CHAPTER 6

John R. Hayes

Carnegie-Mellon University

Herbert H. Clark

Stanford Univeristy

EXPERIMENTS ON THE SEGMENTATION OF AN ARTIFICIAL SPEECH ANALOGUE

The word, as an "object" of language, seems almost as real to our auditory experience as a ball does to our visual experience. Yet, on close examination, words appear to be only segments of the continuous flow of speech we hear around us. The word is, nevertheless, real. To quote Sapir (1921):

> Linguistic experience, both as expressed in standardized, written form and as tested in daily usage, indicates overwhelmingly that there is not, as a rule, the slightest difficulty in bringing the word to consciousness as a psychological reality. No more convincing test could be desired than this, that the naive Indian, quite unaccustomed to the concept of the written word, has nevertheless no serious difficulty in dictating a text to a linguistic student word by word; he tends, of course, to run his words together as in actual speech, but if he is called to a halt and is made to understand what is desired, he can readily isolate the words as such, repeating them as units. He regularly refuses, on the other hand, to isolate the radical or grammatical element, on the ground that it "makes no sense."

The experiments we are reporting here are concerned with the ontogeny of the word as a psychological unit; how do we come to perceive that the utterances of our language are segmented into words?

The segmentation problem might first appear to be no problem at all; language can be segmented because it is already segmented – either by pauses or by other well defined markers which identify the beginnings and endings of the words. While this suggestion is an obvious one, it has received surprisingly little empirical support. Consider the case for pauses, for example. Phenomenologically, we usually think of the words in speech as being separated by very short pauses, just as the words in print are separated by spaces. Examinations of sound spectrographs (for example, Lieberman, 1967), however, show that words in sentences most often flow into one another without any intervening pauses. We might say, however, that we perceive the segmentation because we know that it is possible to pause between the words. But this brings us back to the beginning problem: we have to know the segmentation before we can know where it is possible to pause.

Hayes, J. R. & Clark, H. H. Experiments on the segmentation of an artificial speech analogue. In Hayes, J. R. (1970), *Cognition and the Development of Language*. New York: Wiley.

A second candidate for segmentation is stress. As Jakobson, Fant, and Halle (1963) note, "in languages where the stress is bound to the initial. . . syllable. . .it functions as a boarder mark which denotes the beginning (or end) of the word." English, of course, is not such a language. In English, the rules for stress within a word depend in a very complex way on the abstract function of that word within the sentence as a whole (see Chomsky and Halle, 1968). In this sense, stress is a very imperfect indicator of word boundaries in English; one must know the syntactic rules themselves before one can use the stress rules in order to segment a sentence correctly. Lieberman (1965), indeed, has shown that linguists can judge the stress values of the words within sentences with very high consistency; however, these stress values are apparently not present in the acoustic signal itself, but are inferred by the linguists from their grammatical knowledge of English. Stress is therefore not in the spoken sentence itself, but in the mind of the listener; as such, it is of little help for the initial segmentation of speech.

Another possible segmentation device depends on the differences in the distribution of phonemes at the beginnings and endings of words. These differences could be used as cues for segmentation. For example, the "ng" sound is common at the end of English words but never occurs at the beginning. This kind of cue, however, does not seem very promising, since the overlap in the distributions is more striking than the differences.

As yet, none of the proposed segmentation markers, either separately or in combination, has been shown to be adequate to account for the phenomena of segmentation. Since the search for segmentation markers has been rather discouraging, we have turned in this paper to the examination of an alternative class of theories which do not require that a marker be present whenever a segment is perceived. We will call these the "recognition" theories. Recognition theories hold that a word is perceived as a segment not because markers currently delimit its beginning and end, but rather because the listener has recognized a pattern of phonemes (in the speech stream) that constitutes a word.

The speaker of any language must *learn* the phoneme patterns that constitute words. If segmentation is accomplished by a recognition mechanism, then he must necessarily learn the patterns from his experience with the language. It is widely recognized, in fact, that segmentation does depend on familiarity with the language that is being segmented (Neisser, 1967).

There are many possible mechanisms that would depend on linguistic experience in identifying words. We will define three types and distinguish them according to the kind of linguistic experience they employ.

1. Any *bracketing mechanism* identifies as words those phoneme sequences which have frequently been "bracketed" by segmentation markers. It should be noted, however, that this mechanism does not require that the markers be

present in the utterance currently being segmented, but only that they have been present in the listener's previous linguistic experience.

2. The *reference mechanisms* are those that identify as words the patterns of phonemes that have been strongly correlated with an external event or reference. Osgood and Sebeok (1965) have proposed such a mechanism, for in accounting for the unity of words, they say, "The word *apple* is heard. . .in a variety of constructions. . .but it is associated with a common perceptual sign and/or proximal experience (e.g., is accompanied by seeing and/or manipulating the same object apple)."

3. *Clustering mechanisms* are able to detect the recurrent patterns that constitute words, even without the aid of pauses or of meanings. These abstract mechanisms are able to measure crude correlations and to differentiate between the strong interphoneme correlations found within words and the weaker correlations across word boundaries. They identify as words the strongly correlated clusters of adjacent phonemes in the speech stream.

Harris (1955) has proposed just such a procedure for use by linguists in discovering the words of an unfamiliar language. His procedure is best described by working through one of his examples. He considered the utterance "He's clever," and its phonemic representation /hiysklever/. He then turned to informants in the language (as the source of his correlational information) to determine how many different phonemes can follow /h/. He found that there were 9 different phonemes that could act as successors to /h/. In the same way, Harris found that 14 phonemes could follow /hi/, 29 could follow /hiy/, 29 could follow /hiyz/, 11 could follow /hiyzk/, 7 could follow /hiyzkl/, and so on. High successor counts correspond to little constraint from the context and hence to low correlations among phonemes. Harris interpreted high successor counts as indicating that a segment boundary had been crossed. Hence, he located these boundaries just before the peaks in the successor count, as in the sample utterance /hiy.z.klever/, where the periods indicate boundaries.

We are especially interested in the clustering mechanism for three reasons. First, both the clustering mechanism and the bracketing mechanism have a property not shared by the reference mechanism. Since they operate solely on the auditory input, these two mechanisms have a capability of segmenting words before the words have been associated with a referent or have acquired a meaning, and therefore may play a role in the acquisition of reference and meaning. Suppose, for example, that a child is to learn the meaning of the word *milk* by understanding sentences such as "Drink your milk," "Here's our milk," and "Want some cold milk?" in the presence of milk. The child's task will be very much easier if the sentences are segmented into words before he starts correlating, since he would be able to eliminate such potential correlates as "rmilk" and "ilk" without testing them. Presumably, the task of acquiring grammatical information would also be simplified if sentences were segmented beforehand.

One could also argue from music that a reference mechanism, if such a mechanism exists at all, could not exist prior to a more abstract clustering or bracketing mechanism. Music has much the same superficial structure as does language in that it consists of notes, figures, leitmotivs, phrases, melodies, themes, and the like. For almost all kinds of music, there is nothing we could call the "referent" of a musical phrase, yet we can distinctly recognize the various hierarchical musical units – the recurrent themes of a fugue, the repeated occurrences of rhythmical patterns, and so on. This argues that there must be a nonreferential mechanism for the segmentation and recognition of at least some sound sequences.

The second reason for interest in the clustering mechanisms is an esthetic one. We originally became interested in this mechanism through analogy with a visual phenomenon. Even to a careful observer, an animal who remains motionless may merge perfectly into its background, but it will be quite visible as soon as it moves. When it moves, the correlations among the elements within its visual boundaries are much stronger than correlations between it and its surroundings. Given a clustering mechanism, the same difference in correlation would establish the boundaries of auditory objects, that is, words. We very much prefer to appeal to such general cognitive processes in explaining linguistic phenomena than to more special processes which apply uniquely to language.

Finally, we feel that it is scientifically more useful to try to demonstrate the existence of the clustering mechanism than of the bracketing mechanism. If patterns can be detected without segmentation markers, then we would assume that they can certainly be detected with segmentation markers. Thus, if we can demonstrate a clustering mechanism, we can very likely also demonstrate a bracketing mechanism.

THE EXPERIMENTS

We have set our goals in the following experiments at a very modest level. Our aim is merely to demonstrate the existence of a clustering mechanism in humans and to uncover some of its properties. We want to show that humans, given experience, can in fact identify recurrent patterns in a previously unfamiliar language and that they can do so simply by listening to the language without the aid of segmentation markers, reinforcements, referents, or teachers.

To demonstrate these points clearly, we had to meet the stringent requirement of (1) finding a language in which we could be sure that there were no segmentation markers of any kind, and (2) insuring that the speaker of the language would insert no extraneous markers, as he might do, for example, by unintentionally modifying the stress pattern at word boundaries. To satisfy these requirements, we constructed an artificial language entirely without segmentation markers and hired a computer to speak it to the subjects.

The Speaking Program

All of the sounds used in our experiments were combinations of square waves generated by a DDP-116 computer. To produce an audible square wave, the computer cycled repeatedly through a counting loop, alternately turning a voltage on or off. The modulated voltage was then amplified and fed either to a tape recorder or to a loudspeaker that transformed it into audible sound. The resulting sound resembled a rather poor quality organ or oboe. Changing the setting of the counting loop enabled the pitch of the tone to be changed in steps or to be made to vary slowly or rapidly to produce glides, warbles and trills. The available frequencies ranged from 250 to 2000 c.p.s. The computer could operate four such variable counting loops simultaneously. Thus, it was capable of accompanying itself in a quartet, a task which our artificial languages in fact required it to perform.

The stimulus materials which we presented were organized at three levels. By analogy with natural language, we will call these the level of the "phoneme," the level of the "word," and the level of the "language." The artificial phonemes were complex sounds in three or four voices which lasted between 0.2 and 0.3 second. Typical phonemes are shown in Fig. 6.1. We divided the phonemes into two classes in rough analogy to the vowels and consonants of human speech. Our artificial consonants were phonemes that changed markedly from beginning to end, such as the phoneme in Fig. 6.1A. The artificial vowels were the ones which, like that shown in Fig. 6.1B, changed little from beginning to end, that is, maintained approximately the same quality of sound throughout its duration.

The "words" were fixed sequences of 6 to 8 phonemes. All of the words were constructed by starting with a consonant and alternating consonants and vowels in the phoneme string. A word, then, might consist of consonant # 3, vowel # 5, consonant # 2, vowel # 1, and so on.

A language can be defined as a fixed set of words and their rules of combination. In the experiments described here, the languages were simply finite-state grammars with equal transitions between each word and another word and between each word and itself; in other words, the probability of any one of the four words occurring after any given word, including itself – that is, sampling with replacement – in this sequence was equal. What the subject (*S*) actually heard was a continuous stream of words, one right after the other without pause in between words, for a duration of about 45 minutes.

Experiment 1

The purpose of the first experiment was simply to attempt to demonstrate that humans can segment our artificial language into its component words by listening to it.

Procedure. The language used in Experiment 1 consisted of four words. Three of them were 8 phonemes long and one was 6 phonemes long. The phonemes used in constructing the words were 12 consonants and 7 vowels. Thus, many phonemes were used in more than one word.

The experiment involved two phases, a listening phase and a testing phase. In the listening phase, the *S* listened for about 45 minutes to the artificial language. The *S*s were told that they were to be tested after the listening phase to determine if they could recognize what they had been listening to. They were also told that they would win a prize of $5.00 if they got a perfect score on the test. The *S*s were not told that the language consisted of four words nor were they told that the language consisted of units three to four seconds in length. At no time during either the learning or the testing phase did the experimenter (*E*) give the *S*s any information concerning the correct segmentation of the language.

The premise of our testing procedure was that *S*s who had segmented the language appropriately would know where pauses ought to be placed, even though they had never heard pauses in the language before. In particular, they should be able to recognize that a string of phonemes with pauses placed at the word boundaries was more "familiar" than a string with pauses placed in other locations, that is, in the middle of words.

The test consisted of 40 items. Each test item consisted of two sets of four words, an "A" set and a "B" set. One of the sets consisted of the four words of the language in random order separated by pauses at the word boundaries. The other set consisted of four new words constructed from exactly the same string of 30 phonemes as the first set. There were just two changes: (a) each pause was shifted two phonemes to the left, and (b) the last two phonemes (29 and 30) were placed first. All of the words generated by this procedure consisted of sequences of phonemes that could occur in the original language and that had in fact been presented many times during the listening period. Furthermore, any combination of these words would yield a sequence of phonemes which was legal in the original language. The new set of words, then, contained no unusual transitions from one phoneme to another that might have served as a cue for discrimination.

The *S*'s task in the test was to indicate for each test item which set of words sounded most similar to the language he had heard in the listening phase. The positions of the two sets were counterbalanced over the 40 test items.

Results. The results provide clear evidence of segmentation. The distribution of test scores (shown in Fig. 6.2) was biased strongly in the "correct" direction. A sign test (Siegel, 1956) on the number of *S*s achieving better (or worse) than the chance score of 20 revealed a bias in the expected direction at the 0.02 level of confidence.

Fig. 6.2 also shows that none of the *S*s come even close to achieving a perfect score on the test. The effect obtained with 45 minutes of listening was a small but measurable *tendency* to recognize the appropriate segmentation.

Our initial hope that the 40 test items would provide independent measures of *S*s' discrimination proved to be quite unfounded. A control *S*, given the test without the listening period, achieved a score (of 10) which would have been significantly poorer than chance if the successive judgments were independent. *S*s who had not been exposed to the materials in the listening phase could still adopt a consistent response strategy and score either well above or well below chance. They could not, of course, determine the *direction* in which their scores differed from chance. Therefore, we treated the scores on the 40-item test simply as indicating the direction in which each *S* differed from chance.

Discussion

By serving as pilot subjects, the experimenters were able to obtain a subject's-eye-view of the segmentation process. In the experimenters' subjective view, the process seems to proceed roughly as follows. At first, the sound stream seems quite amorphous and featureless. After perhaps a minute of listening, an event – perhaps a phoneme or part of a phoneme – stands out from the stream. When

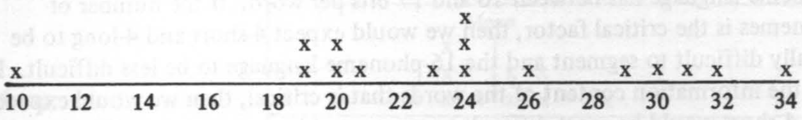

Fig. 6.2. Distribution of scores in Experiment 1.

the event has recurred several times, the listener may notice that it is typically preceded or followed by another event. The combination of events can in turn be related to events that happen in its neighborhood. Recognition of a word, then, seems to proceed from perceptually distincitive foci outwards in both directions toward the word boundaries. Presumably, the process would tend to stop at the word boundaries because the correlations across the boundaries are weak.

If segmentation does proceed by relating a growing focus to neighboring event, then we would expect that any operation which reduces the uniqueness of the relation between the focus and its neighbors would impede segmentation. One way to reduce the uniqueness of this relation is to reduce the number of phonemes. If there are few phonemes, each must occur in several different combinations. We must expect, then, that a language constructed from many phonemes will be easier to segment than a language that is constructed from few phonemes. The second experiment was conducted to test this hypothesis.

Experiment 2

Procedure. In the second experiment, we employed three languages – a 16-phoneme language, and two four-phoneme languages. Each language consisted of four words. In the 16-phoneme language, the consonants were selected from a list of 12 alternatives, and the vowels from a list of four alternatives. In the four-phoneme languages, consonants and vowels were each chosen from lists of two alternatives. These lists were sublists of those used for the 16-phoneme language, and they were chosen so that the phonemes in each language would be as distinct as possible.

The words of the 16-phoneme language and of one of the four-phoneme languages (to be called 4-short) were six phonemes in length. The words of the other four-phoneme language (to be called 4-long) were 12 phonemes in length. We included the language 4-long in the experiment as a test of the hypothesis that the critical factor in segmentation is the information content of the words rather than the number of phonemes in the language. Notice that 4-short has six bits of information per word, 4-long has 12 bits per word, and the 16-phoneme language has between 16 and 17 bits per word. If the number of phonemes is the critical factor, then we would expect 4-short and 4-long to be equally difficult to segment and the 16-phoneme language to be less difficult. If it is the information content of the words that is critical, then we would expect that 4-short would be most difficult to segment, 4-long next most difficult and the 16-phoneme language least difficult.

The three languages were each constructed as in Experiment 1, that is, as finite-state languages with all transition probabilities equal. Each language was presented to a different group of *S*s in the same manner as in Experiment 1, but in this case the tests were shortened from 40 items to 24.

Results. The results are shown in Fig. 6.3. Sign tests on the numbers of *S*s who achieved scores above (or below) the chance score of 12 revealed significant segmentation at the 0.05 level of confidence for the 16-phoneme language, but not for either of the four-phoneme languages. Mann-Whitney tests revealed significant differences in segmentation scores between the 16-phoneme language and each of the four-phoneme languages at the 0.05 level of confidence, and no significant difference between the two four-phoneme languages.

Discussion

The results are consonant with the hypothesis that difficulty of segmentation depends on the number of phonemes in the language, and they are not consonant with the hypothesis that difficulty depends on the information content of the words. The results, therefore, provide support for our position that segmentation is a process which proceeds by relating a growing focus to neighboring events.

We hope to test this position further in future work by attempting to identify and to control the occurrence of the highly distinctive events with which the focus starts. If our position is correct, it should be possible to control the order in which the words of a language are segmented by controlling the occurrence of these events in the words.

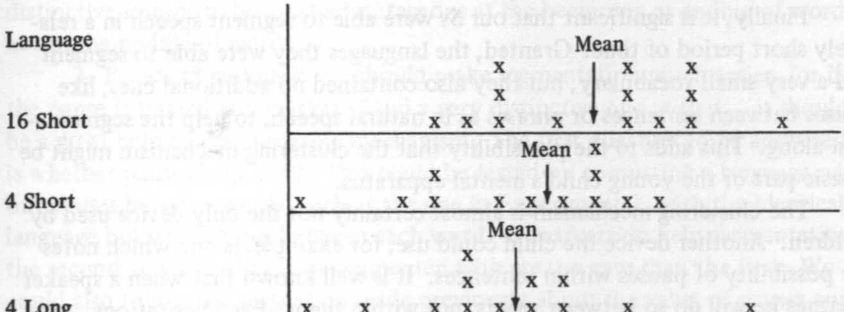

Fig. 6.3 Distribution of scores by language in Experiment 1.

DISCUSSION

These two experiments demonstrate that humans do, in fact, have a clustering mechanism able to segment artificial speech. It is a mechanism which (a) can segment completely unutterable sounds, (b) works on "speech" that has no semantic and no significant syntactic structure, and (c) requires relatively little time – about three-quarters of an hour in our experiments – to come to at least some parsing of the speech. It seems to us that these are important properties of a mechanism that would be useful to a child first trying to sort out the sounds he hears around him. That is, if we could assume that very young children relied on such a mechanism – at least in part – we could make substantial progress toward understanding the very first stages in the child's acquisition of language.

It is important, first, that the sounds that our subjects parsed into words were speechlike, although unutterable. This situation is analogous to the one the child faces in his first encounters with language. It is very unlikely that he has the repertoire of sounds necessary to produce the first words he is able to understand. To segment his parent's speech successfully, he need only be able to *hear* the various distinctions that are being made in the phonological structure. Then, by listening for the correlations, which is what the clustering mechanism does, he is able to decide just where the word boundaries are in the adult speech.

It is just as important, however, that this mechanism is able to operate completely independently of semantic and syntactic considerations – except for the simple syntactic constraint that speech *is* made up of words. This frees the initial segmentation mechanism in the child from necessarily being a complicated device which, say, "associates" what is said with what is seen, as Osgood's reference mechanism would have it do, or which makes and tests hypotheses about the syntactic structure of speech.

Finally, it is significant that our *S*s were able to segment speech in a relatively short period of time. Granted, the languages they were able to segment had a very small vocabulary, but they also contained no additional cues, like pauses between sentences or phrases as in natural speech, to help the segmentation along. This adds to the plausibility that the clustering mechanism might be a basic part of the young child's mental apparatus.

The clustering mechanism is almost certainly not the only device used by children. Another device the child could use, for example, is one which notes the possibility of pauses within sentences. It is well known that when a speaker breathes he will do so between words, not within them. Even hesitations – the interruptions in speech typical of someone trying to think of the next word – will almost always occur at word boundaries (Maclay and Osgood, 1959). A mechanism that is able to note these pauses and hesitations, then has additional information for the segmentation of speech into words. It could be argued, however, that this device is really only a specialized clustering mechanism. In noting speech interruptions, it is still only attending to the high intercorrelations of the sounds within words, that is, to the integrity of the word as a unit, just as a clustering mechanism would do. In similar mechanisms, however, the child could make use of other kinds of information about words, such as, phonological and syntactic properties, in conjunction with the simple clustering mechanism.

This conception of how the child first comes to segment speech leads to several suggestions for future work in segmentation of artificial speech. Most of the following suggestions are aimed at a clustering mechanism used in conjunction with other phonological and syntactic features of speech.

1. Very little is known about what makes a phoneme "distinctive" in our artificial speech, but it is certain that we hear particular sounds that stand out over the background of the other sounds. In several preliminary studies, we have attempted to identify the more distinctive sounds in our language and have done so with some success. Further experiments are needed, however, to show whether the distinctiveness of these phonemes can be helpful to the segmentation process. Our conception of the ongoing process of word perception is that the listener starts with a distinctive sound and then begins to notice its correlations with the sounds around it. Now, if each word of a language began with a distinctive phoneme, the listener would have the greatest opportunity to notice (a) that the sounds before that phoneme were not correlated with it and (b) that those after it were. This would immediately give him the word boundaries. But if the phoneme came in the middle of the word, he would not perceive the boundaries so quickly, for they would emerge only after he had built up all the correlated sounds around the distinctive phoneme. Thus, we would expect distinctive sounds to be most advantageous at the beginning or ending of words, not in the middle of words.

2. Pauses of various kinds should make segmentation much easier, for if the pause is treated as a marker – and a very distinctive one at that – it should be a great help to the clustering mechanism. The first question to be answered is whether pauses help at all. This could be tested by comparing a language without pauses between words, such as the one in Experiment 1, with the identical language but with pauses between each word. If markers do help segmentation, the second language should be segmented with greater ease than the first. We could also follow the suggestion made previously about the value of pauses and hesitations in natural language. One could compare one language with no pauses with another in which pauses occurred between words some part of the time. Our prediction would be that the second language should be more easily parsed than the first.

3. Adult speech abounds with prefixes, suffixes, and inflections, but children in their first attempts to speak ignore them completely. The tendency to ignore them could occur because children really do not perceive affixes and inflections as an integral part of the words they are attached to; instead, children hear them first as segmentation markers.

This conception of the function of affixes and inflections for the child can be tested in an artificial speech analogue. Two languages could be constructed so that the first has no inflection whatsoever, like the language in Experiment 1, and the second has one or more affixes that are tacked onto each of the words in the language interchangeably. The affixes in the second language would serve almost the same function as pauses in that both affixes and pauses occur very frequently in the language, always at word boundaries, and uncorrelated with any particular word in the language. The single test for both languages would be one that does not include affixes, but which is otherwise analogous to the text in Experiments 1 and 2. Even though the test for the second language would not include the affixes, the second language should show better segmentation than the first, for the affixes, acting as highly redundant markers, allow the segmentation to take place more quickly.

4. Other grammatical properties should presumably help in the segmentation process too. One hypothesis, for example, is that languages with phrase-structure grammars should be easier to parse than those with finite-state grammars. Furthermore, in languages with phrase-structure grammars, the word boundaries which coincide with the greatest number of constituent boundaries should be the most quickly perceived, for it is at these boundaries that there is the greatest possibility of noticing the lack of correlation between the sounds on either side of the boundaries. Intonation is another natural language property which undoubtedly aids segmentation. Although intonation contours are a problem for our artificial speech synthesizer as it now exists, the presence of such contours – or their analogues – should also help in the perception of the surface structure constituent boundaries, and hence in the segmentation of words within these constituents. It should be noted that these grammatical considerations can be tested independently of, or in conjunction with, the previous considerations of affixes, pauses, or distinctive sounds.

Finally, we sound one note of warning: the conclusions from these experiments on adults may possibly not be applicable to children at all. The present experiments, and their accompanying discussion, argue only for the *plausibility* of the mechanism in children's acquisition of language, not for its precise existence. To complete the case, experiments like the present ones must be repeated, under the appropriate conditions, on children of about a year old. The technicalities of these experiments, of course, would be very difficult, yet there are available methods that make such experiments feasible. This, too, we leave to the future.

In conclusion, the present experiments have demonstrated the existence of a very primitive clustering mechanism for the segmentation of artificial speech into words. This mechanism appears to operate only on the information available in the intercorrelations between the successive sounds in the speech stream; it identifies as words the clusters of sounds that consistently recur in an unbroken sequence. We have suggested some of the implications of this kind of mechanism for the ontogeny of segmentation in the child and some of the directions research on this problem might take in the future. In a word, however, it is clear, that, in our quest for the child's quest of his first word, we have barely uttered the first word.

REFERENCES

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

Harris, Z. S. From phoneme to morpheme. *Language,* 1955, 31, 190-222.

Jakobson, R., Fant, G., and Halle, M. *Preliminaries to Speech Analysis.* Cambridge, Mass: M.I.T. Press, 1963.

Lieberman, P. *Intonation, Perception, and Language.* Cambridge, Mass.: M.I.T. Press, 1967.

Lieberman, P. On the acoustic basis of the perception of intonation by linguists. *Word,* 1965, 21, 40-55.

Maclay, H., and Osgood, C.E. Hesitation phenomena in spontaneous English speech. *Word,* 1959, 1, 19-44.

Neisser, J. *Cognitive Psychology.* New York: Appleton-Century-Crofts, 1967.

Osgood, C. E., and Sebeok, T.A. *Psycholinguistics: A Survey of Theory and Research Problems.* Bloomington, Ind.: Indiana University Press, 1965, pp. 55-56.

Sapir, E. *Language.* New York: Harcourt, Brace & World, 1921.

Siegel, S. *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill, 1956.

REFERENCES

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

Harris, Z. S. From phoneme to morpheme. *Language,* 1955, 31, 190-222.

Jakobson, R., Fant, G., and Halle, M. *Preliminaries to Speech Analysis.* Cambridge, Mass: M.I.T. Press, 1963.

Lieberman, P. *Intonation, Perception, and Language.* Cambridge, Mass.: M.I.T. Press, 1967.

Lieberman, P. On the acoustic basis of the perception of intonation by linguists. *Word,* 1965, 21, 40-55.

Maclay, H., and Osgood, C.E. Hesitation phenomena in spontaneous English speech. *Word,* 1959, 1, 19-44.

Neisser, J. *Cognitive Psychology.* New York: Appleton-Century-Crofts, 1967.

Osgood, C. E., and Sebeok, T.A. *Psycholinguistics: A Survey of Theory and Research Problems.* Bloomington, Ind.: Indiana University Press, 1965, pp. 55-56.

Sapir, E. *Language.* New York: Harcourt, Brace & World, 1921.

Siegel, S. *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill, 1956.