XVIII. LINGUISTICS

Academic and Research Staff

Prof. Roman Jakobson
Prof. Morris Halle
Prof. A. Noam Chomsky
Prof. Sylvain Bromberger
Prof. Jerry A. Fodor
Prof. Merrill F. Garrett
Prof. Kenneth L. Hale
Prof. James W. Harris
Prof. R. Paul V. Kiparsky
Prof. Wayne A. O'Neil
Prof. David M. Perlmutter
Prof. Krystyna Pomorska
Prof. John R. Ross
Dr. Sandra Chung
Dr. George N. Clements
Dr. Barbara C. Lust
Dr. Dorothy Siegel
Dr. Susan M. Steele
Dr. Edward C. T. Walker
Frederic M. Katz
Scott Soames

Graduate Students

Anthony E. Ades
Avery D. Andrews
Sarah J. Bell
Jill L. Carrier
Jeffrey Cohen
Michael A. Cohen
David B. Duncan
Leland M. George
John A. Goldsmith
Harrold W. Gray
Shosuke Haraguchi
Earl Hofer
La Verne M. Jeanne
Daniel Kahn
Ellen J. Kaufman
Mary-Louise Kean
Wha-Chun M. Kim
Laura E. Knecht
David I. Leslie
Philip S. Le Sourd
Mark Y. Liberman
Marcia Linebarger
Carolyn L. Marsh
Susan E. Martin
Juan Mascaró
Robert C. May
Paul R. Platero
Alan S. Prince
Tanya Reinhart
Alberto M. Rivas
Joel Rotenberg
Ivan A. Sag
Alan W. Sparer
Craig L. Thiersch
Erich F. Woisetschlaeger

RESEARCH OBJECTIVES

National Institute of Mental Health (Grant 3 PO1 MH13390-08S1)

National Institutes of Health (Grant 5 TO1 HD00111-10)

Morris Halle

The ultimate objective of our research is to gain a better understanding of man's mental capacities by studying the ways in which these capacities manifest themselves in language. Language is a particularly promising avenue because, on the one hand, it is an intellectual achievement that is accessible to all normal humans and, on the other hand, we have more detailed knowledge about language than about any other human activity involving man's mental capacities.

Scientific descriptions of languages have for a very long time followed a standard format. A number of topics are almost invariably discussed; for example, pronunciation, the inflection of words, word formation, the expression of syntactic relations, word order, and so forth. Moreover, the manner in which these have been treated has also been quite standard. While traditional grammars have many shortcomings, their great practical utility is beyond question; generations of students have acquired adequate command of innumerable languages with the help of grammars of the standard type. A plausible inference that might be drawn from this fact is that languages are somehow not very different from one another and the traditional standard format has succeeded in capturing essential aspects of what all languages share in common. Accordingly, much of the research of our group has been devoted to studying the common framework that underlies different languages, the general principles that are exemplified in the grammar of different languages. Results strongly indicate that this assumption is indeed correct as far as the linguistic evidence is concerned.

The preceding discussion leads quite naturally to the question, "What evidence from outside of linguistics might one adduce in favor of the hypothesis that all languages are constructed in accordance with a single plan, a single framework?" It seems to us that the most striking evidence in favor of the hypothesis is, on the one hand, the rapidity with which children master their mother tongue, and on the other hand, the fact that even a young child's command of his mother tongue encompasses not only phrases and utterances he has heard but also an unlimited number of phrases and utterances he has not previously encountered. To account for these two sets of facts, we must assume that in learning a language a child makes correct inferences about the structural principles that govern his language on the basis of very limited exposure to the actual sentences and utterances. In other words, we must assume that with regard to matters of language a child is uniquely capable of jumping to the correct conclusions in the overwhelming majority of instances, and it is the task of the student of language to explain how this might be possible.

A possible explanation might run as follows. Assume that the human organism is constructed so that man is capable of discovering only selected facts about language and, moreover, that he is constrained to represent his discoveries in a very specific fashion from which certain fairly far-reaching inferences about the organization of other parts of the language would follow automatically. If this assumption is accepted, the next task is to advance specific proposals concerning the devices that might be actually at play. The obvious candidate is the theoretical framework of linguistics, for while it is logically conceivable that the structure of language might be quite distinct from that of the organism that is known to possess the ability to speak, it is much more plausible that this is not the case, that the structures that appear to underlie all languages reflect quite directly features of the human mind. To the extent that this hypothesis is correct – and there is considerable empirical evidence in its favor – the study of language is rightly regarded as an effort at mapping the mysteries of the human mind.

A. Pi ka pu: THE PERCEPTION OF SPEECH SOUNDS BY PRELINGUISTIC INFANTS

National Institute of Mental Health (Grant HD 05168-01, 02, 03)
M.I.T. Sloan Fund for Basic Research
Grant Foundation

Jerry A. Fodor, Merrill F. Garrett, Sarah L. Brill

Introduction

William James thought that the experiential world of the neonate is a "blooming, buzzing confusion." But the real world, as Robert Louis Stevenson remarked, "is full of a number of things" (emphasis ours). Much of classical cognitive psychology is concerned with how this gap is bridged; how development shapes the aboriginal sensory flux to produce the adult ontology of properties, persons, events, and objects. This process is sometimes called the child's construction of reality.

Many psychologists no longer believe that the classical story is correct, even in outline. In particular, they doubt that the child is ever faced with an unstructured flow of sensations in which the real world must somehow be discovered. With this revised view, the basic psychological mechanisms impose perceptual structure right from the beginning; cognitive development is primarily the elaboration of these structural commitments.

Whichever view one takes, clearly it is an open question what the precise nature of infant perception is and how it relates to that of adults. The research reported here touches on an aspect of this question: To what extent do infants classify speech sounds in the way that adults do? Do prelinguistic infants recognize the patterns of identity and differences in terms of which adults taxonomize the speech signal? The results of our experiments with infants from 14 to 18 weeks old suggest that it is likely that they do. Differential reinforcement of head orientations to members of a randomly repeating series composed of three different CV syllables (e.g., /pa/, /pi/, /ku/) produced significantly greater resistance to extinction when the reinforced stimuli shared a phone (e.g., /pa/ and /pi/) than when they did not (e.g., /pa/ and /ku/). This result is interpretable on two assumptions: (i) infants, like adults, find 'disjunctive' concepts relatively hard to learn; i.e., learning is faster when the positive discriminative stimuli satisfy a uniform description; (ii) infants, like adults, recognize phonetic identities: i.e., the infant's judgment of similarity between syllables is responsive to the number of phones that the syllables share. On these assumptions, a learning task in which /pa/ and /pi/ are the positive discriminative stimuli ought to be mastered more readily than a learning task for which /pa/ and /ku/ (or /pi/ and /ku/) are the positive discriminative stimuli; for, while the reinforcement conditions are disjunctive in the latter cases, they are homogeneous in the former case (all and only reinforced signals begin with /p/). The results suggest, therefore, that a triple of syllables such as /pa/, /pi/ and /ku/ is phonetically cross-classified for the infant just as it is for the adult: /pa/ and /pi/ are heard as having more in common with one another than either does with /ku/.

The significance of this pattern of findings can best be appreciated in the context of work on the psychophysics of adult speech perception. A long series of studies (notably those at Haskins Laboratories) have rendered it implausible that a perceived phone can be defined by a given set of contemporaneous acoustic features. That is, more often than not, researchers have failed to find acoustic invariants corresponding to the phenomenal identity of the phone. The evidence indicates:

(a) A heterogeneous set of acoustic tokens may be assigned to the same phonetic type.

(b) The processes involved in such assignments often involve the categorical representation of acoustic continua.

(c) The acoustic features relevant to such assignments are often widely distributed across the speech signal; ordinal relations among perceived phonetic segments do not, in general, parallel temporal relations among substretches of the speech stream.

In particular, the character of the acoustic representative of the consonantal segment of a CV syllable is typically heavily determined by the character of the acoustic representative of the vowel segment. Acoustically, a CV syllable is likely to be a relatively fused and undifferentiated object, even though phonetically (and perceptually) it is analyzed as a sequence of discrete and more or less independent elements. (For extensive discussion see Liberman et al.[1])

It is plausible to think of the phonetic/perceptual representation of the speech stream as a constancy, engendered by the operation of some sort of decoding mechanism that the adult applies to the acoustic input. The complexity of the computations performed by this mechanism is suggested by the extraordinary difficulties encountered thus far in attempts to develop artificial phone recognizers. If this view of adult speech perception is correct, it implies one of two theories of the ontogenesis of speech perception (paralleling the two accounts of general cognitive development mentioned above). Either the relevant perceptual constancies must be learned, or they are part of the initial equipment that the infant brings to the process of internalizing the rules of his language. Theories of the first sort have stressed the importance of the child's experience of relatively large samples of speech in working out the distributional patterns (patterns of partial contrast between meaningful utterances) on which the linguistic justification of a phonetic analysis eventually rests. Or, in the case of 'motor theories,' they have stressed the importance of the child's monitoring of his own verbalizations in facilitating his discovery of the relation between waveforms and phonetic strings. In either case, however, such theories suggest (as nativistic theories do not) that the mastery of the phonetic/acoustic correspondence ought to be the consequence of a relatively extensive linguistic apprenticeship; one would hardly expect to find it available to prelinguistic infants.

Recent experiments (e.g., Eimas, Siqueland, Jusczyk, and Vigorito[2]) have suggested, nevertheless, that infants, like adults, impose a categorization on speech signals, and, more significantly, that apparently the category boundaries for infants and adults correspond. In our experimentation we set out to explore the independent question whether infants, like adults, respect the existence of internal syllabic structure. Can infants, despite their lack of experience with distributional properties of the language, and their lack of experience with the output of their own articulatory system, nevertheless abstract from the acoustic contamination of one part of a syllable by the rest, and thus recover a linguistically relevant representation of the syllable as a sequence of distinct phones? Does the infant hear the same pattern of phonetic identity and difference among parts of syllables that the adult hears?

Design and Procedure

The paradigm used to test infants' perception of speech, briefly, is the following: Infants (14-18 weeks old) were presented with a series of 60 single-syllable sounds; 3 different syllables occurred in the series (e.g., /pa/, /pi/, /ku/) and there were 20 occurrences of each syllable. The syllable types occurred in semirandom order from a sound source at either the infant's right or left; the locus of a particular syllable type varied randomly (except for the constraint that, summed across trials, the syllable occurred left or right equally often). For two of the three syllable types in the series, a reinforcing visual array located at the apparent sound source followed each occurrence of these syllable types.[3] Figure XVIII-1 is a schematic representation of the test situation. The test session lasted approximately one-half hour and was repeated on 10 separate days.

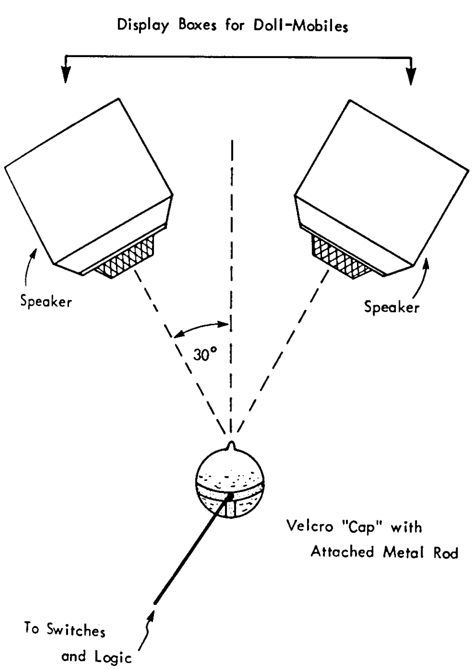

Fig. XVIII-1. Test conditions.

The variable of interest is the incidence of anticipatory head turns: those correctly orienting head turns that occur in the interval following presentation of a syllable, but prior to the illumination of the visual array. This type of head turn may be taken as an index of the infant's ability to predict occurrences of the reinforcing visual array as a function of the character of the syllable. The incidence of such anticipatory head turns[4] for each of the syllable types in the stimulus series may be compared for groups in which the two syllables that are reinforced have consonants that are perceptually distinct for adults (e.g., /pi/ and /ka/) and those in which the reinforced syllables have consonants that are perceptually the same for adults (e.g., /pi/ and /pu/).

Consonants were counterbalanced across vowel environments and yielded 3 syllable sets.[5] The two reinforcement conditions for each such set yield 6 experimental groups. Experiments on 6 infants were run in each of these groups. The design is outlined as follows:

		"Same Phones" Conditions	"Different Phones" Conditions
		visual array presented:	visual array presented:
Syllable series A:	/pi/, /ka/, /pu/	for /pi/ & /pu/	for /pi/ & /ka/
Syllable series B:	/pa/, /ku/, /pi/	for /pa/ & /pi/	for /pa/ & /ku/
Syllable series C:	/pu/, /ki/, /pa/	for /pu/ & /pa/	for /pu/ & /ki/

Results and Discussion

The performance of infants was compared in terms of the proportions of total anticipatory responses to each syllable type which occurred in the last 5 of the 10 sessions. If presentation of the visual array affects changes in the rate of anticipatory responses, reinforced syllable types will show a larger proportion of such responses in the later sessions than will unreinforced syllable types, and this difference should manifest itself regardless of the initial rates of response for particular syllable types. Note, however, that this presupposes that infants are capable of distinguishing among the various syllables in the stimulus series, and of connecting the appearance of the visual array with particular syllable types. Furthermore, with the view that infants are also capable of appreciating the partial phonetic identities among the syllables, one would expect the differences between proportions of response to reinforced and unreinforced syllable types to be greater in the "same phones" group than in the "different phones" group. If on the other hand, each of the syllable types in a series is heard by the infants as equally distinct from each of the other two, no differences as a function of syllable grouping should be found.

A two-factor analysis of variance (factor 1, reinforced vs unreinforced syllables; factor 2, syllable grouping by same or different consonants) with repeated measures on the first factor[6] shows a significant effect for reinforcement ($F = 5.68$, df 1, 34), and for the interaction of syllable grouping with reinforcement (i.e., presentation of the visual array) ($F = 4.14$, df 1, 34). Tests for differences among the means show that the effects in the reinforcement factor are due to the "same phones" group ($t = 2.94$, df 17) with no significant effects in the "different phones" group ($t = .26$, df 17). Table XVIII-1 lists the mean second-half proportions for reinforced and unreinforced syllables in both groups.

Table XVIII-1. Proportions of anticipatory head turns in last 5 test sessions for same and different phone conditions.

SYLLABLE GROUPING	REINFORCEMENT CONDITIONS	
Same Phones Condition	/pa/ & /pi/ /pi/ & /pu/ .428 /pu/ & /pa/	/ku/ /ka/ .358 /ki/
Different Phones Condition	/pa/ & /ku/ /pi/ & /ka/ .413 /pu/ & /ki/	/pi/ /pu/ .404 /pa/

The strongest reflection of the effect of syllable grouping on the infants' ability to predict appearances of the visual array, of course, is the interaction. This indicates that the difference between anticipatory responses to reinforced and unreinforced syllables was significantly greater in the "same phones" group than in the "different phones" group, even though the direction of the relation between the means in the two groups was the same.

These results parallel those that we obtained earlier[7] with a variation of the current paradigm and a different pair of consonants. In the earlier experiment we used the consonants /p/ and /g/ in a design like the one described for the present experiment.[8] The results were that infants in the "same phones" condition showed significantly greater retention of the head-turn response across test sessions than did infants in the "different phones" condition.

The stops /p/ and /g/ (like /p/ and /k/) have the virtue of yielding cases of what might be called "acoustic overlap" across vowel environments. Thus, for example, the noise burst appropriate to the acoustic representative for the percept /p/ in the environment /i/ is the same as that appropriate to /g/ (or /k/) in the environment /a/ (see Schatz[9] and Liberman, Delattre, Cooper, and Gerstman[10]). The pair /p/ and

/g/ (unlike /p/ and /k/) also have the disadvantage that they are distinct in both place of articulation and voicing. We do not consider that the availability of such a cue necessarily renders the infants' perceptual task (that of selecting among the various available acoustic parameters just those that yield the perceptually relevant adult taxonomy) less exacting. Nevertheless, it is true that for the single contrast between /p/ and /g/, a relatively simple system keyed only to voicing oppositions could perform the relevant sorting. The present results with /p/ and /k/ cannot be accounted for in the same way, although in this experiment there must also be some acoustic parameters that will appropriately sort the syllables. In other words, we assume that the infants' performance is not occult. The significant fact is that the parameters the infants apparently do select are those that yield the adult classification of percepts.

The results of the present experiment and the earlier one both indicate that infants from 14 to 18 weeks old are capable of a segmental analysis of syllables (i.e., they respond to their internal structure as a sequence of phones). It seems clear that infants with no experience in producing articulate speech and with no experience of the distributional features of language can, nevertheless, appreciate the perceptual identity of stop consonants across vowel environments, and to this extent at least their naive perceptual analysis of speech signals corresponds to that of sophisticated adults.

Footnotes and References

1. A. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, "Perception of the Speech Code," Psychol. Rev. 74, 431-461 (1967).
2. P. Eimas, E. Siqueland, P. Jusczyk, and J. Vigorito, "Speech Perception in Infants," Science 171, 303-306 (1971).
3. The visual array was a doll-mobile mounted in an internally illuminated box. The mobile rotated slowly during its display, was illuminated immediately after a correctly orienting head turn and remained on for 4.5 seconds. If no orienting head turn occurred within a 3-second interval following syllable presentation, the visual display was turned on automatically for 4.5 seconds.
4. No more than a single such turn was counted per presentation of a syllable. If the infant oriented to the sound, looked away, and then returned, only one anticipatory turn was counted. This is relevant only for the unreinforced trials, since in reinforced trials the initial head turn toward the stimulus triggered the visual array.
5. Syllables were recorded by a female voice on a studio quality recorder. Ten examples of each of the six syllable types were selected from a larger recorded set; multiple copies of these 10 examples were used to construct the stimulus tapes. The stimulus syllables, by adult standards, were all "good" examples of their type, and within each type, the "normal" acoustic variation found in repetitions of the same phone occurred.
6. This analysis simplifies some details of the experiment. The variable 'syllable series' (resulting from counterbalancing consonants across the vowel environments) is not represented. There is no indication, however, of a vowel grouping effect (e.g., of greater effects of reinforcement when /i/ and /u/ are paired – regardless of consonants – than when /i/ and /a/ are paired). There are differences, however, among the three syllable series in the degree to which the syllable grouping effects appear; the means for the "same phones" condition were higher than those for the

"different phones" condition in all three syllable series, but the effects were greatest in the pa-ku-pi and the pu-ki-pa groups. The pi-ka-pu group showed a superiority of reinforced over unreinforced syllables for both same and different phone conditions. Interpretation of differences among the three syllable series is risky, given that it is a comparison among subjects with only 6 subjects in each group. Our own guess is that with larger subject groups the differences among the syllable series would diminish.

One might also analyze the contrast between reinforced and unreinforced syllables in greater detail than first half of the session vs second half. A priori, there is no reason to expect that infants would not be able to make the appropriate sorting of the syllables sometime during the first two or three sessions. For some of the infants there is indication that this is the case. Our analysis simply focuses on those trials in which the contrast is sharpest across the full set of infants.

7. J. A. Fodor, M. F. Garrett, and Diana B. Shapero, "Discrimination among Phones by Infants," Quarterly Progress Report No. 96, Research Laboratory of Electronics, M.I.T., January 15, 1970, pp. 180-183.

8. In the earlier experiment we used visual scoring of infant head turns and a different reinforcing visual stimulus (motion pictures). Infants were also seated on their mothers' laps during the test sessions rather than in an infant seat.

9. C. D. Schatz, "The Role of Context in the Perception of Stops," Language 30, 47-56 (1954).

10. A. M. Liberman, P. C. Delattre, F. S. Cooper, and L. J. Gerstman, "The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants," Psychol. Monog. 68 (8, Whole No. 379) (1954).