

On the relation of speech to language

†Alvin M. Liberman and Doug H. Whalen

There are two widely divergent theories about the relation of speech to language. The more conventional view holds that the elements of speech are sounds that rely for their production and perception on two wholly separate processes, neither of which is distinctly linguistic. Accordingly, the primary motor and perceptual representations are inappropriate for linguistic purposes until a cognitive process of some sort has connected them to language and to each other. The less conventional theory takes the speech elements to be articulatory gestures that are the primary objects of both production and perception. Those gestures form a natural class that serves a linguistic function and no other. Therefore, their representations are immediately linguistic, requiring no cognitive intervention to make them appropriate for use by the other components of the language system. The unconventional view provides the more plausible answers to three important questions: (1) How was the necessary parity between speaker and listener established in evolution, and how maintained? (2) How does speech meet the special requirements that underlie its ability, unique among natural communication systems, to encode an indefinitely large number of meanings? (3) What biological properties of speech make it easier than the reading and writing of its alphabetic transcription?

Taking speech in the narrow sense – as the production and perception of the sounds that convey phonetic structure – one finds two very different views of its relation to language. The more conventional holds that speech is merely a vehicle, bearing no organic relationship to the linguistic baggage it carries. On that view, speech is produced and perceived by processes that are not specialized for language but rather serve horizontally the broadest possible variety of behaviors, linguistic and non-linguistic alike. The outcomes of those primary processes are then presumably sent on to language proper, a separate domain where they find the mental machinery capable of the heavy lifting required by phonology, morphology and syntax. Preferring a name that reflects the nature of a theory rather than its currency, we will call the conventional view ‘horizontal’ (as in Ref. 1). The other, less conventional view is that the biological roots of language run deep, penetrating even to the level of speech and to the primary motor and perceptual processes that are engaged there. Seen from that perspective, speech is a constituent of a vertically organized system, specialized from top to bottom for linguistic communication. Such a view may be appropriately called ‘vertical’.

To evaluate these strongly contrasting views, we propose here to determine which of the two provides the more coherent and plausible account when challenged by several simple and seemingly obvious biological considerations. As we will try to show, those considerations provide decisive tests of any theory of speech, yet they do not normally figure in the cal-

culations of the theorists, nor have they been permitted to ruffle the implicit assumptions that guide (or misguide) almost all applied language research, including that which is aimed at determining how to convert fluency in speech to fluency in the use of its alphabetic transcription. Indeed, bringing those considerations to notice is one purpose of this paper. But first, we will give a brief description of the theories.

The horizontal viewpoint

The first assumption of the horizontal view is that the elements of speech are sounds. That is not merely to say the obvious, which is that speech exploits an acoustic medium, but rather to identify sounds as the primitives that are exchanged when linguistic communication occurs. The invariant acoustic patterns that might justify such an assumption have, indeed, been claimed for a variety of phonetic segments, including, for example, stops², nasals³, and the voicing contrast of fricatives⁴. However, most students of language, and virtually all of those interested in its many applications, simply take for granted that sounds are the elements of speech. It is not usual, therefore, to find that premise developed or defended explicitly as a basis of the horizontal theory, though it is that. And if few bother to dismiss imaginable alternatives, it is, apparently, because alternatives do not spring readily to mind.

As for what should be the second assumption, the one that would concern the articulatory gestures that produce the elemental sounds, there is, again, little in the way of explicit

†A.M. Liberman
(deceased).

D.H. Whalen is at
the Haskins
Laboratories,
270 Crown Street,
New Haven,
CT 06511-6695,
USA.

tel: +1 203 865 6163
fax: +1 203 865 8963
e-mail: whalen@
haskins.yale.edu

conceptual development or consideration of reasonable alternatives. Conspicuously absent, at all events, is the idea – important for the purposes of this review – that those gestures might be specialized for language. One is left to infer, then, that they are not, that their biology is no more distinctly linguistic than that which underlies the motor processes that curl the toes or twiddle the thumbs. Perhaps the most explicit statement of this view is that of the linguist Bjorn Lindblom who has written:

Speakers adaptively tune phonetic gestures to the various needs of speaking situations, and languages make their selection of phonetic gesture inventories under the strong influence of motor and perceptual constraints that are language independent and in no way special to speech. (Ref. 5, p. 7)

In further dismissing the possibility that speech production is a specialized adaptation, Lindblom asks why, if this is so, do ‘inventories of vowels and consonants nevertheless show evidence of being optimized with respect to motoric and perceptual limitations that must be regarded as biologically general and not at all special to speaking and listening.’ (p. 7.)

Unlike Lindblom, most horizontalists give speech production short shrift, assigning to it a secondary role in which its chief function is simply to articulate those sounds that fit comfortably the language-independent properties of the ear^{6–9}. Consistent with that view is the particularly close attention that horizontalists pay to the processes of perception that speech production must serve.

Accordingly, the third assumption – the one that is made most explicitly by all the horizontalists and that has, therefore, provided the occasion for a lively clash of experiments for and against the horizontal and vertical views – is that perception of speech is like the perception of all sounds^{7–11}. On that assumption, there is at the level of primary perception, as at the corresponding level of production, no specialization for language. Rather, phonetic perception is presumably accomplished by the standard all-purpose equipment of the auditory system. The result is that the primary representation of the listener is composed of non-linguistic auditory primitives and the patterns they form. One looks, then, to the language-independent auditory system to meet such particular perceptual requirements (for example, the categorical perception of phonetic elements) as the demands of phonetic communication impose. As for the development of speech in the biological world, it is as though evolution had simply appropriated certain properties of the ear that happened to lie conveniently to hand, selecting those that could most effectively be made to serve the novel function of communication by phonetic means. It should be emphasized at this point that the concern of this essay is only with the production and perception of phonetic structure, for there can surely be no argument about the purely auditory (hence horizontal) nature of such paralinguistic aspects of the signal as the register, loudness and timbre of the speaker’s voice.

A fourth assumption is, or is not, necessary, depending on what the horizontalist assumes about the biology of those grammatical processes that lie beyond speech, and are therefore regarded by almost everybody as deserving of the name ‘language’. Thus, some who prefer a horizontal view of speech elect an equally horizontal view of language^{12–15}, holding

that the general motor and auditory representations of speech are passed on without change to correspondingly general processes of the cognitive machinery, the latter having been bent by intellectual exertions of one kind or another to the grammatical functions appropriate (but presumably not unique) to language. On that view, the three assumptions identified above would suffice. However, most horizontalists assume, at least tacitly, that language occupies a specialized domain and therefore accept the need to translate the linguistically undistinguished results of the primary processes into that domain. For that purpose, they must look to a second stage, beyond primary action and perception, where the secular motor and auditory results of the early processing are, by some necessarily cognitive translation, invested with phonetic privileges and so made usable by the special processes of language. That translation is variously accomplished by matching the primary auditory percept to phonetic templates^{7,16,17}, by associating it with the abstract ‘distinctive features’ so familiar to linguists^{9,18}, or, as we are left to infer by some treatments, simply by giving it a phonetic name^{19,20}. For the purposes of this review, these are distinctions without a difference; in assuming the need for translation to linguistic status, however done, the horizontalist accepts that language is a specialization, while at the same time denying that speech is part of it.

The vertical alternative

The first assumption of the vertical view is that the phonetic elements of speech, the true primitives that underlie linguistic communication, are not sounds but rather the articulatory gestures that generate those sounds^{21–25}. Those units, which serve as the primary representations of speaker and listener alike, are presumed to be something like the coordinative structures that Turvey and others have assumed^{26–29}. As expressed in the peripheral structures, the gestures are changes in the cavities of the vocal tract – openings or closings, widenings or narrowings, lengthenings or shortenings. It is supposed on the vertical view that the processes underlying the gestures evolved with language, specifically in the service of a linguistic function; hence, the gestures are phonetic *ab initio*, requiring no cognitive translation to make them so²¹. Coincident with the evolution of the phonetic mechanisms was what Studdert-Kennedy has called the ‘articulation’ of the vocal tract³⁰ – that is, the division of vocal-tract action into independently controllable components (the lips, the blade, body and root of the tongue, the velum, and the larynx). That articulation, which is unique to our species, opened the way for the selection of gestures that could be overlapped, interleaved, and merged – that is, co-articulated. The biological payoff was fast phonetic action from relatively sluggish organs²³. There was also in evolution a descent of the larynx, creating a more spacious pharyngeal cavity and the possibility for a larger inventory of vowels³¹.

The gestures are combined in various ways to form the phonetic segments, as if the gestures were the atoms of language and the segments its molecules. The function of the segments is presumably to provide the sharp boundaries that words must have in the lexicon, boundaries that are commonly lacking in fluent speech, where co-articulation across words is common. The subsegmental gestures are nonetheless

important, however, because they increase the possibilities for co-articulation beyond what could be done with discrete segments. For the particular purposes of this paper, the difference between segment and gesture is of no great importance, so for convenience, in what follows we will not distinguish them. At any event, we will not try to explain what we do not yet understand, which is how the one is gathered up so as to form the other.

The second assumption is that the articulation and co-articulation of the gestures is controlled by a species-specific specialization for language – a phonetic module^{22,32,33}. The inventory of specialized phonetic gestures it controls is severely limited, both in the style of movement of the gesture (manner of production) and the surface of the vocal tract that is the target (place of articulation). As already implied, those restricted gestures form a natural class, specifically phonetic in nature, that stands apart from the non-phonetic movements (e.g. chewing, swallowing, moving food around in the mouth) made by the same organs. A critical function of the phonetic module is to take advantage of the articulated vocal tract so as to produce a great deal of overlapping, interleaving and merging of the gestures, while precluding those temporal relationships among the gestures that would cause the acoustic consequences of the one or the other to be obscured³⁴.

The third assumption, which is about perception of the elemental gestures, takes account of the fact that co-articulation creates a complex relationship between the acoustic signal and the phonetic structure it conveys: as is well known, there is a lot of context-conditioned variation in the acoustic information for any phonetic segment, as well as a lack of correspondence in segmentation between acoustic signal and phonetic structure. Unraveling that complex relationship between signal and message is the business of the same phonetic module that produced it, for that module incorporates the constraints necessary to process the signal so as to recover the very gestures that were, by their co-articulation, responsible for its apparent complications^{21–23}. (As indicated at the beginning of this section, Fowler^{25,35,36} holds with the verticalists that the elements of speech are gestures, but she takes their perception to be instances of what all proper perceptual systems do; that is, to perceive the distal object or event. No specialization for speech perception is necessary.) The result is that the listener effortlessly perceives the gestures the speaker produced, not because the process is simple, but because the phonetic module is so well adapted to its complex task.

The biological considerations

How do speakers and listeners know what counts; how are they able to communicate in the same code; and how did production and perception develop together in evolution? Among the requirements that are imposed on all natural forms of communication, whether linguistic or not, there is one that is, at once, the most obvious and most important, yet, except for a particularly insightful treatment by Rizzolatti and Arbib³⁷, it has escaped the notice of theorists. On the assumption that a thing is more likely to be noticed if it has a name, Mattingly and Liberman³⁸ have called it the requirement for ‘parity’, and have suggested that it has three facets.

The simplest facet applies even to one-way communication between speaker and listener. There, in what is the

least constraining exchange, communication can succeed only if the two parties have a common understanding about what counts – what counts for the speaker must count for the listener. Thus, both must recognize, either perceptually or cognitively, that /ba/ counts but a sniff does not. The obvious problem for the horizontal theory is that it provides no basis for such common recognition at a perceptual level. Because /ba/ and the sniff both elicit auditory percepts, the difference is just that one is phonetically significant, the other not. But that difference can be appreciated only at some cognitive remove, for it is only there that it can be discovered whether or not the auditory percept is connected to a phonetic unit. The horizontalist should wonder, then, how the phonetic unit evolved independently of its manifestation in action or perception, how one set of percepts rather than some other was crowned with phonetic significance, and how it was determined which auditory percepts were to wear which phonetic crowns. Because the horizontalists have not been concerned to raise the question, one is left to infer the wholly implausible answer that might reasonably follow from their assumptions: that parity must have been established by deliberate agreement among our earliest-speaking ancestors, as if the units of speech were not products of nature, but artifacts, just like the letters of the alphabet.

Seen vertically, however, the speaker produces gestures that are managed by machinery that evolved in connection with a specifically communicative (linguistic) function, from which it follows, as we proposed earlier, that the gestures are phonetic by their very nature. We also proposed that they are recovered from the acoustic signal by the phonetic component of the larger specialization for language. Hence, their perception as phonetic elements is immediate and distinct, leaving the listener in no doubt that they count for linguistic purposes. There is no need for an appeal to cognitive connections between an initial auditory representation and some more abstract phonetic unit to which it is somehow linked, because, as should be said again, the initial representation is already phonetic.

To see the phonetic module even more clearly as a component of the larger language module, a theorist does well to note how far its processes resemble those of syntax. Consider, then, how the phonetic module presumably goes about deciding that some signal does or does not contain information about phonetic structures. The important point is that it can hardly do that by reference to the surface properties of the acoustic signal, as might be seen most readily in the case of sine-wave speech. There, the formants and the band-limited noises (of the fricatives) are replaced by sinusoids that follow the centers of the formants and the noises. Those sinusoids have no common fundamental, no common movement, nor indeed, any acoustic property that would make them coherent from an auditory standpoint. Yet, if they describe phonetically telling trajectories of the articulators, listeners hear phonetic structures^{39,40}. Thus, the only basis for perceptual coherence is at the same time gestural and phonetic. That being so, we should not expect to identify surface acoustic characteristics that will reliably distinguish three sine waves that cause the listener to perceive phonetic structures from three acoustically similar sine waves that will be heard as unrelated tones. In that respect, phonetic

perception is like perception of syntax. As Mattingly and Liberman have put it:

[T]he signal is speech if and only if the pattern of phonetically significant articulatory gestures that must have produced it can be reconstructed. We call this property 'generative detection', having in mind the analogous situation in the domain of sentence processing. There, superficial features cannot distinguish grammatical sentences from ungrammatical ones. The only way to determine the grammaticality of a sentence is to parse it – that is, to try to regenerate the syntactic structure intended by the speaker. (Ref. 38, p. 785)

Surely it is appropriate, if speech is a component of the larger specialization for language, that its perception should rely on the same kind of synthetic processing that characterizes syntax. It is equally appropriate that the gestures should have evolved with language in the service of an exclusively phonetic function, and that they are, therefore, like syntactic structures, linguistic by their very nature.

The second facet of the parity requirement applies to two-way communication – that is, communication in which speaker and listener exchange roles. In this case, it is essential that the phonetic representation in the brain of the one party be replicated in the brain of the other; otherwise they cannot communicate in the same code. But on the horizontal view, the primary representation of the one party is, at any given moment, purely motor, whereas the representation of the other is purely auditory. Those representations are in no way alike, except that neither has anything to do with language. Where, then, do the parties find the common ground on which they must stand, and what reserves that ground exclusively for events that have communicative significance? We suppose the horizontalist would say that the very dissimilar motor and perceptual representations are connected to the same phonetic unit. But that runs foul of the troubling questions, raised earlier, about the nature and origin of such a phonetic unit and the process by which certain motor and auditory representations are selected for connection to it. The vertical view, alternatively, permits us to see that the parties conduct their business in the common currency of phonetic gestures: the gesture that is the unit of production for the speaker is the unit of perception for the listener. There is no need for either to refer grossly dissimilar motor and auditory representations to phonetic units to which they bear a relationship that is, from any point of view, wholly arbitrary.

The third facet of the parity requirement takes into account that the biological development of speech required a process of co-evolution. That is, production and perception must have marched through evolutionary development in perfect step, or else the system would have ground to a halt. Unfortunately, we suggest, for the plausibility of the horizontal view, its proponents maintain quite firmly that the motor and perceptual representations of speech are distinctly different, that such connections as exist between them were established in the experience of speaker/listeners as they perceived the auditory consequences of their own articulatory maneuvers. But connections that depend on learned associations of that kind do not become part of what is handed down in evolution. How, then, can the horizontalist view

account for the fact that the production and perception of speech did co-evolve? A plausible account is, yet again, readily available on the vertical view, because it holds that the primary representations of speaker and listener are exactly the same; accordingly, a change in one is inevitably mirrored by the same change in the other.

What is biologically unique about speech?

Speech is not only a product of evolution, but also species-specific. Not that speech is biologically unique in that respect, for surely each species is endowed for the purpose of communication with its own specialization, a specialization that provides, among other things, a basis for the parity that all communication systems must have. Each such specialization exploits some very specific variety of signal patterns in one or more possible media. Thus, some creatures use acoustical signals, some optical, some mechanical, some chemical, and electric fish do it electrically. But whatever the signal pattern or medium, all the non-human systems have a critically important property in common: every signal is meaningful, and signals are never rearranged to convey new meanings. So, short of waiting for evolution to provide a new signal-meaning pair, there is no way non-human animals can ever say anything new. Their communication systems are closed.

Language is different in a profoundly important way, for language is open or generative, capable in principle of encoding an infinite number of meanings. One necessary component for generativity is that the sign hold an arbitrary relationship to the thing signified^{41,42}. An even more important difference between language and other natural forms of communication arose when evolution for the first time divorced the form of the signal from semantic function^{30,43,44}. That is, through evolution, signals were created having a distinctly communicative cast, as had happened for other species, but now without meanings assigned to them. That novel evolutionary step opened the way for speech to follow the combinatorial strategy, or what Abler has called the 'particulate principle of self-organizing systems'⁴⁵. In all particulate systems, as Studdert-Kennedy has characterized them³⁰, a few discrete units 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. That principle is at work in chemical compounding, in genetics, and in language, serving in all three domains to achieve infinite ends with finite means. Thus, in speech, a few meaningless segments are variously combined and permuted to form an indefinitely large and expandable vocabulary of meaningful words, which, in turn, provides the necessary base for the further application of the principle to the formation of propositions through the combinatorial processes of syntax.

Implications of the particulate principle

The particulate principle imposes several requirements on speech, two of which are of particular concern for the evaluation of the horizontal and vertical points of view. The more obvious requirement is that the particles be commutable, which is to say that, as produced and perceived, they must be discrete, invariant and categorical. That is an absolute requirement, which cannot be compromised in any way. The other requirement has to do with rate: given the characteristics 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 will typically run to considerable lengths. It is essential, therefore, that the particles be produced and perceived rather rapidly if they are to be organized into the larger units of the language hierarchy. In fact, the consonants and vowels that are formed by the particles are delivered in speech at about 10 or 12 per second on average, and for short stretches the rate rises even higher⁴⁶.

How do discrete, invariant and categorical elements meet the rate requirement? Consider now that if the elements of speech were sounds, as the horizontal view has it, then the particulate nature of speech would necessarily be manifest at the acoustic surface. In that case, speech would be an acoustic alphabet. But speech cannot be an acoustic alphabet, for if it were, we could speak only as fast as we can spell. Perception would be similarly limited, as it would require listening to spelled speech. No one should have been surprised, then, to discover, as Cyril Harris did many years ago⁴⁷, that a tape recording of speech cannot be broken up into phoneme-size segments and then rearranged to yield an intelligible permutation. Neither should anyone have been surprised to learn from the very early research on speech that it is, in fact, not an acoustic alphabet but something more like what the vertical view takes it to be. On the vertical view, the particulate (hence alphabetic) nature of speech is represented by the elemental gestures, which are, as they must be, discrete, invariant and categorical. They are produced and perceived as rapidly as language requires because they are co-articulated, which preserves their commutability at the level of the gestures but not in the acoustic signal.

How is perceptual form fit to phonetic function?

That speech is normally co-articulated is established beyond dispute, as are many of the consequences for the acoustic signal. Indeed, that part of the vertical view is so widely accepted that it can hardly be regarded as unconventional. What is just as widely rejected, however, is the essential vertical premise that the elements of speech are phonetically significant gestures, not sounds. The corollary of that premise, also explicitly rejected, is that a specialization for language processes the acoustic signal and yields a primary percept that is distinctly phonetic, not auditory. It is around that issue – of whether the primary processes of speech perception are generally auditory or specifically phonetic – that many dozens of experiments have revolved. Indeed, there are so many particular experiments and so many correspondingly particular results that favor one answer or the other, that it becomes very hard, and ultimately futile, perhaps, to seek a final score that identifies a winner (see also Ref. 48). What can be done more profitably, we suggest, is to take a broader view and thereby see in the light of some generally agreed characteristics of speech which theory appears the more plausible.

Evaluating the theories

Consider, first, the widely accepted consequence of co-articulation, which is that information for a phonetic segment is typically overlapped with information for other segments in the string and distributed over a relatively large span of the

signal. Thus, for example, information for the stop consonant /p/ in the syllable /spi/ is in the spectral shaping of the fricative noise (it would be shaped differently for /sti/); in the silent interval that separates the fricative noise from the vocalic part of the syllable (that would be different for /si/); in the formant transitions at the start of the vocalic section that also provide information about the vowel (they would be different for /spa/); and in the vowel itself (the shaping of the fricative noise and the formant transitions would be different for /spu/) (Refs 49–51). In that example, the information for the second segment in a syllable that comprises three segments is spread from the beginning of the syllable to its end. Or take the case of a syllable like /bag/. When pronounced rapidly, but nonetheless clearly, the information for the vowel will probably extend through the entire syllable, overlapping grossly with both the initial and final consonants. That the information for a vowel positioned between two consonants can extend through the entire syllable is shown most dramatically, perhaps, in the case of the ‘silent center’ vowels. In that paradigm, the experimenter removes from the acoustic signal everything except the consonant–vowel transitions that occupy only the initial and final 50 ms (Refs 52–54). Even with this drastically altered, unnatural-sounding syllable, the vowel is correctly identified. Those brief pieces of sound convey information sufficient for the identification of all the segments in the syllable. Is it plausible, then, to imagine an auditory system that might have evolved, independently of language, to do what phonetic perception does in those cases; that is, to segment into discrete percepts a signal in which information for each of the unit percepts is grossly overlapped with information for the others, or (as in the case of /spi/) to integrate into a single perceptual unit (/p/) information that is spread across the adjoining units? What could possibly have selected for such auditory characteristics?

We suggest that they did not evolve in the auditory system just in case of the possibility that speech would come along and find them useful. And would they not have been dysfunctional for non-speech perception, causing continuously variable acoustic signals that reflect continuously varying events to be divided, perceptually, into discrete and disconnected units that would in no way reflect the physical reality? Or, in other cases, to integrate into a unitary percept information relevant to distinct and successive physical events? On the vertical view, however, it is possible to see that all of the aforementioned effects of phonetic perception are what one would expect if it is the gesture that is produced and perceived. For then, in the case of /spi/, the broadly spread acoustic cues are the common products of the same gestural segment; in the case of /bag/ there are three discrete but overlapped gestures; and in the case of the ‘silent center’ vowels, information about the vowel gesture is spread throughout the syllable, just as it is in /bag/. Thus, the phonetic specialization processes the gestures that are appropriate for language without in any way interfering with the ability of the auditory system to offer a veridical representation of the acoustic world.

It is also widely accepted that the phonetic segments – most obviously the consonants – tend to be perceived categorically, and, further, that the categorical perception of each segment is invariant across all the conditions that cause the

acoustic signal to vary. The consistent finding about categorical perception is that, given stimuli that are adjacent on some acoustic continuum between two consonants, there is an increase in discriminability at the phonetic boundary⁵⁵. That increase establishes two categories, one on either side of the perceptual discontinuity that the increased discriminability reflects. (The increase in discriminability is normally not so great as to indicate that listeners can discriminate only as well as they can identify, which would be the case if perception were perfectly categorical. But the shortfall may be a consequence of the limitations of acoustic synthesis. As it is used in experiments, such synthesis typically changes only one of the relevant acoustic cues, causing the percept to depart progressively from its proper phonetic form and to be perceived therefore as having more and more of a non-phonetic cast; a result of this is that discrimination of a purely auditory kind creeps in. A remedy does not come easily, not only because there is presently available no articulatory synthesizer that would change all the acoustic cues appropriately, but, more importantly, because the articulatory routes from one consonant to another are not continuous. That is to say that production is no less categorical than perception, which is, of course, precisely the point.)

On the horizontal view, there have been two very different explanations of categorical perception. One is that 'categoricalness' is a consequence of the cognitive component of the two-stage process that the horizontalists must assume if their primary non-linguistic auditory percepts are to be given phonetic status. Thus, several investigators have assumed that the listener attaches a phonetic label before the echo of the primary auditory percepts has faded^{19,20,56,57}. It is, therefore, the label, not the primary percept, that is categorical. However, as in all other attempts to reconcile a horizontal view with the requirement of parity, there is no basis for understanding, even in the most general terms, where the phonetic labels come from, or how it is decided which labels go with which auditory percepts.

The more common explanation of categorical perception is that the discontinuity is a property of the auditory system. To evaluate that explanation, it is relevant first to reflect on the widely recognized fact that the acoustic boundaries between segments are not fixed. Even if we accept that some aspects of the acoustic signal are invariant for consonants, there remain the formant transitions – sufficient cues for most stop-consonant perception – which invariably change with phonetic context⁵⁸. The transitions also move as a function of position in the syllable, most dramatically when, as cues for initial and final consonants, they are mirror images⁵⁹. Lexical stress and rate of articulation affect the phonetic boundaries for the voicing distinction⁶⁰⁻⁶². Further, it is also surely relevant that, although the acoustically defined dimensions might be the same across many languages, the placement of the phonetic boundaries varies⁶³⁻⁶⁵. Given these variations in the salient acoustic cues, the number of category-determining discontinuities in the auditory system is beyond enumerating. One wonders, then, what would have selected for those numerous auditory discontinuities? Again, it was surely not in anticipation of their usefulness for a behavior – speech – which had yet to appear on the biological scene. In any case, would not those auditory categories be seriously maladaptive,

grossly distorting physical reality by causing perception of continuous non-speech acoustic events to appear discontinuous? And is it not implausible to suppose that speakers are able to manage their limited articulatory possibilities so as to make the acoustic result conform to the many categories defined by those highly variable discontinuities? On the vertical view, by contrast, one sees that the perceptual boundaries are exactly where the conveniences of articulation and co-articulation put them, not where the properties of the auditory system would have set them down.

The horizontal view also appears the more implausible when one considers that phonetic boundaries are typically marked not by one acoustic cue but by several. For example, the differences among the voiced stops are signaled by the second-formant transition, but also by the third, and there is no way a speaker can control these independently⁶⁶. If the speaker has managed her articulations so as to produce signals appropriate to the second-formant boundary, what is the probability that the acoustic result would happen to be appropriate also for the third-formant boundary, and, moreover, that this happy but highly improbable coincidence would occur for all phonetic segments, for all contexts, and for all positions in the syllable?

Still another difficulty for the horizontal view arises from the fact that there is, within limits, perceptual equivalence among the several acoustic cues for each category, no matter how acoustically diverse they may be. Thus, the difference between /slit/ and /split/ can be cued by varying either the silent interval following the fricative noise or the starting points of the formant transitions. Experiments have shown that exactly the same perceptual effect is produced by either of these manipulations – that is, listeners cannot tell whether the phonetic difference was produced by varying the silence or the first-formant transition^{67,68}. There is a strong presumption that a similar equivalence holds for all such cases – that is, whenever there are several acoustic cues for the same segment, which is virtually always. How plausible is it, then, to suppose that the auditory system evolved so as to provide the perceptual equivalences of those numerous and acoustically disparate cues? The patently maladaptive consequence would be that many very different acoustic signals representing different events in the non-speech world could not be discriminated. The vertical view has no problem with the aforementioned facts, for the acoustic cues, no matter how numerous or disparate, are perceptually equivalent and therefore appropriate to the same phonetic segment because they reflect the same phonetic gesture.

Speech versus reading and writing

How do we account for the biological gulf that separates speech from the reading and writing of its alphabetic transcription? The preliterate child is a prodigy of phonologic development. Commanding thousands of words, he readily produces their phonologic structures when speaking, and just as readily parses them when listening. Thus, he exploits the particulate principle quite naturally, without its ever having been taught to him, and without his having to be aware of the principle or of the remarkable ability it makes available to him. For the skillful use of that principle in speech, it is enough to be a normal member of the human race and to

have been exposed to a mother tongue. By contrast, applying the particulate principle to the task of reading and writing is not an automatic outgrowth of the natural capacity for language but an achievement of a distinctly intellectual kind. To understand the reading process, one would think it critical, therefore, to know exactly where the biological difference lies. Yet, in all the vast literature on reading, that question is never answered or even asked (but see Refs 69,70), certainly not by that overwhelming majority of researchers who explicitly or tacitly accept a horizontal view. A likely reason for this serious omission is that, given the horizontal assumption, no reasonable explanation is possible. Consider that, on any view, the relationship between the alphabet and speech is entirely arbitrary: the visual percepts evoked by the alphabetic characters are of no use until they have been translated into the linguistic units they symbolize. Accordingly, reading is always a translation, and translation is, by its nature, effortful and deliberate. But on the horizontal view, the same requirement is imposed on speech. As we saw earlier, on that view, the sounds of speech are thought to evoke auditory percepts which, like the initial visual percepts of the reader, become linguistically useful only after translation into the phonetic segments they happen to convey. Indeed, one might expect from the horizontal view that reading and writing would be easier than speech. After all, the alphabetic characters are clearer signals than the sounds of speech; the hand and fingers are more versatile effectors than the tongue; and the eyes are more accommodating receptors than the ears.

Even so, putting aside the notion that speech should be harder if one accepts the horizontal view, one must wonder why reading is, at least, not equally easy. Why does mastery of speech not fully prepare the speaker or listener for the seemingly trivial task of substituting the letters of the alphabet for the sounds of speech? An important and proper answer was provided by I.Y. Liberman and Shankweiler⁷¹, who brought to notice that, contrary to what the horizontalist must suppose, the letters of the alphabet do not correspond to sounds but to the underlying, co-articulated gestures, for those are the true phonemic constituents. Unfortunately for the would-be reader, they are less readily available to consciousness. The research that was stimulated by that application of the vertical view revealed that pre-literate children typically do, in fact, lack awareness of phonemic structure and, consequently, find themselves unable to fathom and properly use an alphabetic transcription⁷². Thus began the investigations into phonemic awareness that have proved so fruitful for an understanding of how children read and why some cannot^{73–76}.

At this later stage in the development of the vertical view, we see yet another reason why pre-literate children lack phonemic awareness. For we now more clearly understand that the primary motor and perceptual representations are already phonetic, requiring no translation from some generally motor or auditory form. Those representations are therefore immediately in the language domain, hence perfectly suited for further processing by the other components of the larger specialization for language. Requiring no attention to be spent on translation, the primary representations receive none. To develop the phonemic awareness that reading and writing call for, the child must therefore learn to put his attention

where it has never had to be. Indeed, he must overcome a previously appropriate habit of overlooking the meaningless phonemes in favor of the meaningful morphemes and words^{77,78}.

As for why phonemic awareness is not necessary for speech, we need only suppose, as is consistent with the vertical view, that the speaker thinks of the word, which is presumably in the lexicon as a phonemic structure, and then leaves it to the phonetic module to select and coordinate the distinctly phonetic gestures. However, when the speaker undertakes to write a word, the phonetic module is struck dumb, leaving the speaker to rely, if he can, on his conscious awareness of the phonemic structure of the word he would write. As for the listener, he relies on that same phonetic module to process the sounds and represent, without auditory mediation, the phonemic structures that identify the words. All this is to say simply that speech does not require phonemic awareness for the same reason that it does not produce it.

How special is speech?

The claim from the vertical standpoint that speech in the narrow sense is a specialization has seemed to some to call for the application of Occam's razor. Why have a special system when something more general might do⁷⁹? The answer is that a more general process will *not* do. At the very least, there must be parity, which requires that signals with communicative significance belong to a special mode, as it were, where they are clearly marked for their distinct communicative function. That requirement applies to all animal communication, not just to the one we humans enjoy. Would we, then, dare assume about a non-human creature that it perceives communicative signals exactly as it perceives all others, recognizing their special significance only by some secondary cognitive process similar to that which the horizontalist assumption attributes to human speech perception? Presumably not. So, in assuming a cognitive translation for speech, the horizontalist puts humans at odds with the biology of communication as it is evident in all other species. Which way, then, do we want Occam's razor to cut?

To see how speech perception can be put in a class of perceptual specializations, and so made to appear less exceptional from a biological point of view, we should first take note of one of its most apparent but least remarked characteristics, which is that phonetic units do not have an end organ of their own. Accepting Berkeley's explicit treatment of the matter in his *New Theory of Vision*⁸⁰, many psychologists assume that a primary percept is evoked only by an appropriate end organ. It follows that a phonetic representation must be a translation from the ordinary auditory primitives that own the ear as their end organ, and presumably stand, therefore, as the only primary percepts that can be evoked by the acoustic signals to which the ear responds. However, there are other percepts that are primary, yet, like speech, have no end organ, and hence no labeled line to peripheral equipment that is dedicated to their needs. The most relevant, perhaps, is sound localization. There, the acoustic cues are interaural differences of time and intensity, but nobody takes those differences to be the primary percepts that are then cognitively translated into location. Rather, it is understood that there is a system

specialized to process the interaural cues and represent them immediately as location. Visual perception of depth presents a similar case in that information about binocular disparity is processed by a system specialized to represent it immediately as depth.

These systems all depend on what Konishi has called a ‘central synthesis’⁸¹. As Mattingly and Liberman have put it³⁸, such systems are ‘heteromorphic’, in that the percept is incommensurate with the stimuli, but only when the stimuli conform reasonably to the ecological circumstances for which the perceptual system is specifically adapted; otherwise, they assume a ‘homomorphic’ form. Thus, except when the interaural time and intensity differences far exceed what is ecologically possible, the stimuli for sound localization are not heard homomorphically as sounds that arrive at the ears at different times or with different intensities, but heteromorphically, as location in azimuth. Stereopsis presents a similar case: except when the binocular disparity is far beyond normal limits, the stimuli for visual depth are not perceived homomorphically as disparate images but heteromorphically as phenomenal depth. For speech, the stimulus is an ensemble of several resonances that change their spectral positions more or less continuously, yet, except when those changes do not reflect the trajectories of phonetically significant gestures, they are not perceived homomorphically as continuously changing timbres, but heteromorphically as a string of discrete and categorical segments.

The heteromorphic specializations have several characteristics in common with phonetic perception. One is a plasticity that allows them to be calibrated by the environment. Thus, the specialization for stereopsis, depending as it does on binocular disparity, must be recalibrated as the child’s head grows and the distance between the eyes increases. Much the same must happen in the case of sound localization, as growth causes a change in the distance between the ears and therefore in the interaural time difference. Such calibration represents a kind of learning. It is not the kind of learning that psychologists commonly study, which is unfortunate for our purposes, because it is the kind that occurs also in speech. For in speech, as in the other cases, the specialization is calibrated by the environment. The necessary and sufficient condition for appropriate phonetic calibration is simply exposure to the right environment; the required perceptual ‘learning’ is effortless and, for neurologically normal children, inevitable.

Speech perception also shares with the heteromorphic specializations an elasticity that allows them to respond in their usual way to stimuli that are, within limits, ecologically impossible. Thus, viewers will perceive depth even when the binocular disparity is made to correspond to a greater interocular difference than a real head could ever provide. Beyond a certain disparity, however, the limit of elasticity is reached, and the viewer perceives not only heteromorphic depth but also homomorphic double vision (diplopia). With further departures from what is ecologically possible, depth perception ceases entirely, leaving only the homomorphic (diplopic) representation⁸². Phonetic perception has been shown experimentally to behave in a strikingly similar way^{33,83,84}. Ecologically impossible speech-like stimuli were created by dividing the signal into two parts that could not have come from the

same source. One part presented as sinusoids the cues critical for the distinction between two consonants. These were connected to an acoustic base that conformed to the normal resonances of speech. In isolation, the sinusoids sounded like non-speech whistles that differed in pitch. The base was perceived as a consonant–vowel syllable, but in the absence of the critical cues, the consonant was ambiguous. To control the evidence for two independent sources, and thus for ecological plausibility, the experimenters varied the intensity of the sinusoids, while holding the intensity of the base constant⁸⁴. When the intensity of the sinusoids was very low, the elasticity of the module enabled it to accommodate them, with the result that the signals engaged the phonetic system and caused listeners to perceive the heteromorphic consonants correctly. But at those levels, the whistles that the sinusoids produced in isolation were identified at a chance level; indeed, they were not even heard.

It is noteworthy that an analogous result has been obtained by Eimas and Miller⁸⁵ with two- to four-month-old infants. One supposes about both cases that the phonetic percepts could hardly have been a translation from auditory percepts, because the auditory percepts had not yet become identifiable. Further increases in the intensity of the sinusoids in the experiment by Xu *et al.*⁸⁴ strained the elasticity of the phonetic module, causing the listeners to hear simultaneously (and correctly) both the consonants and the whistles. That is, exactly the same acoustic signal produced in the same brain, at the same time, two very different percepts, one distinctly phonetic, the other not. Finally, at still higher intensities, where the elasticity of the module could no longer accommodate the strong evidence for two sources, the sinusoids no longer engaged the phonetic system, producing only the homomorphic whistles and the ambiguous syllabic base.

That kind of experiment not only relates phonetic perception to stereopsis, another perceptual specialization that lacks an end organ and produces a heteromorphic percept, but also shows quite directly that the primary perceptual response to speech is phonetic and independent of its auditory counterpart. The fact that the consonants were correctly identified at levels of intensity where the whistles were not has been called ‘phonetic precedence’³³. It nicely illustrates the exquisite sensitivity of the phonetic module when it does what it was specifically adapted to do. As we have seen, that adaptation was to several requirements that language meets. Had the auditory system been bent to accommodate those requirements, it would have become useless for the purpose of rendering accurately the sounds of the non-linguistic world. The biological solution was the evolution of a distinct phonetic mode as part of the larger language mode, not in the higher reaches of the cognitive machinery, but down among the nuts and bolts of action and perception.

Conclusion

We have examined two starkly contrasting theories about the production and perception of the sounds that convey phonetic structure: the more conventional, horizontal theory, which holds that those processes begin with ordinary (non-linguistic) motor and auditory representations that are then connected by purely cognitive means to language proper; and the less conventional, vertical theory, according to which the

primary representations are immediately phonetic gestures of the articulatory apparatus, having been produced in a specialized phonetic mode that serves as the basis of the larger specialization for language. The aim of this essay was to show that the vertical theory provides the more plausible answers to important questions of a biological kind, questions that are, for unaccountable reasons, rarely asked. Thus, it is possible to see how, by creating distinctly phonetic motor structures to serve as the ultimate constituents of language, the phonetic specialization enables speech to meet the requirements for parity, as well as those for particulate communication, while also giving it a biological advantage over the reading and writing of its alphabetic transcription.

We have yet to discover exactly how the phonetic motor structures find expression as coordinated movements of the articulators; how, despite elaborate overlapping and interleaving, they are organized into precisely bounded segments; and how the inverse transform from sound to motor structure is accomplished. A consequence of these gaps in our knowledge is that we presently claim for the vertical view only that it heads the theoretical enterprise in the right direction.

Acknowledgements

The writing of this paper was supported by NIH grants HD-01994, DC-02717 and DC-00403 to Haskins Laboratories. We thank Carol A. Fowler, Michael Studdert-Kennedy, Len Katz, Mark Liberman, Min Kang, Richard Ivry, Michael Turvey and Donald Shankweiler for helpful comments.

References

- 1 Fodor, J.A. (1983) *The Modularity of Mind*, MIT Press
- 2 Blumstein, S.E. and Stevens, K.N. (1979) Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *J. Acoust. Soc. Am.* 66, 1001–1017
- 3 Kurovski, K. and Blumstein, S.E. (1984) Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *J. Acoust. Soc. Am.* 76, 383–390
- 4 Stevens, K.N. et al. (1992) Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *J. Acoust. Soc. Am.* 91, 2979–3000
- 5 Lindblom, B.E. (1991) The status of phonetic gestures. In *Modularity and the Motor Theory of Speech Perception* (Mattingly, I.G. and Studdert-Kennedy, M., eds), pp. 7–24, Erlbaum
- 6 Sussman, H. (1989) Neural coding of relational invariance in speech perception: human language analogs to the barn owl. *Psychol. Rev.* 96, 631–642
- 7 Diehl, R.L. and Kluender, K.R. (1989) On the objects of speech perception. *Ecol. Psychol.* 1, 121–144
- 8 Kuhl, P.K. and Miller, J.D. (1975) Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* 190, 69–72
- 9 Stevens, K.N. (1981) Constraints imposed by the auditory system on the properties used to classify speech sounds: evidence from phonology, acoustics, and psychoacoustics. In *The Cognitive Representation of Speech* (Myers, T. et al., eds), pp. 61–74, Elsevier
- 10 Darwin, C.J. (1984) Auditory processing and speech perception. In *Attention and Performance* (Vol. X) *Control of Language Processes* (Bouma, H. and Bouwhuis, D.G., eds), pp. 197–209, Erlbaum
- 11 Kluender, K.R. et al. (1988) Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *J. Phonet.* 16, 153–169
- 12 Bates, E. et al. (1988) *From First Words to Grammar: Individual Differences and Dissociable Mechanisms*, Cambridge University Press
- 13 Bloomfield, L. (1933) *Language*, Holt
- 14 Piaget, J. (1952) *The Origins of Intelligence in Children*, International Universities Press
- 15 Skinner, B.F. (1957) *Verbal Behavior*, Appleton-Century-Crofts
- 16 Klatt, D.H. (1980) Speech perception: a model of acoustic-phonetic analysis and lexical access. In *Perception and Production of Fluent Speech* (Cole, R.A., ed.), pp. 243–288, Erlbaum
- 17 Massaro, D.W. (1987) *Speech Perception by Ear and Eye: A Paradigm for Psychological Enquiry*, Erlbaum
- 18 Jakobson, R. et al. (1963) *Preliminaries to Speech Analysis*, MIT Press
- 19 Fujisaki, H. and Kawashima, T. (1971) *A Model of the Mechanisms for Speech Perception: Quantitative Analysis of Categorical Effects in Discrimination*, Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo
- 20 Pisoni, D.B. (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept. Psychophys.* 13, 253–260
- 21 Liberman, A.M. (1996) *Speech: A Special Code*, MIT Press
- 22 Liberman, A.M. and Mattingly, I.G. (1985) The motor theory of speech perception revised. *Cognition* 21, 1–36
- 23 Liberman, A.M. et al. (1967) Perception of the speech code. *Psychol. Rev.* 74, 431–461
- 24 Browman, C.P. and Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology* 6, 201–251
- 25 Fowler, C.A. and Smith, M. (1986) Speech perception as 'vector analysis': an approach to the problems of invariance and segmentation. In *Invariance and Variability in Speech Processes* (Perkell, J.S. and Klatt, D.H., eds), pp. 123–136, Erlbaum
- 26 Turvey, M.T. et al. (1978) Issues in the theory of action: degrees of freedom, coordinative structures and coalitions. In *Attention and Performance* (Vol. VII) (Requin, J., ed.), pp. 557–595, Erlbaum
- 27 Fowler, C.A. et al. (1980) Implications for speech production of a general theory of action. In *Language Production* (Butterworth, B., ed.), pp. 373–420, Academic Press
- 28 Kelso, J.A.S. (1980) *Understanding Human Motor Behavior*, Human Kinetics
- 29 Turvey, M.T. (1990) Coordination. *Am. Psychol.* 45, 938–953
- 30 Studdert-Kennedy, M. (1998) The particulate origins of language generativity: from syllable to gesture. In *Approaches to the Evolution of Language: Social and Cognitive Bases* (Hurford, J. et al., eds), pp. 202–221, Cambridge University Press
- 31 Lieberman, P. (1984) *The Biology and Evolution of Language*, Harvard University Press
- 32 Mattingly, I.G. and Liberman, A.M. (1990) Speech and other auditory modules. In *Signal and Sense: Local and Global Order in Perceptual Maps* (Edelman, G.M. et al., eds), pp. 501–519, John Wiley & Sons
- 33 Whalen, D.H. and Liberman, A.M. (1987) Speech perception takes precedence over non-speech perception. *Science* 237, 169–171
- 34 Fowler, C.A. and Saltzman, E.L. (1993) Coordination and co-articulation in speech production. *Lang. Speech* 36, 171–195
- 35 Fowler, C.A. (1986) An event approach to the study of speech perception from a direct-realist perspective. *J. Phonet.* 14, 3–28
- 36 Fowler, C.A. (1994) Invariants, specifiers, cues: an investigation of locus equations as information for place of articulation. *Percept. Psychophys.* 55, 597–610
- 37 Rizzolatti, G. and Arbib, M.A. (1998) Language within our grasp. *Trends Neurosci.* 21, 188–194
- 38 Mattingly, I.G. and Liberman, A.M. (1988) Specialized perceiving systems for speech and other biologically significant sounds. In *Auditory Function* (Edelman, G.M. et al., eds), pp. 775–793, John Wiley & Sons
- 39 Remez, R.E. et al. (1994) On the perceptual organization of speech. *Psychol. Rev.* 101, 129–156
- 40 Remez, R.E. et al. (1981) Speech perception without traditional speech cues. *Science* 212, 947–950
- 41 de Saussure, F. (1959) *Course in General Linguistics*, McGraw-Hill
- 42 Hockett, C.F. (1958) *A Course in Modern Linguistics*, Macmillan
- 43 Jakobson, R. (1970) Linguistics. *Main Trends Res. Soc. Hum. Sci.* 1, 419–463
- 44 Studdert-Kennedy, M. Evolutionary implications of the particulate principle: imitation and the dissociation of phonetic form from semantic function. In *The Emergence of Language: Social Function and the Origins of Linguistic Form* (Knight, C. et al., eds), Cambridge University Press (in press)
- 45 Abler, W. (1989) On the particulate principle of self-diversifying systems. *J. Soc. Biol. Struct.* 12, 1–13
- 46 Maclay, H. and Osgood, C.E. (1959) Hesitation phenomena in spontaneous English speech. *Word* 15, 19–44
- 47 Harris, C.M. (1953) A study of the building blocks in speech. *J. Acoust. Soc. Am.* 25, 962–969

- 48 Nearey, T.M. (1997) Speech perception as pattern recognition. *J. Acoust. Soc. Am.* 101, 3241–3254
- 49 Whalen, D.H. and Samuel, A.G. (1985) Phonetic information is integrated across intervening non-linguistic sounds. *Percept. Psychophys.* 37, 579–587
- 50 Repp, B.H. and Mann, V.A. (1981) Perceptual assessment of fricative-stop co-articulation. *J. Acoust. Soc. Am.* 69, 1154–1163
- 51 Mann, V.A. and Repp, B.H. (1980) Influence of vocalic context on the perception of /S/–/s/ distinction: I. Temporal factors. *Percept. Psychophys.* 28, 213–228
- 52 Strange, W. (1987) Information for vowels in formant transitions. *J. Mem. Lang.* 26, 550–557
- 53 Strange, W. et al. (1979) Acoustic and phonological factors in vowel identification. *J. Exp. Psychol. Hum. Percept. Perform.* 5, 643–656
- 54 Verbrugge, R.R. and Rakerd, B. (1986) Evidence of talker-independent information for vowels. *Lang. Speech* 29, 39–57
- 55 Liberman, A.M. et al. (1957) The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368
- 56 Crowder, R.G. (1973) Precategorical acoustic storage for vowels of short and long duration. *Percept. Psychophys.* 13, 502–506
- 57 Pastore, R.E. et al. (1977) Common-factor model of categorical perception. *J. Exp. Psychol. Hum. Percept. Perform.* 3, 686–696
- 58 Delattre, P.C. et al. (1955) Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Am.* 27, 769–773
- 59 Liberman, A.M. et al. (1952) The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Am. J. Psychol.* 65, 497–516
- 60 Lisker, L. (1986) 'Voicing' in English: a catalogue of acoustic features signalling /b/ versus /p/ in trochees. *Lang. Speech* 29, 3–11
- 61 Green, K.P. and Miller, J.L. (1985) On the role of visual rate information in phonetic perception. *Percept. Psychophys.* 38, 269–276
- 62 Summerfield, Q. (1982) Differences between spectral dependencies in auditory and phonetic temporal processing: relevance to the perception of voicing in initial stops. *J. Acoust. Soc. Am.* 72, 51–61
- 63 Abramson, A.S. and Lisker, L. (1970) Discriminability along the voicing continuum: cross-language tests. In *Proc. 6th Int. Congr. Phonetic Sci.* (Hála, B. et al., eds), pp. 569–573, Academia
- 64 Lahiri, A. et al. (1984) A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: evidence from a cross-language study. *J. Acoust. Soc. Am.* 76, 391–404
- 65 Werker, J.F. and Tees, R.C. (1984) Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am.* 75, 1866–1878
- 66 Liberman, A.M. et al. (1958) Some cues for the distinction between voiced and voiceless stops in initial position. *Lang. Speech* 1, 153–167
- 67 Best, C.T. et al. (1981) Perceptual equivalence of acoustic cues in speech and non-speech perception. *Percept. Psychophys.* 29, 191–211
- 68 Fitch, H.L. et al. (1980) Perceptual equivalence of two acoustic cues for stop-consonant manner. *Percept. Psychophys.* 27, 343–350
- 69 Liberman, A.M. (1998) When theories of speech meet the real world. *J. Psycholinguist. Res.* 27, 111–122
- 70 Liberman, A.M. (1989) Reading is hard just because listening is easy. In *Brain and Reading (Wenner-Gren Symposium Series 54)* (von Euler, C. et al., eds), pp. 197–205, Macmillan
- 71 Liberman, I.Y. and Shankweiler, D.P. (1979) Speech, the alphabet, and teaching to read. In *Theory and Practice of Early Reading* (Resnick, L.B. and Weaver, P.A., eds), pp. 109–132, Erlbaum
- 72 Shankweiler, D. et al. (1979) The speech code and learning to read. *J. Exp. Psychol. Hum. Percept. Perform.* 5, 531–545
- 73 Byrne, B. and Fielding-Barnsley, R. (1989) Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *J. Educ. Psychol.* 81, 313–321
- 74 Brady, S.A. et al. (1983) Speech perception and memory coding in relation to reading ability. *J. Exp. Child Psychol.* 35, 345–367
- 75 Brady, S.A. and Shankweiler, D.P., eds (1991) *Phonological Processes in Literacy: A Tribute to Isabelle Y. Liberman*, Erlbaum
- 76 Brady, S. et al. (1994) Training phonological awareness: a study with inner-city kindergarten children. *Ann. Dyslexia* 44, 26–59
- 77 Byrne, B. (1996) The learnability of the alphabetic principle: children's initial hypotheses about how print represents spoken language. *Appl. Psycholinguist.* 17, 401–426
- 78 Byrne, B. and Liberman, A.M. (1999) Meaninglessness, productivity, and reading: some observations about the relation between the alphabet and speech. In *Reading Development and the Teaching of Reading: A Psychological Perspective* (Oakhill, J. and Beard, R., eds), pp. 157–173, Blackwell
- 79 Klueder, K.R. and Greenberg, S. (1989) A specialization for speech perception? *Science* 244, 1530
- 80 Berkeley, G. (1709) *An Essay Towards a New Theory of Vision*, Aaron Rhames for Jeremy Pepyal
- 81 Konishi, M. (1978) Ethological aspects of auditory pattern recognition. In *Handbook of Sensory Physiology: Perception* (Vol. VIII) (Held, R. et al., eds), pp. 289–309, Springer-Verlag
- 82 Richards, W. (1971) Anomalous stereoscopic depth perception. *J. Opt. Soc. Am.* 61, 410–414
- 83 Vorperian, H.K. et al. (1995) Stimulus intensity and fundamental frequency effects on duplex perception. *J. Acoust. Soc. Am.* 98, 734–744
- 84 Xu, Y. et al. (1997) On the immediacy of phonetic perception. *Psychol. Sci.* 8, 358–362
- 85 Eimas, P.D. and Miller, J.D. (1992) Organization in the perception of speech by young infants. *Psychol. Sci.* 3, 340–345

Editor's note

Sadly, Alvin Liberman passed away during the final stages of the preparation of this manuscript. The paper was under revision at the time of his final illness, and it is therefore published with only minor changes (as recommended by the referees and with the co-author's agreement) as a lasting testimony to his contribution to the field.

TICS online – making the most of your personal subscription

- High quality printouts (from PDF files)
- Links to other articles, other journals and cited software and databases

All you have to do is:

- Obtain your subscription key from the address label of your print subscription
- Then go to http://www.trends.com/free_access.html
- Click on the large 'Click Here' button at the bottom of the page and you will see one of the following:
 - (1) A BioMedNet login screen. If you see this, please enter your BioMedNet username and password. If you are not a member, please click on the 'Join Now' button and register. Once registered you will go straight to (2) below.
 - (2) A box to enter a subscription key. Please enter your subscription key here and click on the 'Enter' button.
- Once confirmed, go to <http://tics.trends.com> and view the full text of TICS

If you get an error message please contact Customer Services (info@current-trends.com) stating your subscription key and BioMedNet username and password. Please note that you do not need to re-enter your subscription key for TICS – BioMedNet 'remembers' your subscription. Institutional online access is available at a premium.

If your institute is interested in subscribing to print and online please ask them to contact ct.subs@rbi.co.uk