Evidence for a Special Speech-Perceiving Subsystem in the Human

A. M. Liberman and D. B. Pisoni
Haskins Laboratories Inc.
270 Crown Street, New Haven, CT 06511, USA

Abstract. If we want to discover whether man is specialized to process speech so as to recover phonetic segments, we must, of course, make the appropriate comparisons with nonhuman animals. To promote that undertaking, we here identify a distinctive characteristic of phonetic (as opposed to auditory) perception, and we describe some phenomena of human speech perception, appropriate for testing with animals, in which that characteristic seems to be exhibited. The distinctive characteristic is that the perceptual process is constrained as if by 'knowledge' of what vocal tracts do when they make linguistically significant gestures. The distinctive phenomena are taken from instances of stop-consonant perception. There, the role of a necessary acoustic cue--silence--is to inform the listener that the speaker closed his vocal tract, as he must if he is to produce a stop consonant; and the equivalence in perception of very different acoustic cues--temporal vs. spectral, for example--is to be accounted for on the ground that, though presumably unrelated in auditory perception, they are the distributed results of the same articulatory gesture.

If it were possible to perceive the words of language simply as auditory patterns--that is, without regard to their constituent phonetic elements--then neither phonetic structure nor its perception would be of great biological interest. But such a nonphonetic strategy would, in practice, severely limit the number of words a listener could identify and immensely complicate the processes by which he extracts those words from the stream of speech.

INTRODUCTION

Our assignment is to ask whether biologically special processes might be necessary for perceiving the phonetic structure of language. It seems appropriate, then, to consider those facts about speech and its perception that imply the need for such processes, and to sharpen the point by imagining experiments with nonhuman animals and human infants aimed at finding out whether they hear speech as we do.

Among those processes that might be specialized for the perception of speech, there are at least two kinds. One includes specializations of the auditory system that would serve to extract from the complex speech signal just those parts that carry the linguistically relevant information. We should wonder whether such devices exist if only because there appears to be a need for them: paradoxically, some of the acoustic cues that underlie important linguistic distinctions are among those aspects of the physical signal that are least salient. If there are specializations of that kind, they might be similar to the feature detectors that have been claimed for so many animals. At all events, they would properly belong to the auditory system, however specialized for speech they might be, because they would succeed only in clarifying the signal; they would not decode it. There would remain a peculiar relation between auditory pattern and phonetic message, a relation similar in form and function to those grammatical codes that connect other levels of language in the further reaches of phonology and syntax. Conceivably, there are devices specialized to cope with that peculiar relation, and thus to recover the phonetic message from the sound. If so, they would presumably be different in kind from the specialized auditory devices we have imagined. Indeed, such specialized devices would likely be an integral part of the larger and equally special physiology that comprehends all of the grammatical link between sound and meaning. Hence, we will distinguish specialized grammatical decoders from auditory specializations; the grammatical decoders we are concerned with in this paper would most properly be considered to be specializations of a phonetic sort.

If auditory or phonetic specializations for language do exist, we should expect that, given the appropriate experimental tests, the responses of nonhuman animals would be different from ours; the responses of human infants would, of course, depend additionally on the way those specializations are affected by experience. To promote consideration of how we might, in any case, do relevant research, we will identify several classes of findings with adult human beings that suggest what some of the animal and infant tests might be. But, given the limitations on the length of this paper we will concern ourselves only with the question: Are there specialized phonetic processes?

To provide a proper background for our question, we should remind ourselves of two universal--hence biologically interesting--facts about language. The more obvious is that the structure of language has two aspects: one is formed of meaningful segments (words, phrases, sentences) and governed by the rules of syntax; the other comprises segments that are empty of meaning (phones, syllables, breath groups) and subject to the lawful constraints of phonology and phonetics. Less obvious, but no less universal, is the fact that the shortest of the meaningless segments--the phones, or consonants and vowels, that are the objects of our attention here--are not directly reflected in the sound stream. That is so because of the universal occurrence of coarticulation, the overlapping or even simultaneous production of features from successive phonetic segments. As a consequence, information about those several phonetic segments is carried simultaneously on the same acoustic parameter. Thus, the phonetic message is encoded (not enciphered) in the sound, and in a special way: there is no direct correspondence in segmentation between message and sound, hence no acoustic (or auditory) criterion by which the speech stream can be directly divided into segments that correspond to the phones; and the acoustic cues for any particular segment will vary, often in apparently peculiar fashion, according to the other segments with which it is encoded and simultaneously conveyed.

It is the existence of that universal (among human beings) and special code that most generally bespeaks the need for special phonetic processes. Accordingly, a decent concern for the importance of putting first things first dictates that we should want most urgently to know how well nonhuman animals cope with its most general characteristics. Can they, for example, appreciate, even tacitly, that speech does consist of commutable segments, that "bad" and "dab" are simply different permutations of the same three segments, or that words like "grew" and "ilk" share no segments but have the same number? Unfortunately, the animal tests appropriate to those most general, and possibly most telling, questions are often impossible in practice, or so nearly so as to discourage even the most intrepid investigators. With that in mind, and in the hope that relevant experiments of some kind might nevertheless be done, we will set considerations of logical priority aside and give special emphasis to those less general and more simple--yet still apparently special--characteristics of phonetic perception for which the appropriate animal tests might be feasible. And in order to crowd as many of those characteristics as we can into our alloted space, we will, to the greatest extent possible, deal with a single and simple acoustic cue, silence. (For further discussion of relevant data and issues, see [1,8, 9,14,15,19-20].

SOME ACOUSTIC CUES HAVE PERCEPUTAL EFFECTS THAT MAY BE UNIQUELY PHONETIC

Consider the following easy-to-obtain facts about the way we perceive stop consonants and fricatives. Record the syllable /sa/, represented schematically in Figure 1. There you see a patch of band-limited noise, normally produced in the articulation of /s/, followed by a vocalic section. The vocalic section contains, first, the formant transitions caused by the articulatory movement from consonant position to vowel position, then the formant steady-states appropriate to a drawn-out vowel. Since the partial closure for the fricative /s/ is at approximately the same place in the vocal tract as

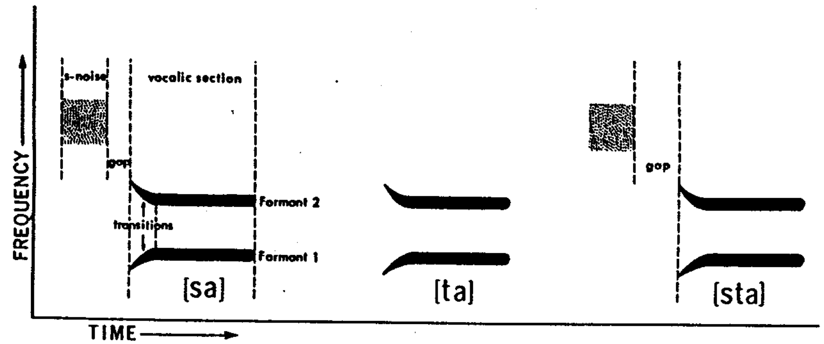

FIG. 1 - Schematic representation of variations on the syllable [sa], illustrating the role of silence in the perception of the stop consonant [t].

the total closure for the stop consonant /t/, the formant transitions of /sa/ are similar to those of /ta/. We find, then, that if we remove the patch of noise, human listeners will, indeed, commonly hear not /a/, but /ta/. Restore the noise, and they will, of course, hear /sa/ again. Move the noise backwards in time so as to open up a gap, or silent interval, of about 60 msec between it and the vocalic section, and they will hear not /sa/, but /sta/.

Thus, silence is a condition for perceiving the stop consonant. But what does silence do? From the point of view of our interests at this workshop, there are at least three possibilities. The first is that the role of silence is explicable in terms of the properties of a generalized auditory system. Consider, once again, the phenomena described above, and see that the formant transitions, which are cues for the stop consonant, might be forward-masked by the noise; in that case, the silence would provide time to evade masking. Given that kind of explanation, we should expect that animals with ears like ours would hear the syllables much as we do.

A second possibility is that we have here the result of an auditory specialization of the kind we referred to earlier. There might, for example, be detectors specialized to extract formant transitions from speech, and these might be disabled by the noise; or there might be specialized auditory devices that produce an interaction between silence and the transition cues. In either case, other animals would not hear the syllables as we do; but in human listeners the observed effects would be found in all of auditory perception, not just in speech.

The third possibility is that the perception of silence is here phonetic rather than auditory. To distinguish phonetic from auditory, we should determine whether the perceiving mechanism is constrained, not only by the properties of the ear, but also by "knowledge", as it were, of what vocal tracts do when they make linguistically significant gestures. To see how just such a constraint might be at work in our example, consider that a speaker cannot say /sta/, as against /sa/, without totally closing his vocal tract and so creating an interval of silence. Given a biologically based link between speech perception and speech production, the absence of silence might therefore signal the listener that the speaker had not closed his vocal tract long enough to have said /sta/; hence /sa/. Such phonetic perception, if it does exist, would be found only in creatures that speak, and then only when they are listening to speech.

We will shortly describe some facts about the perception of speech by human beings that, in the case of the silence cue and others, imply the existence of a phonetic mode of perception. We will look forward then to learning as much as possible about how nonhuman animals hear these same speech sounds, and how both they and human listeners hear the relevant acoustic variations in a nonspeech context. But before we abandon the simple example of /sa-sta/ that we have already offered, we should at least list the several reasons we have for sup-

posing that it is a simple case of phonetic perception: (1) Though perception of the stop can be totally blocked by the s-noise, the transition cues are nevertheless effective in promoting perception of the fricative /s/ [2,7]. Thus, as auditory events, the transition cues "get through"; it is only their (phonetic) interpretations (as /s/ or as /t/) that are affected by the nearness (or farness) of the noise. (2) When the noise of /s/ is put very close to the vocalic syllable /ka/, listeners hear neither /ska/ nor /sa/, but /sja/ (Liberman, Halwes, and Fitch, personal communication). Consider that the transition cues for the stop /k/ and the semivowel /j/ are similar except that the former are more rapid than the latter. We see, then, that the transition cues were interpreted as /ja/, not /ka/, because the gap that is so essential for the production and perception of /k/ was not there; to produce /ja/, the speaker does not close the tract totally, hence he produces no gap. Here, too, it was only the phonetic interpretation of the transitions (as /k/ or as /j/) that was affected. (3) The transition cues in a syllable perceived as /se/ are fully effective in producing selective adaptation of /de/ [6]. This is further evidence that, as auditory events, the transition cues are, in fact, being processed by the perceptual system;they are in no way blocked by the presence of the s-noise. (4) When inserted between s-noise and a vocalic section, silence is sufficient (and not merely necessary) for the perception of a stop consonant; in such cases, there are no ordinary stop-consonant cues (transitions or bursts) to be masked or detected ([4];Summerfield and Bailey, personal communication). (5) In cases like those described in (4), the "place" of the perceived stop--that is, whether it is /b/ or /g/--depends on the nature of the following vocalic section (Summerfield and Bailey, personal communication). (6) In other, analogous cases, the amount of silence necessary to produce a stop-like effect varies according to the tempo of the carrier phase [4]. (7) When the transition cues for the stop are removed from the vocalic speech contexts and presented alone, in which case they are heard as nonspeech "chirps", their identifiability is not at all masked by the preceding s-noise, nor is their perception changed in any qualitative way [5].

DIFFERENT ACOUSTIC CUES PRODUCE THE SAME PHONETIC PERCEPTION IN THE SAME POSITION AND IN THE SAME CONTEXT

We offer the example of the contrast between /slit/ and /split/, diagrammed in Figure 2 so as to show how it can be fashioned out of either of two very different cues--one spectral, the other temporal (Liberman, Halwes, and Fitch, personal communication). The spectral cue is primarily the appropriate set of formant transitions; it is present (plus /p/) in the vocalic section /plit/ and absent (minus /p/)in /lit/. The temporal cue is the silent gap by which the vocalic sections are separated from the s-noise; it is present (plus /p/) when the gap is long and absent (minus /p/) when it is short. In Pair 1, a spectral difference is sufficient to cue the contrast between /slit/ and /split/. In Pair 2, we see how that same contrast is produced by an acoustic cue that is entirely temporal. One asks, of course, what the spectral and temporal cues have in common. In acoustic and auditory terms they appear to be about as different as can be. However, from an articulatory point of view they are related: they are the results of the same gesture. Given that they have the same consequences in phonetic perception, we might suppose that in this case perception is somehow linked to production.

DIFFERENT ACOUSTIC CUES FOR THE SAME PHONETIC PERCEPTION CANCEL OR SUMMATE DEPENDING ON HOW THEY ARE COMBINED

Using the same cues described above, we find that they have opposite effects depending on how they are put together. In Pair 3, we show how to combine the spectral and temporal cues so as to produce the same phonetic distinction (/slit/ vs. /split/) that either cue alone is sufficient to make. We should note that when each cue is so near the perceived phonetic boundary as to be less than perfectly unambiguous, that combination will result in a distinction that is even more

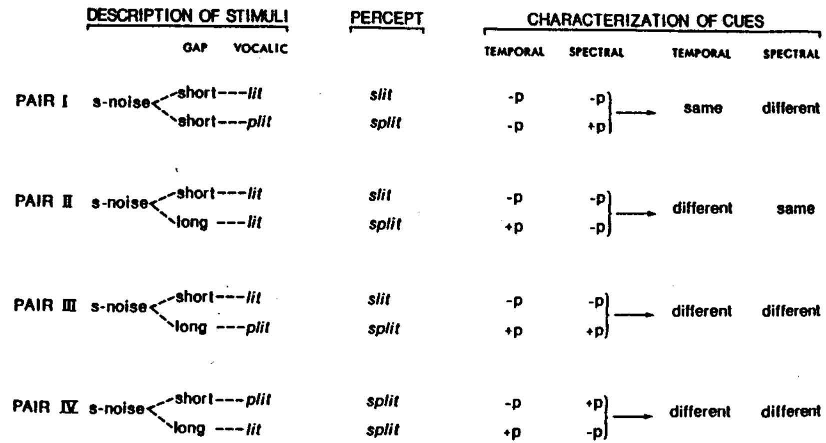

FIG. 2 - Diagrams that show how very different acoustic cues - one temporal, the other spectral - produce the same phonetic distinction, and how both cues together either enhance that distinction or reduce it depending on the way they are combined.

robust. In Pair 4, we see how those same two cues can be combined so as to decrease, not increase, the phonetic difference that is produced by either cue alone. In fact, that combination can, with proper balancing, effectively bring the difference near to zero. Thus, it is as if these cues were vectors; but however we might characterize them, we shall suppose, until animal tests prove us wrong, that their domain is not auditory but phonetic.

THE PERCEPTUAL EFFECTS OF ACOUSTIC CUES ARE SUBJECT TO ECOLOGICAL CONSTRAINTS OF AN APPARENTLY PHONETIC KIND

Acoustic cues can have one phonetic effect or another depending on whether they were produced by one speaker or by two [4,18]. It is as if the listener knew that two vocal tracts can accomplish what one vocal tract cannot. To see an example, imagine the following. We record "now say" and, separately, "shop". Then we place "shop" after "now say". If the physical characteristics of the word "shop" are held within certain limits, it is possible to change "now say shop" to "now say chop" simply by increasing the duration of silence between "say" and "shop". Of course, this is just like the other examples of silence we have described; and in this case, as in the others, we assume that the silence causes the listener to hear the affricate (in "chop") rather than the fricative (in "shop") because an appropriate amount of silence tells him that the speaker closed his vocal tract briefly, as one must to produce the affricate, and as one must not to produce the fricative.

However, two vocal tracts--one saying "now say" and the other "chop"--can produce "now say chop" with no silence at all between "say" and "chop". Thus, with two speakers, the size of the interval of silence provides no useful phonetic information. Experiments reveal that listeners behave accordingly: starting with "now say" and "shop", and given a silent interval appropriate for "chop", listeners do indeed hear "now say chop" if there was only one speaker; but if there were two speakers, then listeners hear "now say shop" at all intervals of silence.

DIFFERENT ACOUSTIC CUES PRODUCE THE SAME PHONETIC PERCEPTION IN DIFFERENT POSITIONS

Consider the voicing distinction between, for example, /b/ and /p/ in three positions in the syllable, as schematized in Figure 3. In initial position, as in /bid/ vs. /pid/, an important and sufficient cue is the so-called voice onset time (VOT), the time interval between release and start of voicing [10, 13]. In intervocalic position, as in /rabid/ vs. /rapid/, an important acoustic cue is the duration of silence between the two syllables [9]. And in final position, as in /ib/ vs. /ip/, an important and sufficient cue is the duration of the preceding vowel [17]. We should note how very different these cues are from an acoustic point of view. We should also note that in each of these cases, as in the earlier example of /slit/ vs. /split/, there are several cues, very different from each other in acoustic terms, but

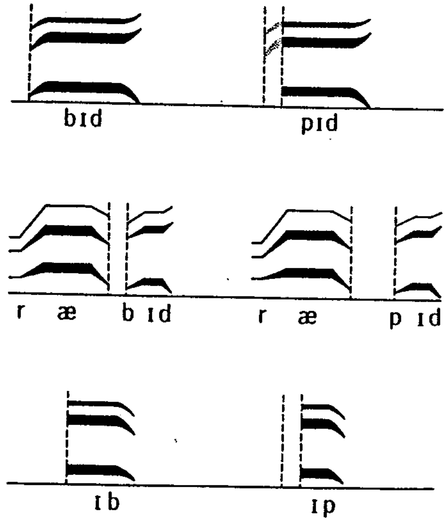

FIG. 3 - Schematic representation of the different acoustic cues that are appropriate for the same phonetic distinction according to its position in the syllable.

equivalent in phonetic perception. That is most dramatically the case with the contrast between /rabid/ and /rapid/, where the effective acoustic cues include, in addition to the duration of intersyllable silence, such variables as the duration of the first syllable, the spectral characteristics (transitions) at the end of the first syllable and at the beginning of the second syllable, the condition of syllable stress (whether trochaic or iambic), the voice onset time in the second syllable, and numerous others (Lisker, personal communication). As in the case of /slit/ vs. /split/, this diversity of acoustic cues is produced by a single articulatory contrast. That such acoustically different cues are more or less equivalent in phonetic perception is further testimony to the link between the way we perceive such phonetic distinctions and the way we produce them.

THE SAME ACOUSTIC CUES ARE PERCEIVED DIFFERENTLY ACCORDING TO THE REMOTE CONTEXT

Consider, again, the perceived difference in voicing between /rabid/ and /rapid/ as cued by the duration of silence. When experimental tokens of such syllables are placed in speech carriers that mimic different rates of articulation, listeners hear the change from /rabid/ to /rapid/ at different durations of silence.

DIFFERENT ACOUSTIC CUES PRODUCE THE SAME PHONETIC PERCEPTION IN THE SAME POSITION BUT IN DIFFERENT IMMEDIATE CONTEXTS

Unfortunately for our purposes, the silence cues do not offer telling examples in this case. We are reluctant on that account alone to omit the kind of context-conditioned variation we would here illustrate, because it is one of the most important consequences of coarticulation, hence one of the most pervasive characteristics of the speech signal. We will, then, turn away from our preoccupation with duration and the sounds of silence just long enough to present a well-worn example of what can happen when we change only one phonetic segment in a syllable that contains three. Take /did/ and /dud/, shown schematically as two-formant (synthetic) approximations in Figure 4 (Delattre, Liberman, and Cooper [3]). We should note first that the lower (first) formant is the same for the two vowels and, indeed, for the two syllables. So we put our attention on the higher (second) formants. There we find information sufficient, in combination with the common lower formant, to tell us that the vowels are /i/ and /u/ and that the consonants are all /d/, not /b/ or /g/. We see, then, that a phonetic difference limited to the middle (vowel) segment does not produce a change in the signal that is limited to the middle portion of the sound; rather, the entire formant changes. Note especially that the transition cues for /d/ are in very different parts of the spectrum for the two syllables --high for /did/ and low for /dud/. Moreover, if the starting point of the second formant for /did/ is lowered so as to coincide with the starting point for /dud/, then listeners will most likely hear, not /did/, but /bid/. Finally, we should note that for corresponding positions in the two syllables, the transitions are opposite in direction: for /did/ they are rising in initial position and falling in final position, but for /dud/ they are falling in initial position and

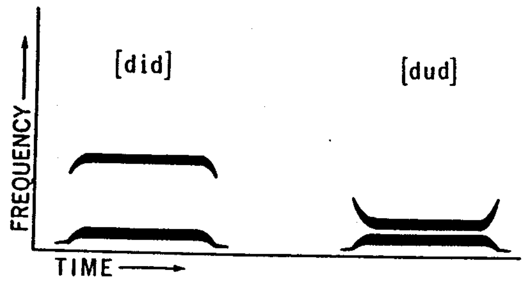

FIG. 4 - Schematic representation of the variation in consonant cues as a function of the vowel with which the consonant is coarticulated.

rising in final position. Though the acoustic cues for the /d/ have very different shapes, the underlying articulatory gesture for the consonant is much the same: a complete closure of the vocal tract produced by touching the tongue tip to the alveolar ridge. The different shapes of the acoustic cues are owing, of course, to the coarticulation of the consonant gesture with the gesture appropriate for the preceding or following vowel. As for the perception, it is as if it rationalized the context-conditioned variation in the acoustic cue and recovered the common articulation.

VERY SIMILAR ACOUSTIC CUES PRODUCE VERY DIFFERENT PHONETIC PERCEPTIONS ACCORDING TO THE CONTEXT

Having considered cases in which cues that lie on different acoustic dimensions have the same phonetic effect, we will now look at a case in which cues on a single acoustic dimension produce perceived contrasts on each of three phonetic dimensions: manner, voicing, and place. We have already seen that the presence or absence of an appropriate silence between the noise of /s/ and the syllable /lit/ will cue the phonetic distinction of manner between /split/ and /slit/. We have also noted that when those same intervals of silence are introduced between the syllables /ra/ and /bid/, they will produce the phonetic distinction of voicing between /rabid/ and /rapid/. We add now that further reductions in the duration of the silence cause the listener to hear a change of place from /rabid/ to /ratid/ [16]. It should be noted that, though the acoustic cues for all three contrasts are nominally on a single physical dimension (duration of silence), the articulatory maneuvers that produce them are not; manner, voicing, and place distinctions are different gestures made by different sets of muscles. The "place" change from /rabid/ to /ratid/ is especially interesting in that connection. As Port [16] found, speakers normally close for a significantly shorter time (hence show a shorter silent interval) when producing the "flap" /t/ in /ratid/ than when producing the labial /b/ in /rabid/. Given that listeners report hearing /ratid/ at the very short silent intervals (even though the spectral cues were appropriate for /rabid/), we should suppose they are once again honoring the extent to which phonetic perception is constrained by tacit knowledge of what a vocal tract can and cannot do when it makes linguistically significant gestures: it is as if the perceptual system 'knew' that the speaker could not have said /rabid/ because his vocal tract had not closed long enough for that.

THE PERCEPTION OF ACOUSTIC CUES IS DIFFERENT IN SPEECH AND NONSPEECH CONTEXTS

Once again we can use the example of the silent gap as a cue for the distinction between /rabid/ and /rapid/. Imagine (1) a set of (speech) stimuli in which the size of the gap is changed in relatively small steps from /rabid/ to /rapid/ and (2) a corresponding set of (nonspeech) stimuli in which the same gaps separate two bursts of noise shaped so as to correspond in amplitude and duration to the syllables of the speech stimuli. For human listeners, discrimination of the gaps is different in the two cases, being more nearly categorical in the speech stimuli [11]. Of course, the cue is (necessarily) in different contexts, as it is in all the speech-nonspeech comparisons that have been made. That is all the more reason to make the same comparisons with animals and infants (see Pisoni [15] for further discussion)

CAN SPEECH BE PERCEIVED WITHOUT REGARD FOR ITS PHONETIC STRUCTURE?

The various examples we have dealt with might test whether adult humans, animals, and infants do, in fact, process the speech signal so as to recover the phonetic structure that is encoded in it. But what if it is true, as some think, that the listener must recover the phones only if he wants to spell or rhyme or alliterate or do something equally elitist? When he is just listening to speech, and trying only to extract whatever meaning it may contain, does he skip the phonology altogether? Can he, in that case, deal directly with the meaningful segments--words are surely the most likely candidates--as holistic auditory patterns? Plainly, it must be possible to do that, but only within limits. We should say what two of those limits might be.

The first limit would be on the number of words that could be identified. Words do have internal phonetic structure, after all, and distinctions among them commonly depend on rearrangements of the constituent phonetic elements. But the acoustic criteria that are appropriate for one class of phonetic segments are not ordinarily appropriate for others, and, within a class, complex adjustments must be made to accommodate the variations with position, context, and speaking rate that we have referred to earlier. Thus, a procedure that works on the auditory patterns only as auditory patterns--that is, in disregard of their constituents--will presumably fail before it has identified all the patterns.

The second limit set by a failure to appreciate phonetic structure derives from the fact that in normal, fluent speech, coarticulation does not respect word boundaries. It follows then that, just as there is no acoustic criterion that divides speech into phone-size units, so also is there none that will reliably divide it into segments that correspond to words. It should be emphasized that this problem is not trivial: as mentioned earlier, a difficulty caused by coarticulation is that information about several successive phonetic segments is carried simultaneously in the acoustic signal and on the same parameter; therefore, the phonetic segments cannot be recovered by simply cutting the continuous acoustic signal into discrete segments. Consider, then, the plight of a creature whose stored lexicon is defined only in auditory terms: applying acoustic or auditory criteria to the stream of speech, he will recover (auditory) segments that bear a random relation to the word-size segments stored in his lexicon; therefore, the number of items he must store is not equal to the several tens of thousands of words, but is rather incalculably larger than that. To get along with a store that comprises only the number of words he knows, the listener must divide the speech stream into segments whose boundaries can be coterminous with the words. Only phonetic segments--or, more properly, their underlying phonologic forms--meet that requirement. If an animal cannot recover the phonetic structure, then he should often have difficulty retrieving the words of his vocabulary from fluent speech. Hence, we should suppose that a creature may not bypass the phonetic structure if he would perceive most of what is said to him.

REFERENCES

[1] Darwin, C.J. 1967. The Perception of Speech. In Handbook of Perception: Vol. 7, ed. by E.C. Carterette and M.P. Freidman. (New York: Academic Press), pp. 175-226.

[2] Darwin, C.J. 1971. Ear differences in the recall of fricatives and vowels. Quart. J. Exp. Psych. 23: 46-62.

[3] Delattre, P.C.; Liberman, A.M.; ånd Cooper, F.S. 1955. Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Am. 27: 769-773.

[4] Dorman, M.F.; Raphael, L.J.; and Liberman, A.M. 1976. Further observations on the role of silence as a cue for stop consonants. J. Acoust. Soc. Am. 59, Suppl. 1: S40 (A).

[5] Dorman, M.F.; Raphael, L.J.; Liberman, A.M.; and Repp, B. 1975. Some maskinglike phenomena in speech perception. J. Acoust. Soc. Am. 57, Suppl. 1: S48(A). Full text in Haskins Laboratories Status Report on Speech Research SR-42/43: 265-276.

[6] Ganong, W.F. 1975. An experiment on "phonetic adaptation." Progress Report (Cambridge: MIT Research Laboratory of Electronics) 116: 206-210.

[7] Harris, K.S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech 1: 1-7.

[8] Liberman, A.M. 1974. The specialization of the language hemisphere. In The Neurosciences: Third Study Program, ed. by F.O. Schmitt and F.G. Worden. (Cambridge, Mass.: MIT Press) 43-56.

[9] Liberman, A.M.; Cooper, F.S.; Shankweiler, D.; and Studdert-Kennedy, M. 1967. Perception of the speech code. Psychol. Rev. 74: 431-461.

[10] Liberman, A.M.; Delattre, P.C.; and Cooper, F.S. 1958. Some cues for the distinction between voiced and voiceless stops in initial positions. Language and Speech 1: 153-167.

[11] Liberman, A.M.; Harris, K.S.; Eimas, P.; Lisker, L.; and Bastian, J. 1961. An effect of learning on speech perception: the discrimination of durations of silence with and without phonetic significance. Language and Speech 4: 175-195.

[12] Lisker, L. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. Language 33: 42-49.

[13] Lisker, L. and Abramson, A.S. 1964. A cross-language study of voicing in initial stops: acoustical measurements. Word 20: 384-422.

[14] Pisoni, D.B. (In press) Speech perception. In Handbook of Learning and Cognitive Processes, ed. by W.K. Estes. (Hillsdale, N.J.: Lawrence Erlbaum Assoc.).

[15] Pisoni, D.B. (In press) Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. J. Acoust. Soc. Am.

[16] Port, R. 1976. Influence of tempo on the closure interval cue to the voicing and place of intervocalic stops. J. Acoust. Soc. Am. 59, Suppl. 1: S41(A).

[17] Raphael, L.J. 1972. Preceding vowel duration as a cue to the perception of word-final consonants in American English. J. Acoust. Soc. Am. 51: 1296-1303.

[18] Raphael, L.J.; Dorman, M.F.; and Liberman, A.M. 1976. Some ecological constraints on the perception of stops and affricates. J. Acoust. Soc. Am. 59, Suppl. 1: S25 (A).

[19] Studdert-Kennedy, M. 1976. Speech Perception. In Contemporary Issues in Experimental Phonetics, ed. by N.J. Lass. (New York: Academic Press) pp. 243-293.

[20] Stevens, K.N. 1975. The potential role of property detectors in the perception of consonants. In Auditory Analysis and Perception of Speech, by G. Fant and M.A.A. Tatham. (New York: Academic Press) pp. 303-330.

[21] Stevens, K.N., and House, A.S. 1972. Speech perception. In Foundations of Modern Auditory Theory, Vol. 2, ed. by J. Tobias. (New York: Academic Press) pp. 1-62.