DOCUMENT RESUME

ED 058 776 FL 002 832

| | |
|---|---|
| AUTHOR | Mattingly, Ignatius G. |
| TITLE | Speech Cues and Sign Stimuli. |
| INSTITUTION | Haskins Labs., New Haven, Conn. |
| SPONS AGENCY | Cambridge Univ. (England). King's Coll.; Department of State, Washington, D.C. Board of Foreign Scholarships. |
| REPORT NO | SR-27-71 |
| PUB DATE | 71 |
| NOTE | 23p.; In "Speech Research," July 1-September 30, 1971 |
| EDRS PRICE | MF-$0.65 HC-$3.29 |
| DESCRIPTORS | Acoustics; *Animal Behavior; Articulation (Speech); Artificial Speech; Auditory Discrimination; Auditory Perception; Behavior Patterns; Information Processing; *Intellectual Development; *Language Development; Language Patterns; Language Research; Neurological Organization; Phonetics; Psycholinguistics; *Signs; Spectrograms; *Speech; Stimuli |

ABSTRACT

Parallels between sign stimuli and speech cues suggest some interesting speculations about the origins of language. Speech cues may belong to the class of human sign stimuli which, as in animal behavior, may be the product of an innate releasing mechanism. Prelinguistic speech for man may have functioned as a social-releaser system. Human language developed as a result of the intellect, which was capable of making a semantic representation of the world of experience and the phonetic social-releaser system. Linguistic capacity--the ability to learn the grammar of a language--was also necessary. Grammar evolved to interrelate the semantic product of the intellect and the phonetic product of the prelinguistic communication system. References are included. (Author/VM)

# Speech Cues and Sign Stimuli

Ignatius G. Mattingly+
Haskins Laboratories, New Haven

The perception of the linguistic information in speech, as investigations carried on over the past twenty years have made clear, depends not on a general resemblance between presently and previously heard sounds but on a quite complex system of acoustic cues which has been called by Liberman et al. (1967) the "speech code." These authors suggest that a special perceptual mechanism is used to detect and decode the speech cues. I wish to draw attention here to some interesting formal parallels between these cues and a well-known class of animal signals, "sign stimuli," described by Lorenz, Tinbergen, and others. These formal parallels suggest some speculations about the original biological function of speech and the related problem of the origin of language.

A speech cue is a specific event in the acoustic stream of speech which is important for the perception of a phonetic distinction. A well-known example is the second-formant transition, a cue to place of articulation. During speech, the formants (i.e., acoustical resonances) of the vocal tract vary in frequency from moment to moment depending on the shape and size of the tract (Fant, 1960). When the tract is excited (either by periodic glottal pulsing or by noise) these momentary variations can be observed in a sound spectrogram. During the transition from a stop consonant, such as [b,d,g,p,k], to a following vowel, the second (next to lowest in frequency) formant (F2) moves from a frequency appropriate for the stop towards a frequency appropriate for the vowel; the values of these frequencies depend mainly on the position of the major constriction of the vocal tract in the formation of each of the two sounds. Since there is no energy in most or all of the acoustic spectrum until after the release of the stop closure, the earlier part of the transition will be neither audible nor observable. But the slope of the later part, following the release, is audible and can be observed (see the transition for [b] in the spectrogram for [bɛ] in the upper portion of Figure 1). It is also a sufficient cue to the place of articulation of the preceding stop: labial [b,p], alveolar [d,t], or velar [g,k]. It is as if the listener, given the final part of the F2 transition, could extrapolate back to the consonantal frequency or locus (Delattre et al., 1955).

+Also University of Connecticut, Storrs.

Acknowledgments: The preparation of this paper was supported in part by the Fulbright-Hays Commission and the Provost and Fellows of King's College, Cambridge. I also wish to acknowledge gratefully encouragement and criticism given by G. Evelyn Hutchinson, Alvin M. Liberman, Benjamin Sachs, Jacqueline Sachs, Michael Studdert-Kennedy, Philip Lieberman, Alison Jolly, Mark Haggard, Adrian Fourcin, and Dennis Fry, but responsibility for errors is mine.

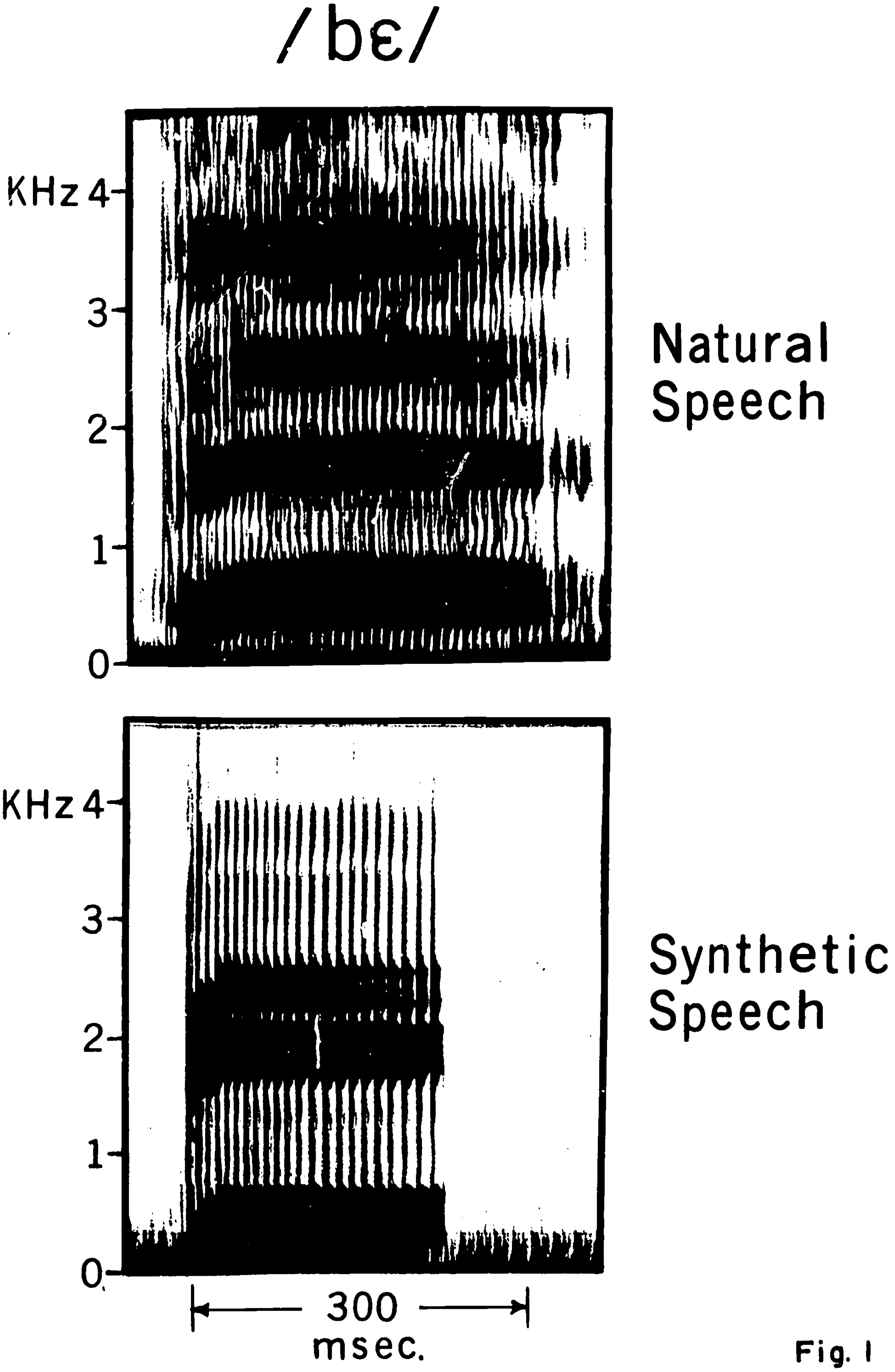


Fig. I

Spectrograms of Natural and Synthetic Speech for [bɛ]

It is possible electronically to synthesize speech which is intelligible, even though it has much simpler spectral structure than natural speech (Cooper, 1950; Mattingly, 1968). In the lower portion of Figure 1 is shown a spectrogram of a synthetic version of the syllable [bɛ]. Synthetic speech can be used to demonstrate the value of a cue such as the F2 transition by generating a series of stop-vowel syllables for which the slope of the audible part of the F2 transition is the only variable, and other cues to position of articulation, such as the frequency of the burst of noise following the release of the stop, or the slope of F3, are absent or neutralized (Cooper et al., 1952). A syllable in a series such as this will be heard as beginning with a labial, an alveolar, or a velar stop depending entirely on the slope of the F2 transition. This is true even though the slope values appropriate for a particular stop consonant depend on the vowel: thus a rising F2 cues [d] before [i], and a falling F2, [d] before [u] (see the patterns in Figure 3).

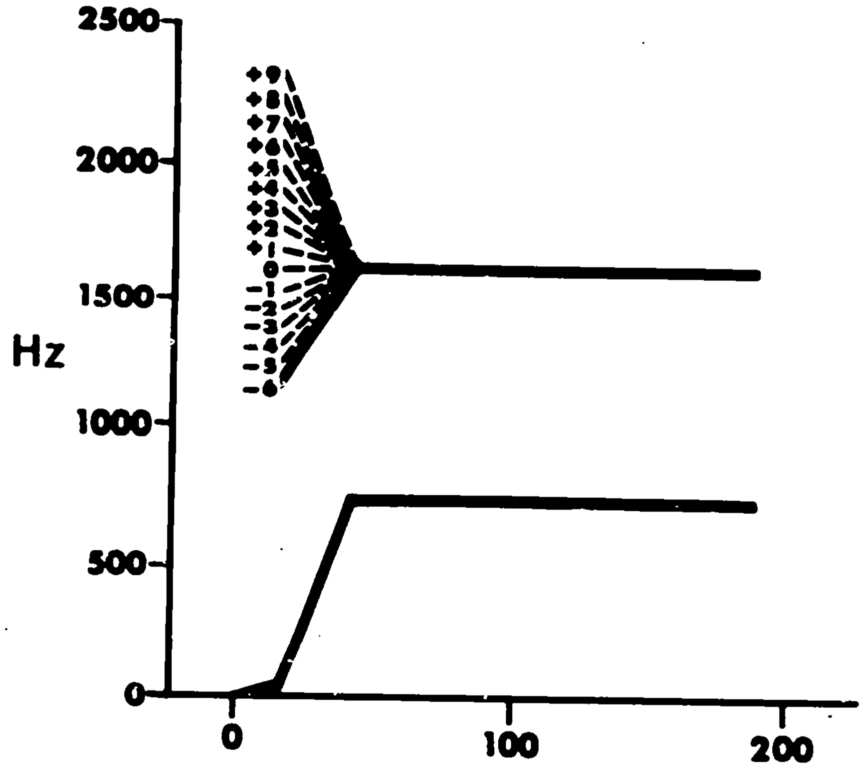
Phonetic distinctions other than place are signalled by other cues. Thus, in English, the cue separating the voiceless, aspirated stops [p,t,k] from the voiced stops [b,d,g] is voice-onset time (Liberman et al., 1958). If the beginning of glottal pulsing coincides with, or precedes, the release, the stop will be heard as [b], [d], or [g], depending upon the cues to place of articulation; if the pulsing is delayed 30 msec or more after the release, the stop will be heard as [p], [t], or [k]. Again, the duration of the formant transitions is a cue for the stop-semivowel distinction (e.g., [b] vs. [w]) (Liberman et al., 1956). A shorter (30-40 msec) transition will be heard as a stop, whereas a longer (60-80 msec) transition will be heard as a semivowel.

Some recent work indicates that human beings may possibly be born with knowledge of these cues. While appropriate investigations have not yet been carried out for most of the cues, the facts with respect to voice-onset time are rather suggestive. Not all languages have this distinction between stops with immediate voice onset and stops with voice onset delayed after release, but for all those that do, the amount of delay required for a stop to be heard as voiceless rather than voiced is about the same (Lisker and Abramson, 1970; Abramson and Lisker, 1970). This constraint on perception thus appears to be a true language universal, and so likely to reflect a physiological limitation rather than a learned convention.

Exploring the question more directly, Eimas et al. (1970), by monitoring changes in the sucking rate of one-month-old infants listening to synthetic speech stimuli, showed that the infants could distinguish significantly better between two stop-vowel stimuli which straddle the critical value of voice-onset time than between two stimuli which do not, even though the absolute difference in voice-onset time is the same. Thus the information required to interpret at least one speech cue appears either to be learned with incredible speed or to be genetically transmitted.

Sign stimuli, with which I propose to compare speech cues, have been defined by Russell (1943), Tinbergen (1951), and other ethologists as simple, conspicuous, and specific characters of a display which under given conditions produces an "instinctive" response: the red belly of the male stickleback, which provokes a rival to attack, or the zigzag pattern of his dance, which

Patterns for a Series of Stop-Vowel Syllables with Systematically Varied F2 Transitions



Note: F2 transitions with low starting points will cause the stop to be heard as [b], those with high starting points as [g], and those in between as [d].

Fig. 2

F1 and F2 Patterns Heard as [di] and [du], Despite the Apparent Difference in the F2 Transition

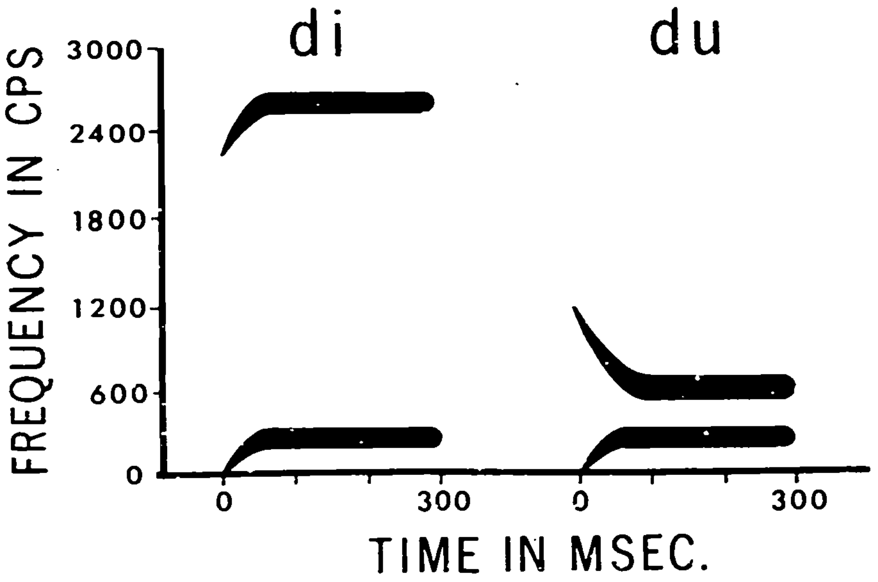


Fig. 3

arouses the female (Tinbergen, 1951); the spots by which the ringed plover identifies her eggs (Koehler and Zagarus, 1937); the red spot on the herring gull's bill, which makes her chicks beg for food (Tinbergen, 1951). These examples are visual, but sign stimuli are found in other modalities also: e.g., the monotone note of the white-throated sparrow's song, by which he asserts his territorial claims (Falls, 1969); or the chemical in the blood from a wounded minnow, which causes other minnows to flee when they scent it in the water (Manning, 1967). Responding properly to sign stimuli is normally of great value for the survival of the individual or the species. As Manning (1967:39) comments, "Sign stimuli will usually be involved where it is important never to miss making a response to the stimulus." It is this circumstance, perhaps, which accounts for the striking properties of sign stimulus perception which we shall be mainly concerned with here: the animal responds not to the display in general but specifically to the sign stimuli, and the strength of the response is in proportion to the number and conspicuousness of the sign stimuli. The perception of a sign stimulus and the response it produces have been attributed by Lorenz (1935) to a special neural "innate releasing mechanism."

The concepts of the sign stimulus and the innate releasing mechanism, as used in early ethological work, have come in for much justified criticism (e.g., Hailman, 1969; Hinde, 1970). It has been argued that sign stimuli cannot be shown to differ in principle from other stimuli; that some purported sign stimuli are not actually specific to particular responses but merely reflect the general capabilities of the animal's sense organs or associated perceptual equipment; that the word "innate" suggests too simple a dichotomy between nature and nurture; and that sign stimuli do not always lead to direct and immediate responses but influence behavior in other ways.

But when all these criticisms are taken into account, there remain some very striking phenomena. There are many cases in which a stimulus is selectively perceived by a particular species and not by others. The selectivity cannot be accounted for simply by an appeal to the general sensory capabilities of the species. The stimulus consistently elicits a direct response (or other specific behavior indicating that the stimulus has been perceived, as in the case of orientation). This response is adaptive. Moreover, in many instances (and in all the examples given above) the stimulus is a character of a display by a conspecific (or symbiotically related) individual; the entire pattern of behavior, consisting of the display and the response, is adaptive.

Displays of this latter sort have been called "social releasers" (Tinbergen, 1951:171). Their component sign stimuli elicit appropriate responses from conspecific individuals in situations important for group safety or for the integrity and continuity of the species. Social releasers include: alarm calls; the "threat behavior" of many species, by which the adaptive ends of sexual fighting are achieved with few actual casualties; the displays which serve as reproductive isolating mechanisms, encouraging intraspecific and discouraging interspecific mating; and the signs by which parents and young identify each other, so that the latter are protected and fed. In all these adaptively important situations, displays composed of sign stimuli serve to authenticate the conspecificity of individuals.

It has also been suggested before that sign stimuli actually occur in human behavior. The facial characteristics and limb movements of babies evoke parental behavior (Tinbergen, 1951). Babies, in turn, respond to adult facial characteristics, notably to eyes and to smiles, and women have a universal flirting gesture (Eibl-Eibesfeldt, 1970). I think that speech cues may also belong to the class of human sign stimuli, despite obvious differences to be discussed shortly. But let us now consider the resemblances.

First of all, the speech cues, like the sign stimuli, do not require a natural context, or even a naturalistic one; the appropriate response can be elicited by drastically simplified models of the natural original. Tinbergen's sticklebacks would respond to an extremely crude model, provided only that it had a red belly, but disdained very naturalistic models which lacked this crucial feature (Figure 4) (Tinbergen, 1951:28). Lorenz (1954: 291, translated by Eibl-Eibesfeldt, 1970:88) makes the general claim that "where an animal can be 'tricked' into responding to simple models, we have a response by an innate releasing mechanism." In the case of speech, most of the complexity of the spectrum can be dispensed with so long as the essential cues are preserved. It has already been mentioned that the simple, two-formant synthetic utterances of Figure 2 are clearly heard by subjects as [b], [d], etc. The natural and synthetic utterances in Figure 1 are linguistically equivalent, even though in the latter only the lower formants appear, and these in a very stylized configuration.

The synthetic utterance is not, however, simply an acoustic cartoon of the natural utterance. Though it shares with a cartoon the appearance of extreme simplicity and emphasis of salient features, it is rather a systematic attempt to represent, consistently but exclusively, the essential acoustic cues, all other details of the signal being discarded or neutralized. The principal loss in such synthetic speech is not intelligibility but only naturalness. This is rather surprising. One might quite reasonably expect that intelligibility would depend crucially on naturalness, that tampering with the observed spectrum of a natural utterance to any degree would alter its linguistic value or cause it not to be perceived linguistically at all. I do not mean to imply that high-quality natural speech would not be more intelligible than synthetic speech, or that sticklebacks would not respond more strongly to a real stickleback with a red belly than to a dummy. In synthetic speech, a host of redundant minor cues, as yet unidentified, are no doubt sacrificed together with the linguistically irrelevant details of the signal. Similarly, in the construction of the dummy, sign stimuli of minor importance have been ignored. But it appears that the dependence of artificial speech cues and sign stimuli on a naturalistic context is very small. Though the listener and (for all we know) the stickleback may be quite aware of the lack of naturalness, neither one appears to be disturbed by it. The relative naturalness of the speech cues and sign stimuli themselves is something else again, as will be seen shortly.

Both speech cues and sign stimuli exhibit what Tinbergen (1951:81), translating Seitz (1940), calls "the phenomenon of heterogeneous summation." The same response can be elicited by separate and noninteracting sign stimuli: thus, either the redness of the patch on the herring gull's bill or the contrast of the patch with the rest of the bill release the chick's pecking response. Moreover, if two stimuli for the same response are present, but one

Stickleback Models Used by Tinbergen

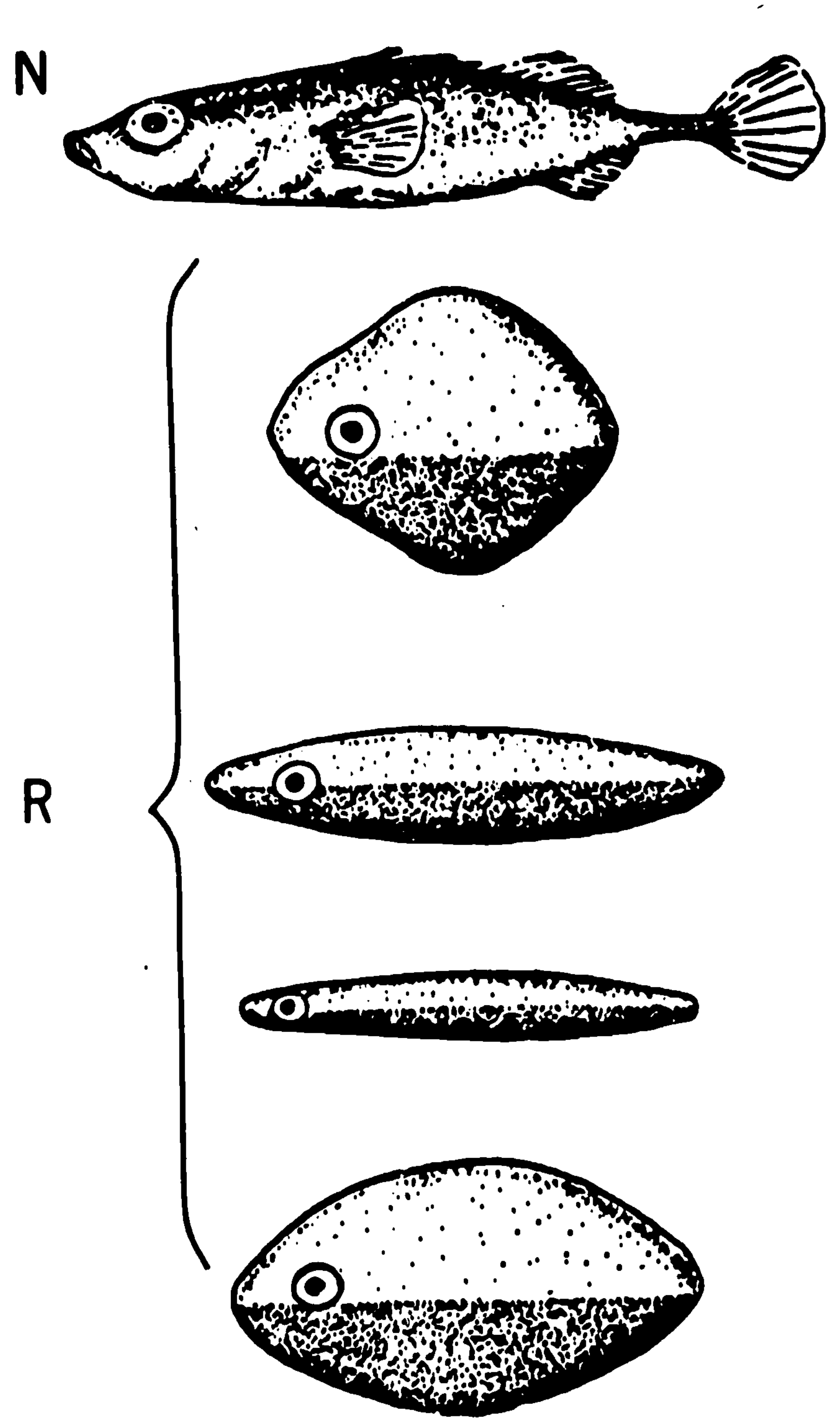


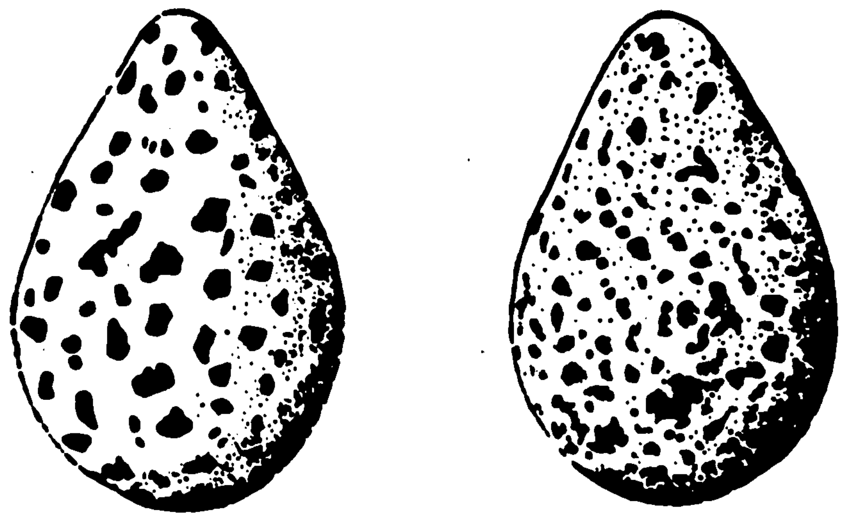
Fig. 4

Note: The fairly realistic model marked N, which lacked a red belly, provoked attack by male sticklebacks much less than the various crude models labeled R, which have red bellies. (After Tingergen, 1951.)

is defective, the second will compensate for the deficiency of the first. A similar principle operates in speech perception. Multiple cues for the same phonetic feature are the rule. For example, point of articulation in stop consonants is cued not only by the F2 transition but also by the F3 transition and by a burst of noise at an appropriate frequency just after release of stop closure (Delattre et al., 1955; Halle et al., 1957; Harris et al., 1958). In medial position, a voiced rather than a voiceless stop is cued by low-frequency periodic energy during closure, by lesser duration of closure, and by greater length of the preceding vowel (Lisker, 1957). Furthermore, the perceptual weight of one cue appears to be independent of that of the others; all combine additively to carry a single phonetic distinction; if a cue is defective or absent, as is very often the case in natural speech, the deficiency is compensated for by the presence of other cues. Thus Hoffman (1958) compared perception of point of articulation for (a) synthetic stop-vowel syllables in which all three cues (burst, F2 transition, F3 transition) were present, (b) syllables in which the burst cue was absent, (c) syllables in which the third formant with its transition was absent, and (d) syllables in which both third formant and burst were absent and only the F2 transition was present. He found that the optimal version of a cue for a particular point of articulation is the same whether presented separately or in combination with other cues; that labeling is most consistent when all three cues are optimal for the same point of articulation; and that an optimal F3 transition would compensate for a nonoptimal burst cue, and conversely. A.M. Liberman (personal communication) points out that speech also carries multiple cues to the sex of the speaker: men's voices differ from women's both in pitch range and in formant frequency range. Thus, neither the perception of speech cues nor that of sign stimuli is a Gestalt (Hinde, 1970).

An optimal speech cue is often not a realistic one; such a cue is the analog of a "supernormal" sign stimulus, such as the pattern of black spots on a white background on the artificial egg (see Figure 5) which the plover prefers to a natural egg with dark brown spots on a light brown background (Koehler and Zagarus, 1937). "The natural situation," Tinbergen (1951:44) observes, "is not always optimal." Similarly, if a human subject is presented with stimuli like those represented in Figure 2, he will hear the first few, those with rising transitions, as [bɛ]. The stimuli with the less steeply sloping transitions are closer to what one observes in instances of [bɛ] in natural speech, while the more extreme transitions are unlikely, perhaps even articulatorily impossible. Yet, in a labeling test, the more steeply rising the F2 transition, the more likely is the subject to hear [bɛ]. Thus the subject will label more consistently not only when more cues are present but also when the cues present are more nearly optimal, i.e., supernormal. Again, vowels spoken in isolation will occupy more extreme positions on the F1-F2 plane than vowels in connected speech (Shearme and Holmes, 1962) and are easier to label than the "same" vowels excised from connected speech. As Manning (1967) says, the failure of a sign stimulus to evolve to the supernormal extreme can usually be explained by considering other functional requirements. Thus the low-contrast, brown-on-brown spotting of the plover's eggs also serves to camouflage them from predators; black on a white background would not be so effective. The vocal tract, likewise, is primarily a group of devices for breathing and eating. A vocal tract which produced supernormal formant transitions and extreme vowels at normal speech rates

The Supernormal Plover Egg with Black Spots on a White Background (at left) Preferred by the Plover to the Normal Egg with Dark Brown Spots on a Light Brown Background (at right)



(After Koehler and Zagarus, 1937, reproduced in Tinbergen, 1951.)

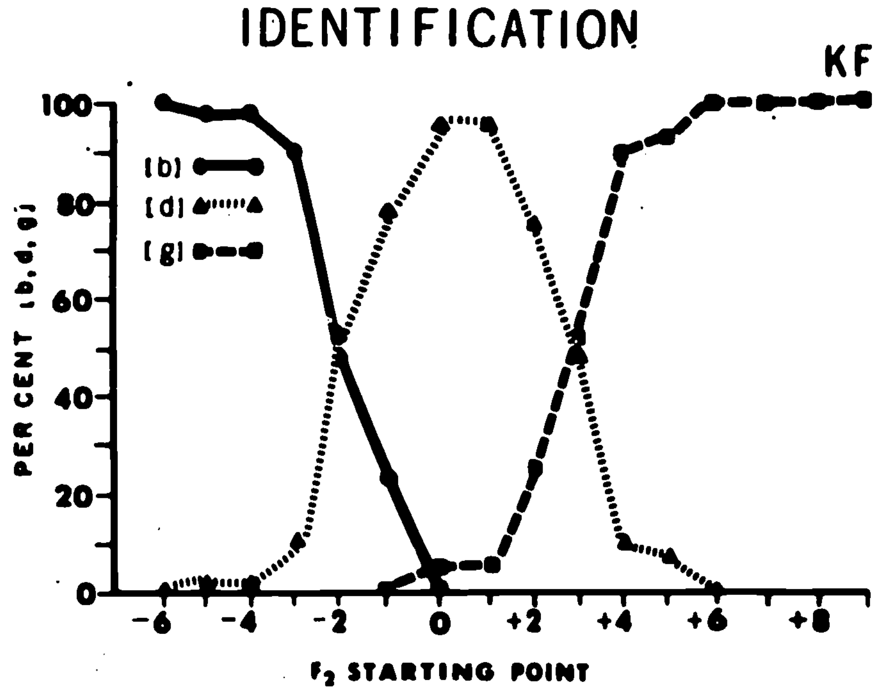
Fig. 5

would probably be unable to perform these primary functions properly. What is more interesting, as Manning goes on to point out, is that the tendency to respond to the sign stimulus has not evolved so as to be perfectly adjusted to the naturally occurring form of the stimulus. Like heterogeneous summation, this must reflect a characteristic of the process by which sign stimuli are perceived, and speech perception must share this characteristic. When we listen to natural speech, presumably we respond best to that combination of cues which approaches the supernormal ideal most closely. Thorpe (1961:98), similarly, has observed that the best natural sign stimulus display is the one which "can come nearest to the supernormal for the largest number of constituent sign stimuli."

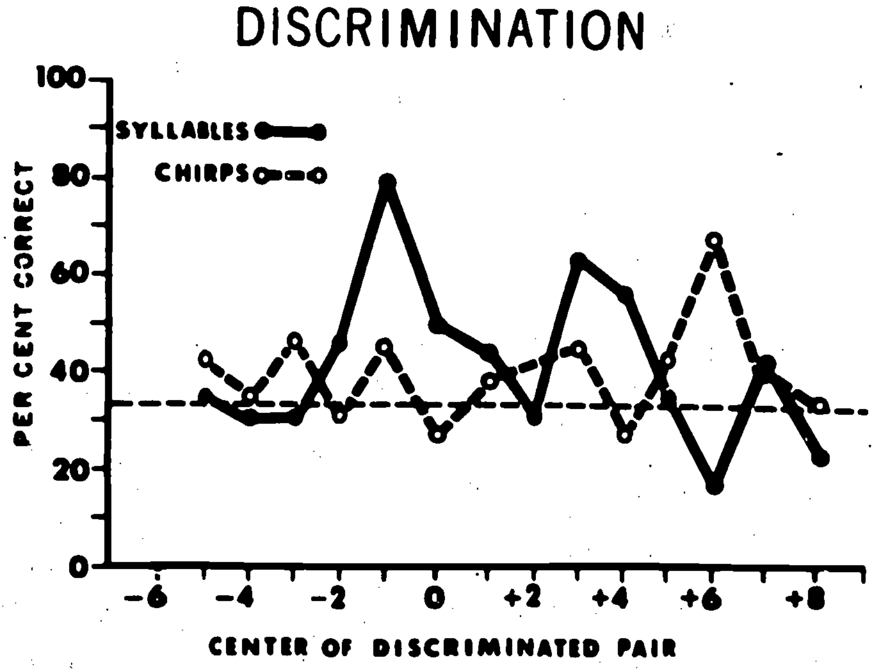
Finally, since the validity of the concept of a specialized neural mechanism to account for the selective perception of and response to sign stimuli is in dispute, the possibility that some such mechanism operates in speech perception is of special interest. The properties which speech perception have in common with sign stimuli point in this direction, for they are not characteristic of human auditory perception in general; so does the possibility of genetic transmission of knowledge of the cues. There is also some other evidence. If we ask a subject to discriminate pairs of stimuli which are adjacent along the acoustic series of stop-vowel syllables with varying F2 transition (Figure 2), he will do very well near the boundaries implied by the cross-over points in his labeling functions and very poorly elsewhere. The upper part of Figure 6 shows the labeling functions of a typical subject; the lower part (solid line) shows his discrimination function for the syllables. He is discriminating categorically (Liberman, 1957). Discrimination of this kind is quite unusual in psychophysical tasks. If we now give the subject a similar discrimination task in which the stimuli are "chirps," i.e., F2 transitions in isolation, without F1 or the steady-state portion of F2 (Figure 7), his discrimination function, represented by the dashed line in the lower part of Figure 6, is quite different. He discriminates better than random for most of the series, but the peaks of the syllable discrimination function are absent. Without a context containing other speech cues, the F2 transition is heard quite differently: there is no indication of categorical perception, and the function is more typically psychophysical (Mattingly et al., 1971).

Additional evidence for a special mechanism comes from experiments in dichotic presentation of speech sounds. If different stop-vowel syllables are simultaneously presented to a subject's two ears, he will be able to report correctly the stimuli presented to the right ear more often than the stimuli presented to the left ear. The effect is attributed to the processing of speech in the left cerebral hemisphere (Kimura, 1961; Studdert-Kennedy and Shankweiler, 1970). No such right-ear advantage is found with nonspeech signals such as musical tones (Kimura, 1964). Experiments by Conrad (1964), Wickelgren (1966), and others suggest that the speech perception mechanism is somehow involved with, and perhaps includes, "short-term memory."

To recapitulate, speech cues have a number of perceptual properties in common with sign stimuli. Their perception does not require a naturalistic context, they obey the law of heterogeneous summation, they are more effective as they approach a supernormal ideal, and there is reason to suppose that a special neural mechanism is involved. Some of these formal properties



Labeling and Discrimination Functions for One Subject for the Series Synthetic Speech Syllables Shown in Figure 2.



The Same Subject's Discrimination Function for the Series of "Chirps" Shown in Figure 7.

Fig. 6

The Pattern for a Series of "Chirps" (Isolated F2 Transitions), Corresponding to the Series of Stop-Vowel Syllables in Figure 2

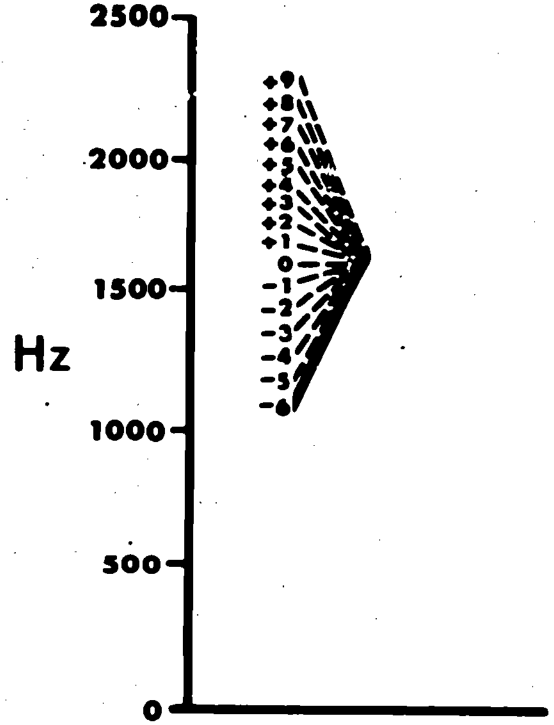


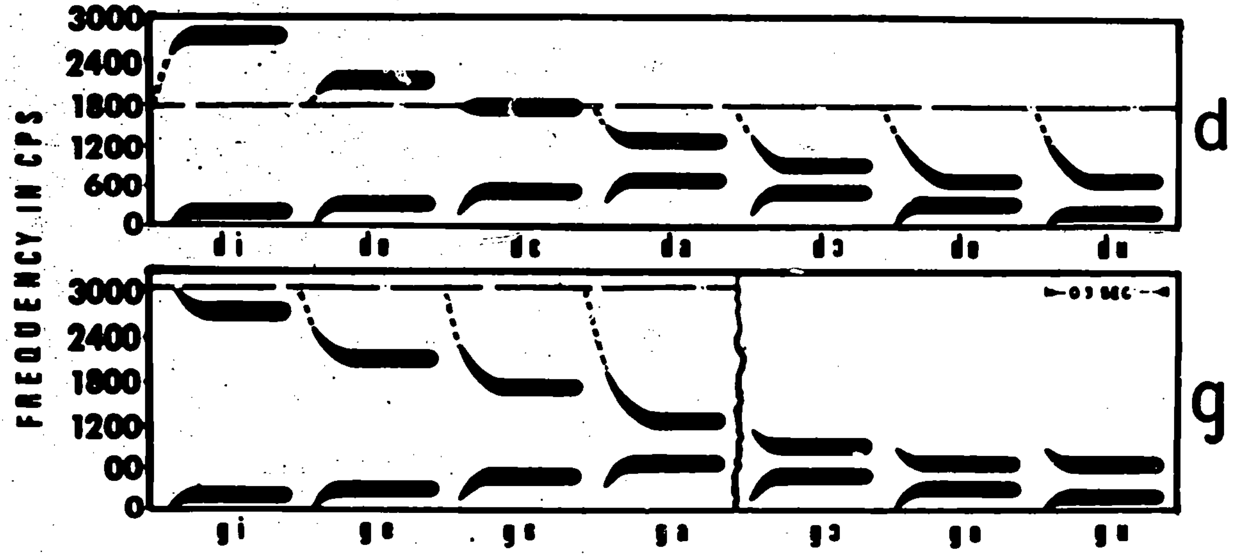
Fig. 7

appear in other situations--heterogeneous summation is a property of human binocular vision for instance--but it is their co-occurrence in both speech and sign stimuli that I find compelling. These properties are shared by the sign stimulus systems of many species, presumably for functional rather than for phylogenetic reasons. Thus, we are led to ask whether speech is in some way functionally similar to a sign stimulus system. But before considering this point, we ought to mention certain rather obvious differences between sign stimuli and the speech cues.

First the speech cues are transmitted at a rate much higher than the sign stimuli of any animal system. The displays in which sign stimuli occur, if not virtually static, are either relatively slow-moving or highly repetitive. But the acoustic events of speech which serve as cues occur extremely rapidly. The speech-perceiving mechanism not only keeps up with these events but is capable, as experiments with speeded speech have demonstrated, of speeds more than three times greater than normal speaking rates (Orr et al., 1965). A further gain in transmission speed is obtained by "parallel processing": the speaker produces and the listener extracts cues for different phonetic distinctions more or less simultaneously from the same acoustic activity (Liberman et al., 1967). Thus in a consonant-vowel syllable, the slope of the transition will carry information about the place of articulation of a consonant, its manner class (stop, fricative, semivowel) and about the quality of the vowel, while the excitation of these same transitions will cue the voicing distinction. The information rate of speech can be as high as 150 bits/second, and the question of the adaptive value of such a high rate arises.

Another difference between speech cues and sign stimuli is implicit in our use thus far of such terms as "place of articulation." Although the speech cues are acoustic events, the phonetic distinctions perceived by the listener are not acoustic but articulatory. Thus, the cues for, say, the alveolar sounds [t,d]--a high-frequency burst, an F2 transition which has a locus at about 1800 Hz, and an F3 transition with a locus at 3200 Hz--seem like a highly arbitrary selection if they are regarded as purely acoustic events. Moreover, the events do not occur synchronously; and, as we have just noted, they are interspersed with cues for other phonetic distinctions. But if these same events are interpreted as acoustic correlates of the simple articulatory gesture which produces [t,d], both the selection of events themselves and their relative timing appears quite straightforward. Another indication of the articulatory reference of the cues is that a series of stimuli may be perceived as belonging to the same phonetic category, even though they are not neighbors on an acoustic continuum, but they must not fail to be close together on some articulatory continuum. Thus the series of stimuli heard as [d] before vowels ordered from high front to low back form both an articulatory and an acoustic continuum,defined (though in somewhat oversimplified fashion) by the [t,d] locus (see the upper portion of Figure 8). But in the case of [k,g] the acoustic continuum is incomplete because the concept of the locus fails to apply consistently; the locus for [k,g] with low back vowels appears to be much lower and less clearly specifiable than for high front vowels (lower portion of Figure 8). Yet the perception is constant because the articulation is similar (Liberman, 1957). Conversely, the series of stimuli in Figure 2, which do form an acoustic continuum, divides into [b,d,g,] because the articulatory reference changes abruptly at

F1 and F2 Patterns for [d] and [g] Followed by a Seeies of Vowels
from High Front to Low Back



Note the discontinuity in F2 for [g].

Fig. 8

two points on the continuum. Because of such phenomena, it seems reasonable to regard speech as an acoustic encoding of articulatory gestures, or rather of the motor commands underlying those gestures (Lisker et al., 1962; Liberman et al., 1963; Studdert-Kennedy et al., 1970). We may call the sequence of motor commands which determines the speaker's output the "phonetic representation." The listener, because of his intuitive knowledge of the speech code, can recover this representation.

The most notable difference between speech cues and sign stimuli is that while sign stimuli typically produce a stereotyped behavioral response, speech cues do not. The reason the response to speech is not stereotyped is of course that unlike sign stimulus displays, a phonetic representation has no fixed significance apart from the linguistic system in which it functions; in itself it is a meaningless pattern, related only quite indirectly to the semantic values of the speakers and hearers. Speech does not stand by itself; it functions as part of language. The meaning of an utterance and the nature of the ultimate behavioral response depend not just on the characters of the stimulus, the environmental context, and the internal state of the perceiver, but also upon something not found in conjunction with any set of sign stimuli-a grammar. By virtue of a system of grammatical rules, shared by speaker and hearer, the speaker can evoke not just a few stereotyped responses but a wide variety, many of which are delayed or covert, and in principle, an infinite range of semantic values can be expressed. The problem is to explain why and how such a powerful system should have evolved.

It is with this problem that most attempts to find precedents for human language in animal behavior have begun. The cries of animals grossly resembling man, as well as animal communications systems which transmit a substantial amount of information even though the physical nature of the signals may be very different from human speech, have been scrutinized by many investigators for linguistic properties. These efforts have consistently failed. The properties treated as linguistic by some investigators have been so abstract--for example, the Hockett-Altmann "design features" (Altmann, 1967; Hockett and Altmann, 1968)--that those characteristics which distinguish language from purposive behavior in general are lost to view (Chomsky, 1968:60) and really fundamental features are placed on a level with trivial ones. Thus Hockett's Design Feature 3, "Rapid Fading," a property shared by all acoustic phenomena, is apparently just as important as DF 13, "Duality of Patterning," which, as we shall see, is truly significant. It is perhaps noteworthy that, according to Hockett and Altmann, the stickleback's communication system, which is of great interest from the viewpoint adopted here, lacks most of the linguistic Design Features.

Other investigators have tried indiscriminately to force the phenomena of animal behavior into standard linguistic categories. In Lenneberg's (1967: 228) words, they have attempted

> to count the number of words in the language of gibbons, to look for phomemes in the vocalizations of monkeys or songs of birds, or to collect the morphemes in the communication systems of bees and ants. In many other instances no such explicit endeavors are stated, but the underlying faith appears to be the same since much time and effort is spent in teaching parrots, dolphins or chimpanzee infants to speak English.

Such efforts, I think, are doomed to failure, and those who have insisted most strongly on the "biological basis of language"--Chomsky and Lenneberg--share this view. Chomsky (1968:62) suggests that human language "is an example of true 'emergence'--the appearance of a qualitatively different phenomenon at a specific stage of complexity of organization." Lenneberg (1967) believes that language has for the most part evolved covertly. In his view, we cannot expect that the steps in the evolution of a characteristic A from some quite different characteristic B will necessarily be manifest. The nature of the process of genetic modification is such that the intervening steps must in many cases remain obscure. This, he suggests, is the case with human language. While Lenneberg's general position on the nature of evolution may well be essentially correct, to take refuge in this position in the case of a particular evolutionary problem, such as the origin of human language, is essentially to abandon the problem.[1]

Despite the lack of precedents for grammar, I think that Chomsky and Lenneberg are perhaps unduly pessimistic and that the parallels between the speech cues and the sign stimuli suggest some interesting speculations about the origins of language.

One of the traditional explanations of language is that it developed from cries of anger, pain, and pleasure (see, e.g., Rousseau, 1755). The difficulty with this explanation is that it does not attempt to account for the transition from cries to names, or for the emergence of grammar. But let us put these problems to one side for the moment and postulate, just as the traditional explanation does, a stage in man's evolution when speech existed independently of language. Such speech, we suppose, had no syntax or semantics. But it was more than just expressive because it had phonetic structure. Its utterances were phonetic representations encoded by acoustic cues. If we ask what function such prelinguistic but structured speech could have had, the parallels we have discussed between speech cues and sign stimuli suggest a possible answer. Since speech is intraspecific, we suggest that it may have been, at this stage of evolution, a social releaser. If this speculation is correct, prelinguistic speech may have served early man as a vehicle for threat behavior, as a reproductive isolating mechanism, and as a means for mutual recognition of human parents and offspring. By means of phonetic representations underlying his utterances, man elicited appropriate behavioral responses from his fellows in each of these crucial situations. It is probably pointless to speculate as to what particular phonetic representations evoked what responses, but it perhaps reflects the primitive function which we

[1] Even if precedents for grammar existed in animal communication, it would be very difficult to learn about them. Most of what we know of the grammatical aspects of human language we know not from observations of human behavior but by virtue of our special status as members of the human species. The work of the linguist depends on the availability to him of the intuitions of speakers of a language that certain utterances are, or are not, grammatical. A member of another species, however intelligent, would find it difficult to deduce the most elementary grammatical concepts by observing and manipulating behavior: he would have, somehow, to consult the grammatical intuitions of a human speaker. We are similarly at a loss when speculating about the possible grammars of animal communication systems.

have attributed to speech that while the segmental aspects of speech have been adapted for linguistic purposes, the prosodic features remain as a primary means of physically harmless fighting, of courting, and of demonstrating and responding to parental affection.

If speech was once a social releaser system, we should expect it to show adaptation in the direction of "communications security." While being as conspicuous as possible on appropriate occasions to conspecific individuals, social releasers should be otherwise as inconspicuous as possible, in particular to prey and to predators. In the case of visual releasers, various camouflaging arrangements are found: outside the courtship period, the stickleback changes the color of his belly to a less noticeable shade and birds hide their brilliant plumage (Tinbergen, 1951). In the case of acoustic releasers, the animal can become silent when this is expedient; the simplicity of this solution is the great advantage of acoustic systems. As for speech, two of the differences we have noted between sign stimuli and speech cues are probably to be interpreted as further adaptations in the direction of security. The rapid rate at which the speech cues can be transmitted means that when necessary, transmissions can be extremely brief, making it so much the more difficult for an enemy to locate the source of the signal. And the fact that the articulatory information conveyed by speech can be perceived only by man means that, from the standpoint of other animals, as Hockett and Altmann (1968) point out, human speech is quite literally a code, concealing not only the phonetic representation but also the fact that there is such a representation and that the speaker is human. Presumably the animals man preyed upon would not have been able to distinguish his speech from the chatter of herbivorous nonhuman primates.

Moreover, if we regard speech as a social releaser system, a natural explanation is available for an old problem. The fact that no other animal except man can speak, not even the primates to whom he is most closely related, has long been a cause for wonder and speculation. But, of course, a social releaser is required, almost by definition, to be species-specific: it must be so if it is to perform its authentication function effectively. It is thus no more surprising that speech should be unique to man than that zigzag dances should be unique to sticklebacks.

Let us now consider how the concept of prelinguistic speech as consisting of a system of phonetic social releasers bears on the problem of the origin of language. Most speculations on this topic suppose that man's unusual intelligence must have been the principal factor in the development of language. The weaker version of this view (which would have been that of many post-Bloomfieldian linguists) assumes that man's intelligence differs from that of animals in degree: he alone is intelligent enough to divide the world into its semantic categories and to recognize their predicative relationships. The structure of his language, insofar as it is not purely a matter of convention, reflects the structure of human experience. The stronger version of this view (which I think it is fair to attribute to Chomsky and his colleagues) assumes that man's intelligence differs in kind from that of other animals and that the structure of language, properly understood, reflects specific properties of the human intellect. Speech, according to either version, serves simply as the vehicle for the abstract structure of language. The anatomy of the vocal tract imposes certain practical constraints on linguistic behavior but has only a trivial relationship to linguistic structure.

The difficulty with this view is not only that it makes no attempt to account for the choice of speech as the vehicle of language, but also that many animals display some degree of intelligence, and a few display intelligent behavior comparable in some ways to man's. One would expect to find some limited linguistic behavior among animals of limited intelligence, or something approximating human linguistic behavior among animals whose intelligence seems to resemble man's. But, as we have seen, precedents of any kind are lacking, and it is argued that language is an instance of evolutionary "emergence."

I wish to suggest a somewhat less drastic alternative to emergence. This is that language be regarded as the result of the fortunate coexistence in man of two independent mechanisms: an intellect, capable of making a semantic representation of the world of experience, and the phonetic social-releaser system, a reliable and rapid carrier of information. From these mechanisms a method evolved for representing semantic values in communicable form.

Before this could happen, a means had to be found for the speaker-hearer to recode semantic representations into phonetic representations, and phonetic representations into semantic representations. Clearly this recoding is a complex process, if only because the intellect, being capable of representing a wide range of human experience, probably has a very large number of categorical features available for semantic representations in long-term memory, while the phonetically significant configurations of the vocal tract can be described in terms of a very small number of categorical features--fifteen or twenty at most (Chomsky and Halle, 1968). It would thus be impossible to accomplish the recoding simply by mapping semantic features onto phonetic features. It was necessary for another mechanism to evolve: linguistic capacity, the ability to learn the grammar of a language.[2] The grammar is a description of the complex but rule-governed relationships, in part universal, in part language-specific, which obtain between semantic representations and phonetic representations. By virtue of his grammatical competence, a person can speak and understand utterances in the language according to the rules of grammar.[3]

[2] In this discussion, I have ignored for simplicity's sake the obvious fact that there are not one but many languages, each with its own grammar. To Rousseau (1755) and von Humboldt (1836), to explain the diversity of human languages was regarded as a problem second in importance only to that of explaining the origin of language. Recently, Nottebohm (1970) has offered the intriguing suggestion, based on an analogy with bird song, that language diversity enables some members of a species to develop traits appropriate to their particular environment without an irreversible commitment to subspeciation.

[3] The account of the organization of grammar given here, necessarily oversimplified, is based on Chomsky (1965, 1966).

One component of the grammar is the lexicon, a list of morphemes with which semantic, syntactic, and phonological information is associated. The stock of morphemes in a language is large but finite, while the number of conceivable semantic representations is infinite. But an infinite number of grammatical strings of morphemes can be generated by the syntactic component of the grammar, and from these, the semantic component can generate a correspondingly infinite number of semantic representations. The phonological component parallels the semantic component: for each string of grammatical morphemes, a phonetic representation can be generated. The speaker's task is thus to find a phonetic representation which corresponds grammatically to a given semantic representation, while the hearer's task is to find a semantic representation corresponding to a given phonetic representation. In both his roles, the speaker-hearer, in order to recode, must determine heuristically the probable input to a grammatical component, given its output and the rules which generate output from input. Very little is known about how he performs these tasks.

For our purposes, however, the important point is that a grammar has an obvious symmetry. There is a core, the syntactical and lexical components, and two other components, the semantic and the phonological, which generate the semantic and phonetic representations, respectively. The nature of the semantic component, and the representation it generates, appear to be appropriate for storage in long-term memory. The nature of the phonological component, and the representation _it_ generates, are appropriate for on-line transmission by the vocal tract. To relate these two representations is the main motivation of the grammar, and its form is determined both by the properties of the intellect and by those of the phonetic social-releaser system. It is thus surely not correct to view speech as if it were merely selected by happenstance as a convenient vehicle for language.

Once the grammar had begun to develop, we should not be surprised to find that it exercised a reciprocal influence on the development both of the phonetic system and of the intellect. In the case of the former, it has been argued very persuasively (Lieberman et al., in press; Lieberman and Crelin, 1971) that the vocal tract of modern man has evolved from something rather like that of a chimpanzee to its present form, with a shorter jaw, a wider and deeper pharynx, and vocal cords for which the tension is more finely controlled, and that these modifications not only have no other discernible adaptive value than to increase the reliability and the richness of structure of human speech but are actually disadvantageous for the vocal tract's primary functions of chewing, breathing, and swallowing. If man's vocal tract has evolved in this way, corresponding modifications must have taken place in the neural mechanisms for production and perception of speech, resulting in the speech code in the form we now know it. The evidence for the development and specialization of the human intellect as a result of its grammatical affinities is, of course, far less concrete, but the very least that can be said is that the capability of symbolizing things and ideas by words permits a degree of conceptual abstraction without which the kind of thinking which human beings regularly do would be impossible.

If the function of a grammar is to serve as an interface between the phonetic and semantic domains, it is hardly surprising that precedents for linguistic behavior have not been found. The speech production and perception system is a highly specific mechanism; so also is the human intellect. Their co-occurrence in man was a remarkable piece of luck; other animals, which on behavioral or physiological grounds appear to be of high intelligence, had no opportunity to develop language because they lacked a suitable pre-existing communications system. Moreover, even if high intelligence and an appropriate communications system had co-occurred in some other species and combined to form a "language," its grammar would be utterly different in form from any human grammar, because the intellectual and communicative mechanisms from which it evolved would be quite different in detail from the corresponding human mechanisms. In the circumstances, the most we can hope for is to understand more about the separate evolution of the intellect and that of the speech code and to interpret human grammars in terms of their dual origin.

To summarize, I have called attention to certain parallels between the speech cues and sign stimuli. These parallels suggest the speculation that prelinguistic speech may have functioned as a social-releaser system, which would explain the fact that speech is species-specific. It is suggested, furthermore, that human language is not simply the product of the human intellect but is rather to be viewed as the joint product of the intellect and of this prelinguistic communications system. Grammar evolved to interrelate these two originally independent systems. Its dual origin explains the lack of precedents for language in animal behavior and its apparent "emergence."

## REFERENCES

Abramson, A. and Lisker, L. (1970) Discriminability along the voicing continuum: Cross-language tests. In Proc. 6th International Cong. Phonetic Sciences. (Prague: Academia).

Altmann, S.A. (1967) The structure of primate social communication. In Social Communication among Primates, S.A. Altmann, ed. (Chicago: Univ. of Chicago Press).

Chomsky, N. (1965) Aspects of the Theory of Syntax. (Cambridge, Mass.: M.I.T. Press).

Chomsky, N. (1966) Topics in the Theory of Generative Grammar. (The Hague: Mouton).

Chomsky, N. (1968) Language and Mind. (New York: Harcourt Brace).

Chomsky, N. and Halle, M. (1968) The Sound Pattern of English. (New York: Harper and Row).

Conrad, R. (1964) Acoustic confusions in immediate memory. Brit. J. Psychol. 55, 75-83.

Cooper, F.S. (1950) Spectrum analysis. J. acoust. Soc. Amer. 22, 761-762.

Cooper, F.S., Delattre, P.C., Liberman, A.M., Borst, J.M. and Gerstman, L.J. (1952) Some experiments on the perception of synthetic speech sounds. J. acoust. Soc. Amer. 24, 597-606.

Delattre, P.C., Liberman, A.M. and Cooper, F.S. (1955) Acoustic loci and transitional cues for consonants. J. acoust. Soc. Amer. 27, 769-773.

Eibl-Eibesfeldt, I. (1970) Ethology. (New York: Holt, Rinehart and Winston)

Eimas, P.D., Siqueland, E.R., Jusczyk, P. and Vigorito, J. (1970) Speech perception in infants. Science 171, 303-306.

Falls, J.B. (1969) Functions of territorial song in the white-throated sparrow. In Bird Vocalizations in Relation to Current Problems in Biology and Psychology, R.A. Hinde, ed. (Cambridge: Cambridge University Press).

Fant, C.G.M. (1960) Acoustic Theory of Speech Production. (The Hague: Mouton).

Hailman, J.P. (1969) How an instinct is learned. Scientific American 221, 6, 98-106.

Halle, M., Hughes, G.W. and Radley, J.-P.A. (1957) Acoustic properties of stop consonants. J. acoust. Soc. Amer. 29, 107-116.

Harris, K.S., Hoffman, H.S., Liberman, A.M., Delattre, P.C. and Cooper, F.S. (1958) Effect of third-formant transitions on the perception of the voiced stop consonants. J. acoust. Soc. Amer. 30, 122-126.

Hinde, R.A. (1970) Animal Behavior. 2nd ed. (New York: McGraw-Hill).

Hockett, C.F. and Altmann, S.A. (1968) A note on design features. In Animal Communication, T.A. Sebeok, ed. (Bloomington: Indiana Univ. Press).

Hoffman, H.S. (1958) Study of some cues in the perception of the voiced stop consonants. J. acoust. Soc. Amer. 30, 1035-1041.

Kimura, D. (1961) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15, 166-171.

Kimura, D. (1964) Left-right differences in the perception of melodies. Quart. J. exp. Psychol. 16, 355-358.

Koehler, O. and Zagarus, A. (1937) Beiträge zum Brutverhalten des Halsbandregenpfeifers (Charadrius h. hiaticula L.). Beitr. Fortpflanzungs biol. Vögel 13, 1-9. Cited by Tinbergen, 1951.

Lenneberg, E. (1967) Biological Foundations of Language. (New York: John Wiley).

Liberman, A.M. (1957) Some results of research on speech perception. J. acoust. Soc. Amer. 29, 117-123.

Liberman, A.M., Cooper, F.S., Harris, K.S. (1963) A motor theory of speech perception. In Proceedings of the Speech Communications Seminar. (Stockholm: Speech Transmission Laboratory, Royal Institute of Technology).

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

Liberman, A.M., Delattre, P.C. and Cooper, F.S. (1958) Some cues for the distinction between voiced and voiceless stops in initial position. Lang. and Speech 1, 153-167.

Liberman, A.M., Delattre, P.C., Gerstman, L.J. and Cooper, F.S. (1956) Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. exp. Psychol. 52, 127-137.

Lieberman, P., Crelin, E.S. and Klatt, D.H. (in press) Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. American Anthropologist.

Lieberman, P. and Crelin, E.S. (1971) On the speech of Neanderthal man. Linguistic Inquiry 2, 203-222.

Lisker, L. (1957) Closure duration and the intervocalic voiced-voiceless distinction in English. Lang. 33, 42-49.

Lisker, L. and Abramson, A.S. (1970) The voicing dimension: Some experiments in comparative phonetics. In Proc. 6th International Cong. Phonetic Sciences. (Prague: Academia).

Lisker, L., Cooper, F.S. and Liberman, A.M. (1962) The uses of experiment in language description. Word 18, 82-106.

Lorenz, K. (1935) Der Kumpan in der Umwelt des Vogels. J. f. Ornith. 83, 137-213, 289-413. Tr. in Instinctive Behaviour, C.H. Schiller, ed. (London: Methuen, 1957).

Lorenz, K. (1954) Das angeborene Erkennen. Natur und Museum 84, 285-295.

Manning, A. (1967) An Introduction to Animal Behavior. (Reading, Mass.: Addison-Wesley).

Mattingly, I.G. (1968) Experimental methods for speech synthesis by rule. IEEE Trans. Audio 16, 198-202.

Mattingly, I.G., Liberman, A.M., Syrdal, A.K. and Halwes, T. (1971) Discrimination in speech and nonspeech modes. Cognitive Psychol. 2, 131-157.

Nottebohm, F. (1970) Ontogeny of birdsong. Science 167, 950-966.

Orr, D.B., Friedman, H.L. and Williams, J.C.C. (1965) Trainability of listening comprehension of speeded discourse. J. educ. Psychol. 56, 148-156.

Rousseau, J.-J. (c.1755) Essay on the origin of languages. Tr. by J. H. Moran in On the Origin of Language, J. H. Moran and A. Gode, eds. (New York: Ungar)

Russell, E.S. (1943) Perceptual and sensory signs in instinctive behavior. Proc. Linnacan Soc. London 154, 195-216.

Seitz, A. (1940) Die Paarbildung bei einigen Cichliden I. Zs. Tierpsychol. 4, 40-84. Cited in Tinbergen, 1951.

Shearme, J.N. and Holmes, H.N. (1962) An experimental study of the classification of sounds in continuous speech according to their distribution in the formant 1-formant 2 plane. In Proc. Fourth Int. Cong. Phonetic Sciences. (The Hague: Mouton).

Studdert-Kennedy, M., Liberman, A.M., Harris, K.S. and Cooper, F.S. (1970) Motor theory of speech perception: A reply to Lane's critical review. Psychol. Rev. 77, 234-249.

Studdert-Kennedy, M. and Shankweiler, D. (1970) Hemispheric specialization for speech perception. J. acoust. Soc. Amer. 48, 579-594.

Thorpe, W.H. (1961) Introduction to: Experimental studies in animal behavior. In Current Problems in Animal Behaviour, W.H. Thorpe and O.L. Zangwill, eds. (Cambridge: Cambridge University Press).

Tinbergen, N. (1951) The Study of Instinct. (Oxford: Clarendon Press).

von Humboldt, Wilhelm. (1836) Linguistic Variability and Intellectual Development. Tr. by G.C. Buck and F.A. Raven. (Coral Gables, Fla.: Univ. of Miami Press, 1971).

Wickelgren, W.A. (1966) Distinctive features and errors in short-term memory for English consonants. J. acoust. Soc. Amer. 39, 388-398.