands of years did not differ from each other holistically, but rather by an internal structure made of a small number of phonological units. The invention that exploited this discovery was the notion that, if these units were to be represented by arbitrary optical shapes, then reading and writing would be possible for all who knew the language, provided they could become reasonably aware of its phonological structure, and so appreciate what the optical shapes were all about. Thus, an alphabet, though based on one of the naturally evolved structures of language, can only have developed in cultural evolution, out of conscious, cognitive effort.

(3) For exactly the same reason that speech could evolve in the race--viz., that parity is intrinsic to the system--it can, and does, develop without formal tuition in the child. Thus, there is not in the child, any more than there was in the race, a need for cognitive intervention to establish the link between phonological structures and the sounds of speech. In the acquisition of speech, it is the phonetic module that 'learns', automatically adjusting its internal precognitive processes according to the stage of its development and the conditions of the linguistic environment (Mattingly and Liberman, in press, b). Reading and writing cannot develop in this epigenetic way for the same reason that they did not evolve in the race: parity must be imposed, and, typically, that requires special experience or instruction to lead the child to the same discovery about words that had earlier been made by those who developed the alphabet.

Only after parity is established can the child properly apply the alphabetic principle, and only by practice can he then make such application automatic.

(4) To explain why fewer children fail at speech than at reading and writing, I need only say what has been the point of the whole talk: reading and writing present a cognitive hurdle that speaking and listening do not. Because of the phonetic module, there is nothing in the child's normal experience with spoken language that necessarily acquaints him with the fact that words have an internal structure. Indeed, the automaticity of the module tends rather to make this fact obscure. Yet it is precisely this fact that must be understood if the alphabet is to make sense, and if its advantages are to be properly exploited. (Elbro, 1988; I. Y. Liberman, 1988; Lundberg, 1988; Olson, 1988).

References

Browman, C.P. and Goldstein, L.M. (1985). Dynamic Modeling of Phonetic Structure. (ed. V. Fromkin). In Phonetic Linguistics. Academic Press, New York.

Elbro, C. (1988). Morphemic Awareness Among Dyslexics. Paper presented at the Seventh Inernational Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Fant, C.G.M. (1962). Descriptive Analysis of the Acoustic Aspects of Speech. Logos, 5, 3-17.

Fowler, C.A. (1986). An Event Approach to the Study of Speech Perception From a Direct Realist Perspective. J. of Phon., 14, 3-28.

Gathercole, S.E. and Baddeley, A.D. (1988). The Role of Phonological Memory in Normal and Disordered Language Development. Paper presented at the Seventh Inernational Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Kean, M-L. (1988). Grammatical Capacity in Developmental Dyslexia: Issues for Research. Paper presented at the Seventh Inernational Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., and Studdert-Kennedy, M. (1967). Perception of the Speech Code. Psychol. Rev., 74, 431-461.

Liberman, A.M. and Mattingly, I.G. (1985). The Motor Theory of Speech Perception Revised. Cogn., 21, 1-36.

Liberman, I.Y. (1988). Phonology and Beginning Reading Revisited. Paper presented at the Seventh Inernational Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Lindblom, B. (1988). On the Evolution of Spoken Language: Some Remarks on the Origin of the 'Phonetic Code'. Paper presented at the Seventh Inernational Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Lundberg, I. (1988). Lack of Phonological Awareness--a Critical Factor in Developmental Dyslexia. Paper presented at the Seventh Inernational Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Olson, R.K. (1988). Defects in Disabled Readers' Phonological and Orthographic Coding: Etiology and Remediation. Paper presented at the Seventh Inernational Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Mattingly, I.G. and Liberman, A.M. (in press, a). Specialized Perceiving Systems for Speech and Other Biologically Significant Sounds. In Functions of the Auditory System. (eds. G.M. Edelman,, W.E. Gall, and W.M. Cowan). Wiley, New York.

Mattingly, I.G. and Liberman, A.M. (in press, b). Speech and Other Auditory Modules. In Signal and Sense: Local and Global Order in Perceptual Maps. (eds. G.M. Edelman, W.E. Gall, and W.M. Cowan). Wiley, New York.

Shankweiler, D. and Crain, S. (1986). Language Mechanisms and Reading Disorder: A Modular Approach. Cognition, 24, 139-168.

Acknowledgments. Preparation of this paper was aided by a grant to Haskins Laboratories (NIH-NICHD-HD-01994). The notions offered here are adapted from the contributions of the many colleagues, at Haskins Laboratories and elswhere, who have been concerned with the role of phonological structures in reading.