Converging Measures of Speech Segmentation in Preverbal Infants

JAMES L. MORGAN
Brown University

Two studies using novel extensions of the conditioned head-turning method examined contributions of rhythmic and distributional properties of syllable strings to 8-month-old infants' speech segmentation. The two techniques introduced exploit fundamental, but complementary, properties of representational units. The first involved assessment of discriminative response maintenance when simple training stimuli were embedded in more complex speech contexts; the second involved measurement of infants' latencies in detecting extraneous signals superimposed on speech stimuli. A complex pattern of results is predicted if infants succeed in grouping syllables into higher-order units. Across the two studies, the predicted pattern of results emerged, indicating that rhythmic properties of speech play an important role in guiding infants toward potential linguistically relevant units and simultaneously demonstrating that the techniques proposed here provide valid, converging measures of infants' auditory representational units.

<table>
<thead>
<tr>
<th>speech perception</th>
<th>speech segmentation</th>
<th>rhythm</th>
<th>language development</th>
</tr>
</thead>
<tbody>
<tr>
<td>conditioned head-turning method</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The basic problem of segmentation is readily apparent to anyone listening to speakers conversing in an unfamiliar language: what one hears is a babble devoid of readily identifiable words, phrases, or sentences. Phenomenologically, the contrast with what one experiences in listening to fluent speech in a familiar language could scarcely be more striking. Solving the segmentation problem is an important challenge facing infants in the initial stages of acquiring language. The development of segmentation skills is absolutely prerequisite to further acquisition of language, for if a language learner cannot break input utterances into their constituent parts, it will not be possible either to learn what individual parts mean or how these parts fit together. Inasmuch as units emerging from segmentation serve as input to further stages of analysis, the nature of analytic capacities recruited in language acquisition is dependent upon the output of segmentation.

Despite the clear importance of segmentation, however, little is known about the nature or development of requisite perceptual capacities. With the exception of a series of studies conducted by Hirsh-Pasek, Jusczyk, and Kemler Nelson (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright Cassidy, 1989), little research has even attempted directly to address perceptual problems of speech segmentation by infants. Without doubt, a major impediment to further work in this area is the lack of reliable methods for assessing infants' representational units.

The premise of this article is that progress in understanding infant speech segmentation must rely on converging results emerging from methods addressing complementary aspects of representational units. Two such techniques for investigating infant speech segmentation will be introduced. Substantive results arising from application of these techniques converge in suggesting that rhythmic properties of speech play an important role in guiding infants toward potential linguistically relevant units. These converging results simultaneously validate the techniques proposed here.

The techniques introduced below are intended to supplement the method developed by Hirsh-Pasek, Jusczyk, and Kemler Nelson. Their work has employed a pause-insertion...
technique adapted from earlier studies of adult segmentation of unfamiliar languages (Hayes & Clark, 1970) in which short pauses are added to natural speech at points that either coincide or fail to coincide with syntactic junctures (i.e., clause or phrase boundaries). Infants' preference for the "coincident" versus "noncoincident" stimuli is assessed in a visual preference paradigm. Initial results reported in Hirsh-Pasek et al. (1987) showed that 7-month-old infants listened longer to clause-boundary-coincident stimuli. Subsequent work in Kemler Nelson et al. (1989) showed that this preference obtains when the stimuli consist of samples of child-directed speech, but not when the stimuli consist of adult-directed speech samples. Recently, Jusczyk et al. (1992) have extended earlier findings to show infant preference at 9 months for phrase-boundary-coincident stimuli as well. A series of studies employing low-pass filtered speech demonstrates that this preference is based on the suprasegmental, or prosodic, aspects of speech.

The pause-insertion technique has many desirable features. It employs natural speech, can be used with infants across a broad range of ages, and can be administered efficiently, as training infants to make novel discriminations or responses is not required. The technique draws its rationale from the notion that units resist interruption. Hence, speech stimuli in which prosodic units are preserved (by inserting pauses only at junctures between units) will sound more natural than speech stimuli in which such units are violated. On this view, patterns of infants' preferences may be interpreted as indicating that (prosodic) clauses and phrases function as early perceptual units.

However, as Jusczyk et al. (1992) noted, these results are also compatible with a simpler explanation. Clause boundaries are commonly cued in spontaneous speech by final vowel lengthening (Bernstein Ratner, 1986; Cooper & Paccia-Cooper, 1980) and discontinuities in fundamental frequency contour (Cooper & Sorensen, 1981). Clause boundaries in child-directed speech (most of which occur at the ends of sentences) are also typically accompanied by pauses (Broen, 1972). Major phrase boundaries are also often cued by final vowel lengthening and frequency discontinuities, both in adult-directed (Cooper & Paccia-Cooper, 1980) and child-directed (Morgan, 1986) speech. Thus, vowel lengthening, frequency discontinuities, and pausing are correlated with one another. When pauses are inserted at points in speech where these other cues are lacking, such correlations are violated. Thus, infants might respond differently depending on whether naturally occurring correlations among prosodic cues are preserved or violated. Hence, the conservative interpretation of the results of studies using the pause-insertion technique is that infants are sensitive to the constellations of acoustic cues that may serve to mark linguistic units in speech. As Jusczyk et al. (1992) concluded, however, "... showing that infants are sensitive to such acoustic correlates is only a first step in understanding how the infant discovers the linguistic units that function in the native language. It remains to show that these sorts of acoustic markers are actually used in segmenting the linguistic input" (pp. 289–290).

To determine what information in input contributes to infants' segmentation of the speech stream, it is necessary to have methods capable of assessing how infants represent their linguistic input—in particular, methods capable of ascertaining how infants have organized sequences of speech sounds in the signal into representational units. In developing such methods, it is reasonable to begin by contemplating fundamental properties of representational units and considering what the processing consequences of these properties may be.

Essentially, the construct "representational unit" entails two complementary properties: external individuality and internal coherence. External individuality means that, at a given level of analysis, the behavior of a group of elements that constitutes a unit is indistinguishable from the behavior of a single element. For example, Miller (1956) argued that working memory is capable of holding a certain number of chunks. A chunk could be a single digit or a large number: Regardless of the complexity of its internal structure, a chunk takes up a fixed amount of working memory. Certainly, Miller's finding that memorial and processing capacities vary according to the fashion in which information is represented is among the most fundamental in all of cognitive psychology. The same phenomenon can be seen at several different levels of linguistic analysis, as morphologi-
organ, 1986) ng, frequency correlated with exerted at points are lacking. Thus, infants responding oniations among olated. Hence, the results of a technique is tstalizations of ark linguistic l. (1992) context infants are es is only a the infant distinc- tion in the ow that these usually used in sp. 289–290).

In input con- of the speech thods capable their linguistics capable of eorganization of the signal into loping such by contem- f representation the process- esses may be.

Presentational properties: l coherence.

At a given group of ele- stinguishable element. For hat working rtain number le digit or a xplexity of s up a fixed ly, Miller's ng capacities ich informa- most funda- lology. The ernal differ- morphologi-
cal word-formation rules operate in a given fashion regardless of the number of phonemes in the stem, rules moving noun phrases from one syntactic position to another apply without regard to the number of words in the phrase, and so forth. The advantage of such hierarchical organization is clear: It buys cognitive economy.

However, not all juxtapositions of elements yield units. A set of digits, for example, can be considered as constituting a number provided that the digits are interpreted as participating in a base system, denoting, say, so many thousands, hundreds, tens, and ones. Similarly, whether a given sequence of words constitutes a noun phrase depends upon the grammatical system in which they participate (a theory of grammar will provide definitions of both noun phrase and constituency). These restrictions on the constituency of units reflect the requirement that they have internal coherence. The properties upon which coherence is defined vary depending on (the theory of) the system of which the units are a part. For example, the Gestalt psychologists argued that similarity was an important factor governing the organization of visual arrays (Koffka, 1935). However, similarity plays a minor role in grammar: Syllables are not sequences of high vowels or alveolar consonants, and noun phrases are not strings of nouns. Rather, at many levels of grammar, properties of alternation, contrast, and opposition tend to govern which elements may be combined to form units: Syllables consist of vocalic nuclei flanked by consonantal onsets and codas, and noun phrases consist of noun heads flanked by specifiers and complements drawn from other grammatical categories. In any event, like external individuality, the property of internal coherence confers clear processing advantages on the perceiver. Knowledge of the system and the bases upon which coherence is defined allows the perceiver to anticipate what will ensue and to allocate processing resources accordingly.

To address early speech segmentation, variants of the conditioned head-turning method widely used in studies of infant speech perception (Aslin, Pisoni, & Jusczyk, 1983; Kuhl, 1985) have been designed to exploit each of these fundamental properties of representation units. First, the discrimination maintenance technique takes advantage of the processing consequences of external individuality. In this technique, infants are first trained to discriminate between a pair of syllables presented in isolation. Then, infants hear these syllables embedded in multisyllabic strings. Across conditions, the prosodic and distributional properties of the added syllables (henceforth, “context syllables”) are systematically varied. In some conditions, these properties are designed to encourage infants to group the context syllables into higher-order units, whereas in others they are designed to discourage this. Inasmuch as combining individual elements into higher-order units (which may then themselves be treated as individuals) tends to conserve processing resources, infants in the former conditions should prove best able to maintain accurately their discriminative responding.

Methods akin to the discrimination maintenance technique proposed here have a long history in psycholinguistics. Miller and Selfridge (1950) used recall to assess the representational organization of linguistic materials, showing that recall improved as the stimulus materials more closely approximated the statistical regularities of English. Savin and Perchinock (1965) gave subjects both sentences and lists of unrelated words to recall and indexed subjects’ representations of the sentences by the numbers of unrelated words they recalled. Infants are not amenable to tasks requiring overt recall of linguistic material. However, discrimination performance in the head-turning paradigm involves a memory component (Kuhl, 1985) and should therefore also be susceptible to variations in stimulus representation.

The discrimination maintenance technique has previously been used in studies of infant speech perception by Goodsitt, Morse, VerHoeve, and Cowan (1984) and Goodsitt, Morgan, and Kuhl (1993). The experiments in Goodsitt et al. (1993) included conditions in which a pair of context syllables occurred in either distributionally regular or distributionally irregular patterns. As predicted, infants receiving the former type of stimuli proved to be better able to maintain the discriminative response on which they had been trained. These results are compatible with the possibility that these infants represented the context syllables as a single unit. If this is so, however, it also ought
to be possible to find effects of internal coherence in comparable circumstances.

As a means of measuring processing consequences of internal coherence of representational units in infant speech segmentation, a noise detection technique is introduced here. In this technique, infants are trained to turn their heads upon hearing a brief buzz. Initially, buzzes are presented in pauses between strings of syllables. After infants have acquired the response, the buzzes are presented at different points within the strings of syllables. As with the discrimination maintenance technique, across conditions, prosodic and distributional properties of the strings are systematically varied to either encourage or discourage infants' grouping of certain syllables into higher-order units. If infants do group particular syllables, then their response latencies may differ depending on whether the buzzes are located between the grouped syllables or between ungrouped syllables. In contrast, if infants do not group syllables into higher-order units, their response latencies should not differ as a function of buzz location.

This technique is an adaptation of the "click displacement" method pioneered by Ladefoged and Broadbent (1960) and widely used in psycholinguistic research over the following decade. The rationale of this method was that the subjective temporal displacement of an extraneous sound superimposed on a speech stimulus (or other internally structured auditory stimulus) is illustrative of the psychological tendency of perceptual units to resist interruption. As used with adults, this technique usually involved auditory presentation of a sentence accompanied by an extraneous click. The subjects' task was to write down the sentence they had just heard, indicating where in the sentence the click had occurred. Subjects commonly indicated that the clicks had occurred at or near boundaries of phrases, even when the clicks occurred in the middle of phrases. However, interpretive complications soon became unmanageable. Because the click-attracting boundaries could be defined (depending on which of the then-viable linguistic theories the researcher adopted) either in terms of the more syntactic surface structure or the more semantic underlying structure, it was not clear which of these had a more profound effect on perception.

Because the subjects were typically involved in a memorial task, moreover, it was not clear which level of processing, encoding or recall, was responsible for the click displacement (though studies by Abrams & Bever, 1969; Bond, 1972; Flores d'Arcais, 1978; and Holmes & Forster, 1970, using on-line tasks, found differences in response latency as a function of objective click location: Clicks within syntactic units were responded to more slowly than were clicks between units). Work by Reber and Anderson (1970) showing substantial effects of response bias (subjects reported hearing clicks at phrase boundaries even when no clicks had been presented) led to the abandonment of this method in adult research.

These concerns are moot when applied to a task involving on-line measurement of young infants' responses to extraneous sounds superimposed on speech stimuli. Neither the syntactic nor semantic structure of such speech can condition infants' responses, because preverbal infants presumably lack the requisite knowledge. Only perceptually available information—segmental, prosodic, or distributional—can affect infants' structuring of speech strings. Moreover, if the primary measure is one of latency, the notion of response bias will not apply.

The noise detection technique has not previously been used in studies of infant speech perception; therefore, its validity as a means of measuring perceptual grouping is unknown. Indeed, measures of latency (RT) have rarely, if ever, been used in studying infant speech perception. However, in a recent series of articles, Haith and his colleagues (Canfield & Haith, 1991; Haith, Hazan, & Goodman, 1988; Haith & McCarty, 1990) have shown that RT can be used as a means of assessing infants' organization of perceptual arrays. In these studies, 3-month-old infants were presented with visual stimuli in sequences that varied in predictability. Infants' latency of fixation on new stimuli was shorter for more predictable sequences, suggesting that infants were able to discern the structures of the sequences and to anticipate what would occur next.

The logic of the noise detection technique differs from this somewhat. If an infant represents a sequence of speech sounds as a unit (with internal coherence), then perception of
Involved in not clear g or recall, spacement over; 1969; 1978; and line tasks, as as a function, as within more slowly Work by substancs reported even when the anchor pheme can preverbal knowledge- or informational—strings. is one of not previ-per of means of unknown. e rarely, pre of letters, & Haith, 1988; Haith 1987; can be organized studies of 3-4th visual development as stimuli sequences, iscern the anticipate technique as a unit perception of the initial sound(s) in a sequence will establish in the infant anticipation of subsequent sounds. If the infant is set to perceive a particular sound, then it will be more difficult for the infant to perceive and react to an unexpected auditory event (it is from such a perceptual set that units gain their tendency to resist interruption). Therefore, infants’ reactions to extraneous noises that occur within units should be slower than reactions to noises occurring between units. This predicted pattern of results parallels that found in studies of adult on-line click detection.

The goal of this article is to bring both the discrimination maintenance and noise detection techniques to bear in studying contributions of rhythmic properties of syllabic strings to infant segmentation. Note that a complex pattern of results is predicted if both techniques provide valid, sensitive measures of segmentation. Under both techniques, strings presented to experimental groups of subjects are distributionally and prosodically structured to encourage assembly of several speech segments into larger units; strings presented to control groups are not so structured. Both techniques predict differential results when larger perceptual units come into play, but they make predictions in opposing directions. That is, discrimination maintenance performance will be enhanced when grouping occurs, due to the attendant cognitive economy that is achieved. However, latency of responses to extraneous noises will be degraded when grouping occurs (and when noises are presented within perceptual groups), due to the tendency of perceptual units to resist interruption. Moreover, under the discrimination maintenance technique, grouping is manifest primarily by a between-conditions main effect; under the noise detection technique, grouping is manifest primarily by a condition by noise location interaction.

These discrimination maintenance and noise detection techniques exploit fundamental, but complementary, properties of psychological units. Converging results arising from these studies will simultaneously validate the methods proposed here and provide information about environmental cues which foster infant segmentation, providing particularly compelling evidence for the nature of infants’ representational units.

EXPERIMENT 1

In this experiment, the discrimination maintenance technique was used to investigate effects of rhythmic and distributional cues on infants’ segmentation of simple strings of syllables. Three conditions were included. In the first, rhythmic properties of individual stimulus strings and distributional properties of speech segments across stimulus strings provided correlated cues to grouping of a pair of syllables. In the second condition, these cues were uncorrelated: Rhythmic properties cued grouping, but distributional properties did not. In the third condition, neither property cued grouping. It was predicted that infants in the first (correlated cues) condition would be better able to maintain a speech discrimination initially trained in isolation than would infants in the third (no cues) condition. In light of the results of Jusczyk et al. (1992) showing that infants are sensitive to constellations of correlated cues, it was also predicted that performance in the second (uncorrelated cues) condition would more closely resemble that of the third condition.

Method

Subjects

Infants approximately 8 months of age were recruited from birth announcements in local newspapers in Minneapolis-St. Paul metropolitan area. Fifty-nine subjects were tested to attain the final sample of 36. Twelve subjects were assigned at random to each of three stimulus conditions. Subjects were excluded from the study for the following reasons: excessive fussiness or crying (3), equipment failure or experimenter error (2), failure to meet the predetermined training criterion in the first session (14), interference from uncooperative siblings (1), and difficulties scheduling subsequent testing sessions within a 3-week period (3). The mean age of the infants completing the study was 234 days (range = 216–252 days). The three sessions required for each infant were scheduled so that the time from the initial training session to the final test session was no more than 21 days and the time from one session to the next was no more than 10 days.

Stimuli

The stimuli used in this experiment were edited versions of the stimuli used in Goodlett et al. (1993). Study 2. The original stimuli included four syllables, one token each of [de], [te], [ke], and [ge], spoken in isolation with a flat intonation contour by a trained adult female. These syllables, which had been digitized at 10,000 samples per second through a 5-kHz low-pass filter, edited, and stored on disk using ILS software, were matched for duration (500 ms), fundamental frequency contour, and pitch. For the present study, these syllables were further edited using custom waveform analy-
six software. The tokens of [de] and [ti] were shortened to 425 ms each by excising 18 and 19 individual glottal periods, respectively; the original 500-ms versions of [ko] and [ga] were retained as is (below, these are referred to as “long versions” and denoted by [ko] and [ga]). Two additional 350-ms versions of [ko] and [ga] were created by excising 37 and 40 individual glottal periods, respectively (below, these are referred to as “short versions” and denoted by [ko] and [ga]). Zero crossings were chosen as the beginning and ending points of each excited portion, so that no transients were created by resplicing the waveforms.

The syllables [de] and [ti] served as contrast syllables; infants were conditioned to turn their heads when they heard one or the other of these, as described below. The syllables [ko], [k], [ga], and [ga] served as context syllables. Between-subjects stimulus conditions were created by combining the two contrast syllables and two of the four context syllables to form sets of four trisyllabic strings. In each string, a contrast syllable occurred in either initial or final position, and the two contrast syllables occurred adjacent to one another. In all stimulus conditions, syllables within the trisyllabic strings were separated by 50-ms silences.

In the invariant trochaic (correlated cues) condition, both the rhythmic pattern and the relative position of the two context syllables were fixed. A trochaic rhythm pattern was used—a long syllable followed by a short syllable. Two subconditions were included: six infants were assigned at random to each. In one, the contrast was [kog]; in the other, the contrast was [gakog]. The sets of strings included in each subcondition are shown in Table 1. In this condition, both distributional and rhythmic properties cued grouping of the context syllables.

In the variable trochaic (uncorrelated cues) condition, the rhythmic pattern over the two context syllables was fixed but the syllables themselves shifted position relative to one another, depending on whether the contrast syllables appeared in initial or final position. The rhythm pattern was again trochaic. Two subconditions were included; six infants were assigned at random to each. The sets of strings included in each subcondition are shown in Table 1. In this condition, rhythmic and distributional cues were uncorrelated: Rhythmic cues indicated that the context syllables should be grouped, whereas distributional cues failed to do so.

In the variable mixed (no cues) condition, the durations of each of the context syllables were fixed and the syllables shifted position relative to one another, depending on whether the contrast syllables appeared in initial or final position. Thus, the rhythm pattern over the context syllables also varied. Two subconditions were included; six infants were assigned at random to each. The sets of strings included in each subcondition are shown in Table 1. In this condition, neither distributional nor rhythmic properties cued grouping of the context syllables.

Apparatus
Infants were tested in a sound-treated laboratory room with an adjoining control booth. Trial duration, stimulus presentation, and delivery of reinforcement were controlled by custom-designed software. Stimuli were presented through a 5000-Hz low-pass filter, a custom programmable attenuator, and a Sansui amplifier which was connected to an Omega loudspeaker located in the testing room with the infant. Reinforcement was provided by illuminating the interior of a smoked Plexiglas box above the speaker and activating a motorized animal figure inside the box.

Procedure
The experiment consisted of three main phases: shaping, criterion, and multisyllabic testing. The shaping and criterion phases were completed in an initial training session; the multisyllabic testing phase spanned two subsequent test sessions. Through all sessions infants were seated on their parents’ laps at a small table. An assistant seated directly across from the infant maintained the infant’s attention at midline by displaying and manipulating an assortment of toys. In the control booth, the experimenter monitored the infant through a one-way mirror and initiated trials upon observing that the infant’s attention was focused at the midline. Reinforcement was delivered contingent on the infant responding to a stimulus change with a head-turn toward the loudspeaker, which was located 90° from midline on the infant’s right. The experimenter served as sole judge of whether the infant turned his or her head during trials, depressing a button to signal to the computer than a head-turn had occurred. Reinforcement began when the experimenter signaled a head-turn and lasted for the duration of the trial. To preclude bias, throughout all sessions, parent, assistant, and experimenter all listened to music over around-the-ear headphones.

In the shaping phase, infants heard the contrast syllables ([de] and [ti]) in isolation. For each infant, one of these syllables was designated as the control syllable; head-turns occurring while this syllable was being played were not reinforced. The other syllable was designated as the target syllable; head-turns occurring while this syllable was being played were reinforced. Assignment of [de] and [ti] as control and target syllables were balanced so that half of the infants in each subcondition received each syllable as the target. Throughout the shaping phase, as a background, the

<table>
<thead>
<tr>
<th>Condition</th>
<th>Subcondition</th>
<th>Contrast-Initial Strings</th>
<th>Contrast-Final Strings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invariant trochaic (rhythmic &amp; distributional cues to grouping)</td>
<td>1</td>
<td>dekog</td>
<td>nkog</td>
</tr>
<tr>
<td>Variable trochaic (rhythmic cue to grouping)</td>
<td>2</td>
<td>dekog</td>
<td>ngog</td>
</tr>
<tr>
<td>Variable mixed (no cues to grouping)</td>
<td>1</td>
<td>dekog</td>
<td>nkog</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>dekog</td>
<td>ngog</td>
</tr>
</tbody>
</table>

Table 1: Trisyllabic Strings Included in Stimulus Conditions
control syllable was continuously repeated at a comfortable listening level (60–62 dB (B) SPL), with 500-ms interstimulus intervals. Periodically, upon the experimenter’s signal, this background was interrupted with a five-token-long trial, allowing a response window of 4.625 s. The background-trial-background sequence of the shaping phase is illustrated in Table 2 for one of the stimulus subconditions.

In the shaping phase, infants were taught to turn their heads toward the loudspeaker when a change from the control syllable to the target syllable occurred. During this phase, all trials were target (change) trials. Initially, the target stimulus was presented at a level 12 db greater than the control stimulus to elicit orienting head-turns. The intensity difference between target and control stimuli gradually decreased as the infant responded correctly to target trials, until both stimuli were presented at equal intensity levels.

At this point, the criterion phase began. In this phase, the computer randomly selected trials to be either target trials or control (no change) trials. The background-trial-background subconditions for target and control trials in the criterion phase are illustrated in Table 2 for one of the stimulus subconditions. Infants were tested until they had either attained the predetermined criterion of responding correctly on seven of eight consecutive trials (by turning on target trials and not turning on control trials) or completed 30 trials. Infants not attaining the criterion within the specified number of trials were excluded from further participation; infants who attained the criterion returned for two additional testing sessions.

The two test sessions each began with a nine-trial review recapitulating the shaping phase—the control and target syllables were presented in isolation, and the intensity of the target syllable was gradually lowered until it was equal to the control syllable intensity. Following this, the multisyllabic testing phase began. In this phase, infants heard the same target/control contrast on which they had been trained, with syllables presented in trisyllabic strings. These trisyl labic strings varied across three between-subject stimulus conditions, as described previously.

In the multisyllabic testing phase, as the background, subjects heard continuous repetitions of control strings presented in blocks of three, with the order of the control syllable and the context syllables fixed within each block. Strings were separated by 1,000-ms silences. Across blocks, the order of control and context syllables changed randomly, with the constraint that no more than two consecutive blocks could be ordered identically. Upon observing that the infant’s attention was focused at the midpoint, the experimenter could request a trial. At the earliest interstring interval, the background was interrupted, and a trial was initiated. Each test trial consisted of one block of three consecutive identical trisyllables, allowing a response window of 7.125 s. Following the trial, the background was resumed, advancing to the next block.

Computer software arranged the selection of control and target trials in a pseudorandom order in sets of four. Each set included one target trial with the target syllable in initial position and one with it in final position; similarly, there was one control trial with the control syllable in initial position and one with it in final position. Altogether, each test session included 10 target trials and 10 control trials; in half of the trials, the contrast syllable occurred in initial position, and in half, it occurred in final position. Some possible background-trial-background sequences for target and control trials in the multisyllabic testing phase are illustrated in Table 2 for one of the stimulus subconditions.

Results and Discussion

Preliminary between-group analyses were conducted on several variables measured during the training session and the review phase of the first test session. Before the introduction of the multisyllabic contexts, all subjects had received equivalent treatments; therefore, no significant differences among the three groups were observed.

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Examples of Trial in Experiment 1&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stimulus Block</strong></td>
<td><strong>Background</strong></td>
</tr>
<tr>
<td><strong>Experiment Phase</strong></td>
<td><strong>Trial Type</strong></td>
</tr>
<tr>
<td><strong>Shaping</strong></td>
<td>All</td>
</tr>
<tr>
<td><strong>Criterion</strong></td>
<td>Target Control</td>
</tr>
<tr>
<td><strong>Target Initial</strong></td>
<td>dek&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; dek&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; t&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;3&lt;/sup&gt; dek&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;3&lt;/sup&gt; kogade&lt;sup&gt;3&lt;/sup&gt; dek&lt;sup&gt;2&lt;/sup&gt; ...</td>
</tr>
<tr>
<td><strong>Target Final</strong></td>
<td>dek&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;3&lt;/sup&gt; kogade&lt;sup&gt;3&lt;/sup&gt; dek&lt;sup&gt;1&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; dek&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; dek&lt;sup&gt;1&lt;/sup&gt; ...</td>
</tr>
<tr>
<td><strong>Control Initial</strong></td>
<td>dek&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; dek&lt;sup&gt;1&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; dek&lt;sup&gt;1&lt;/sup&gt; dek&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; dek&lt;sup&gt;2&lt;/sup&gt; ...</td>
</tr>
<tr>
<td><strong>Control Final</strong></td>
<td>dek&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; dek&lt;sup&gt;1&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; dek&lt;sup&gt;2&lt;/sup&gt; dek&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; kogade&lt;sup&gt;2&lt;/sup&gt; dek&lt;sup&gt;2&lt;/sup&gt; ...</td>
</tr>
</tbody>
</table>

<sup>a</sup>Possible sequences are given here only for one of four subconditions in the invariant trochaic condition, in which [t] served as the target stimulus, [de] served as the background stimulus, and [kogade] served as the context.

<sup>b</sup>Superscripts denote repetitions. [de<sup>4</sup>] denotes one or more repetitions of [de].
expected in these analyses. These expectations were borne out: Groups did not differ in total number of trials to completion of the training phase, number of trials to criterion, total percent correct in the criterion phase, or percent correct in the review phase of the first test session; in all instances, \( F(2, 33) < 1.00, \) n.s.

Overall percent correct discriminative head-turning was tabulated for each infant in each test session across all 20 trials. Mean group performance is shown in Figure 1. Percent correct was also tabulated separately for initial and final position contrasts.

An omnibus condition by session by position mixed ANOVA was conducted on these data. This analysis revealed a significant main effect of condition, \( F(2, 33) = 22.16, p < .001, \) and a significant condition by session interaction, \( F(2, 33) = 5.80, p < .01. \) No terms involving position attained significance, and this factor was therefore not considered in subsequent analyses.

Given the results of Goodsell et al. (1993), it was expected that performance in the invariant trochaic condition would improve across sessions as infants learned about the distributional properties of the stimuli but would remain the same in the variable mixed condition. These expectations were fulfilled: The invariant trochaic condition improved significantly across sessions, \( F(1, 11) = 9.66, p < .01, \) whereas the variable mixed condition did not change, \( F(1, 11) < 1.00, \) n.s. It is noteworthy that performance of subjects in the invariant trochaic condition in this study was 10 to 15% higher than that of subjects in the invariant condition of Goodsell et al., Study 2, who heard the same stimuli without rhythmic cues.

Initial expectations for the variable trochaic condition, which received uncorrelated cues from rhythm and distribution, were that performance would not differ from that of the variable mixed condition. The results proved otherwise, however. A Tukey test on scores from the first session \( (\text{CRT} = 13.61, \alpha = .01) \) revealed, contrary to expectations, that the variable mixed condition \( (M = 55.0) \) differed significantly from both the invariant trochaic \( (M = 70.0) \) and variable trochaic \( (M = 68.8) \) conditions; these last two conditions did not differ. A Tukey test on scores from the second session \( (\text{CRT} = 14.81, \alpha = .01) \) showed that the invariant trochaic condition \( (M = 84.2) \) differed significantly from both the variable mixed \( (M = 54.6) \) and the variable trochaic \( (M = 62.9) \) conditions, whereas these latter two conditions did not differ from one another. Across sessions, overall performance in the variable trochaic condition showed an insignificant decline, \( F(1, 11) = 2.91, p > .10. \) Closer inspection of the data, however, hinted at a bimodal pattern: 4 subjects showed large drops across sessions, whereas 4 other subjects showed moderate gains across sessions. Interestingly, these latter 4 subjects were among the 5 youngest subjects included in the condition.

Informal comparison of the present results with those from Goodsell et al. (1993), Study 2, suggests that rhythmic cues may be quite powerful in facilitating infants' grouping of syllables into wordlike units. This suggestion is affirmed by performance of the variable trochaic condition in the first test session, which, despite the lack of distributional support in input for grouping the context syllables, was equivalent to performance in the invariant trochaic condition. A variety of studies indicate that infants have clear sensitivities to prosodic properties of speech stimuli including rhythm (Demany, MacKenzie, & Vurpillot, 1977; Fernald & Kuhl, 1987; Fowler, Smith, & Tassinary, 1986; Mehler et al., 1988). The present results suggest that infants may use rhythm (or perhaps alternative prosodic properties as well) as an initial means of selecting candidates for segmentation. More generally, these results
are consistent with the possibility that infants employ a **bracketing strategy** to segregate portions of the speech stream based on cues to their endpoints, as has been hypothesized by Gleitman and Wanner (1982), Peters (1985), and Morgan (1986).

Beyond this, the present results—in particular, the difference between the invariant trochaic and variable trochaic conditions in the second session—confirm and extend Goodsit et al.'s (1993) finding that infants can use distributional information to form wordlike percepts. Goodsit et al. argued that infants may deploy a **clustering strategy** to aggregate basic elements of perception (syllables or phones) into larger units. A clustering strategy entails observation of the transitional probabilities between basic elements across contexts, information that can only be gleaned from distributional analysis. The present results suggest that infants coordinate these two strategies: As distributional patterns accrue across time, infants may discard those initially bracketed candidates for which corroborating clustering evidence is not found. Younger infants may need additional time to amass and analyze distributional evidence before discarding their early, rhythmically based hypotheses.

**EXPERIMENT 2**

In this experiment, the noise detection technique was used to investigate effects of rhythmic and distributional cues on infants' segmentation of simple strings of syllables. Three conditions comparable to those in Experiment 1 were included. The key prediction of this study was that infants hearing stimulus strings in which rhythmic and distributional properties provide correlated cues to grouping should respond more slowly to noises located between syllables whose grouping is cued than to noises located between syllables whose grouping is not cued. Response latencies of infants hearing stimulus strings lacking such correlated cues should not differ as a function of noise location.

**Method**

**Subjects**

Infants approximately 8 months of age were recruited from birth announcements in local newspapers in the Providence metropolitan area. Seventy-four subjects were tested to attain the final sample of 48. Sixteen subjects were assigned at random to each of three experimental conditions. Subjects were excluded from the study for the following reasons: excessive fussiness or crying (1), failure to meet the predetermined training criterion in the first session (15), interference from uncooperative siblings (3), and difficulties scheduling subsequent testing sessions within a 3-week period (7). The mean age of the infants completing the study was 243 days (range = 226–259 days). The two sessions required for each infant were scheduled so that the time between the two sessions was no more than 10 days.

**Stimuli and Apparatus**

The stimulus syllables used in this experiment were the same as those used in Experiment 1, as were the stimulus conditions. Trisyllabic strings included in each of the stimulus conditions are illustrated in Table 1. Infants were tested in a soundproofed laboratory room; the control room was next door. Trial duration, stimulus presentation, and delivery of reinforcement were controlled by custom-designed software. Stimuli were presented through a 5000-Hz TTE 411APS low-pass filter, a custom programmable attenuator, an Onkyo P-304 preamplifier, and an Onkyo M-504 power amplifier that was connected to an Electrovoice Sentry 100A monitor loudspeaker located in the testing room with the infant. Two smoked Plexiglas boxes containing motorized animal figures and located above the loudspeaker provided reinforcement.

**Procedures**

The experiment consisted of three main phases: **shaping, criterion,** and **testing.** The shaping and criterion phases were completed in an initial training session; the testing phase was completed in a subsequent test session. The physical setup was the same as in Experiment 1, except that the loudspeaker was located 90° from midline to the infant's left. The infant's behavior was monitored by the experimenter in the control room via a closed-circuit television camera and video monitor. As before, the experimenter served as sole judge of whether the infant turned his or her head during trials, depressing a button to signal to the computer that a head-turn had occurred. Through both sessions, all adults listened to music over around-the-ear headphones.

Infants were assigned at random to one of three stimulus conditions. These corresponded to the conditions used in Experiment 1. In this study, because infants were trained to respond to an extraneous noise superimposed upon the speech stimulus rather than to a change in the speech stimulus itself, the notion of **contrast syllable** was irrelevant. In the following, therefore, the term **buffer syllable** will be used instead to distinguish [de] and [t] from the context syllables [ki], [ko], [ga], and [gå].

Infants heard the trisyllabic strings throughout the experiment and heard strings with both buffer syllables equally often. The trisyllabic strings were presented at a comfortable listening level (60–62 dB SPL), with 50 ms between syllables within strings, and 1,000 ms between strings. The software randomly chose which of the four stimulus strings to present in each input block; three repetitions of the string occurred in each block. Upon judging that the infant's attention was focused at midline, the experimenter could request a trial. During the shaping phase, in each trial a **buzz** occurred in the interstring interval. Infants were allowed a window of 2.5 s from the onset of the buzz.
<table>
<thead>
<tr>
<th>Experiment Phase</th>
<th>Trial Type</th>
<th>Background</th>
<th>Trial</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shaping</td>
<td>All</td>
<td>kógáde³ kógar³ tikógar³</td>
<td>kógáde ↓ kógáde kógáde</td>
<td>kógári³ dekógar³ tikógar³</td>
</tr>
<tr>
<td>Criterion</td>
<td>Target</td>
<td>dekógar³ dekógar³ tikógar³</td>
<td>tikógar ↓ tikógar tikógar</td>
<td>kógar³ dekógar³ kógáde³</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>kógar³ tikógar³ dekógar³</td>
<td>dekógar dekógar dekógar</td>
<td>kógáde³ tikógar³ kógar³</td>
</tr>
<tr>
<td>Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context-Internal</td>
<td>First Position</td>
<td>tikógar³ kógáde³ dekógar³</td>
<td>kó ↓ gáti kógar³ kógar³</td>
<td>dekógar³ kógáde³ tikógar³</td>
</tr>
<tr>
<td></td>
<td>Second Position</td>
<td>kógar³ dekógar³ tikógar³ dekógar³</td>
<td>tikógar ↓ gá dekógar dekógar</td>
<td>kógáde³ kógar³ tikógar³</td>
</tr>
<tr>
<td>Context-External</td>
<td>First Position</td>
<td>tikógar³ dekógar³ kógáde³</td>
<td>ti ↓ kógar tikógar tikógar</td>
<td>dekógar³ dekógar³ kógar³</td>
</tr>
<tr>
<td></td>
<td>Second Position</td>
<td>kógáde³ tikógar³ kógar³</td>
<td>kógar ↓ de kógáde kógáde</td>
<td>kógar³ dekógar³ tikógar³</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>dekógar³ kógáde³ tikógar³</td>
<td>kógar³ kógar³ kógar³</td>
<td>tikógar³ kógáde³ dekógar³</td>
</tr>
</tbody>
</table>

a Possible sequences are given here only for one of two subconditions in the invariant trachae condition, in which (kógar) served as the context.

b Superscripted numbers indicate repetitions. ↓ indicates the location of the buzz within a trial.
in which to turn toward the loudspeaker to receive reinforcements; head-turns occurring after this period were not reinforced. Initially, the buzz was 100 ms in duration and 12 db louder than the speech stimuli; across shaping trials, the buzz was gradually shortened to 20 ms, and its intensity was lowered until it was equal to that of the speech. A possible background-trial-background sequence for one of the stimulus subconditions in the shaping phase is shown in Table 3.

At this point, the criterion phase began. In this phase, the computer randomly selected trials to be either target (buzz) trials or control (no buzz) trials. Possible background-trial-background sequences for target and control trials in the criterion phase are illustrated in Table 3. Infants were tested until they had either responded correctly on seven of eight consecutive trials (by turning on target trials and not turning on control trials) or completed 30 trials. Infants not attaining the criterion within the specified number of trials were excluded from further participation; infants who attained the criterion returned for a test session.

The test session began with a nine-trial review recapitulating the shaping phase—the buzz was presented in the intersyllabic interval, and the duration and amplitude of the buzz were gradually reduced. Following this, the testing phase began. In this phase, infants heard the same strings in the same fashion as before. However, during test trials, the buzz occurred in an intersyllabic interval within a string, either between the two context syllables ("context-intern-
al") or between a context syllable and a buffer syllable ("context-external"). As before, infants were allowed a period of 2.5 s beginning with the onset of the buzz in which to turn to receive reinforcement.

The testing phase included 24 trials total—8 control trials and 16 target trials. Each of the four strings was presented in four target trials. In two of these, the buzz occurred in the first intersyllabic interval; in the other two, the buzz occurred in the second intersyllabic interval. In half of the target trials, the buzz occurred in context-intern al position; in the remaining trials, it occurred in context-external position. Possible background-trial-background sequences for each trial type in the testing phase are illustrated in Table 3.

The final four subjects in each condition were videotaped using a Panasonic AG5040 time-lapse recorder interfaced to the computer. The computer initialized a timing signal on the recorder at the onset of the buzz in each target trial. Tapes were later reviewed frame-by-frame (60 frames per second) to ascertain the precise point at which the infant had fully turned toward the loudspeaker; the value of the timing signal was read off this point on the tape. The times thus obtained averaged almost 700 ms less than those obtained via the experimenter’s button press (because the time required for the experimenter’s judgment and press was eliminated), but for all subjects these times were highly correlated with the on-line RTs (range = 0.83-0.91).

Results and Discussion

As in Experiment 1, preliminary between-group analyses were conducted on variables measured during the training session and the review phase of the test session, including total number of trials to completion of the training phase, number of trials to criterion, total percent correct in the criterion phase, mean RT to target trials in the criterion phase, and percent correct in the review phase of the test session. No significant differences among the three groups were expected in these analyses. In all instances, these expectations were borne out.

Percent correct head-turning was tabulated for each infant across all 24 test trials (grand $M = 12.6$, $SD = 9.2$) and analyzed in a one-way between-group ANOVA. No differences were observed. Percent correct head-turning was also tabulated separately for the 16 target trials, broken down into context-intern al and context-external trials and analyzed in a condition by trial type mixed ANOVA. Again, no significant effects were observed. Thus, infants in all three conditions showed equal facility in detecting the buzzes; organization of the stimuli did not affect accuracy of responding.

For each subject, mean head-turning latency was computed for each of the four trial type (context-intern al or context-external) by position (first or second intersyllabic interval) combinations. All but two subjects had at least two head-turns in each combination (maximum four). To minimize the influence of extreme RTs, a root mean square procedure was employed in calculating the average. Mean group performance, collapsed across position, is shown in Figure 2.

An omnibus condition by trial type by position mixed ANOVA was conducted on these data. This analysis revealed significant main effects of condition, $F(2, 45) = 5.74, p < .01,$
and position (RTs to second-position buzzes were shorter than those to first-position buzzes), $F(1, 45) = 4.15, p < .05$. No interaction terms involving position attained significance, and this factor was therefore not considered in subsequent analyses. Most importantly, as predicted, this analysis also revealed a significant condition by trial type interaction, $F(2, 45) = 4.69, p < .05$. As Figure 4 shows, subjects in the invariant trochaic condition displayed different response latencies depending upon location of the buzzes with respect to potential units, whereas subjects in the two variable conditions showed no such differences.

Tukey tests were conducted to compare the three conditions on scores from context-external and context-internal trials. As expected, no difference was observed among the groups on context-external trials. On context-internal trials (CRt = 60.1, $\alpha = .05$; CRt = 76.7, $\alpha = .01$), however, subjects in the invariant trochaic condition ($M = 1294.7$) showed significantly longer latencies than did subjects in either the variable trochaic ($M = 1223.5$) or the variable mixed ($M = 1193.4$) conditions; the two variable conditions did not differ from one another. As in Experiment 1, however, closer inspection of the data from the variable trochaic condition again hinted at a bimodal pattern. Whereas most subjects in this condition had RTs on context-internal and context-external trials that differed little, 5 subjects had mean RTs on context-internal trials that were at least 75 ms longer than their RTs on context-external trials. These latter 5 subjects were among the 8 youngest subjects included in the condition, again suggesting that older and younger infants may differ in their abilities to process and integrate prosodic and distributional cues to grouping.

Although the discrimination maintenance technique had been used in previous studies of infant speech perception, the noise detection technique debuted here. Although this technique proved to be adequately sensitive for present purposes, some ways of enhancing its sensitivity are readily apparent. The means used to measure latency—invoking a button press by an experimenter making on-line judgments when head-turns occurred—were quite crude. Fine-grained coding of videotapes from several sessions showed, not surprisingly, that including the experimenter’s RTs added substantial variability to the measures of latency (although there was no evidence of bias from this source). Moreover, infants’ head-turning itself contributed variability to the data. Some infants move more quickly than others. Also, from trial to trial, because some delay between the request for a trial and the beginning of a trial was unavoidable, infants’ orientation at the onset of the buzz with respect to the loudspeaker varied: On trials where infants were oriented away from the loudspeaker, head-turns took longer to complete. The head-turns observed here were smooth ballistic movements; measuring latency to the initiation of movement would eliminate these latter two sources of variability.

GENERAL DISCUSSION

Provided that the discrimination maintenance and noise detection techniques both provide valid, sensitive measures of the organization of infants’ representations of strings of speech sounds, it was predicted that a complex pattern of results would be obtained. First, in the discrimination maintenance technique, it was predicted that subjects should perform more accurately under stimulus conditions in which particular groupings of speech sounds were reliably cued. In Experiment 1, subjects in the invariant trochaic group, in which these conditions were met, performed significantly better than subjects in either of the variable groups, in which these conditions were not met. Second, in the noise detection technique, it was predicted that subjects should show different response latencies as a function of noise location only under stimulus conditions in which particular groupings of speech sounds were reliably cued. In Experiment 2, the predicted condition by trial type interaction was obtained: Subjects in the invariant trochaic condition showed different latencies across trial types, whereas subjects in the variable conditions did not. These results provide evidence validating these techniques as means of measuring early speech segmentation.

Note that the convergence of results from these two techniques is critical for ruling out potential alternative explanations. For example, it might be possible to argue that the stimuli differed in inherent variability across the conditions, with stimuli in the invariant trochaic condition being the least variable. Due to this, stimuli in this condition might be easier to encode, leading to the observed advantage in discrimination maintenance. This explanation


Infant Speech Segmentation

of bias from head-turning data. Some other theories, such as the theory of contingency, do not attribute any particular representational organization to the infant. However, this explanation fails to predict the pattern of within-subject differences observed under the noise detection technique. On the other hand, this pattern of within-subject differences could conceivably be attributed to systematic variation in stimulus-masking properties, again avoiding attribution of specific representational organizations to the infant. However, such variations would not lead to the observed between-subject differences under the discrimination maintenance technique. Therefore, the most parsimonious explanation of the complete pattern of results obtained is that rhythmic and distributional properties of speech may jointly influence the organization of infants’ auditory perceptual representations.

These findings comport well with those arising from applications of the pause-insertion technique used by Hirsh-Pasek et al. (1987) and Jusczyk et al. (1992). Those articles reported that infants’ preferences were associated with prosodic properties of speech stimuli, including speech rhythm. Given the considerable training required to shape infants’ head-turning responses and the repeated exposures required to diagnose accurately infants’ representations, the results of the present studies do not speak to the issue of whether infants use properties like rhythm to form representational units in on-line processing of running speech. However, the present results clearly indicate that infants are capable of forming units based on rhythmic properties of speech under certain conditions of exposure, lending additional plausibility to the claim that infants use such prosodic properties in more natural conditions as well.

In spontaneous speech, prosodic cues of a variety of sorts occur in combination with one another. Quasi-synthetic stimuli of the sort employed here offer the opportunity to pull these constellations of cues apart—presenting individual cues in isolation, setting cues in conflict with one another, or presenting patterns of cues which children do not encounter in their native languages. In future research, manipulations of these sorts may contribute information relevant for discerning how perceptual abilities recruited for segmentation develop and what the role of linguistic experience in such development might be.

The primary goal of the present studies was to demonstrate that the two techniques used can yield converging results on representational units in infant speech segmentation. Therefore, the prosodic characteristics of the stimuli were selected to facilitate grouping of syllables into higher-order units. The particular rhythmic pattern used here was based on a pattern very commonly found in English. Cutler (1990) noted that English polysyllabic words most often have an initial trochee—a strong syllable followed by some number of weaker syllables (duration is one factor contributing to syllable strength). Based in part on this observation, Cutler has proposed that English speakers may use a “metrical segmentation strategy,” identifying strong syllables with beginnings of words and agglomering following weak syllables onto them. Jusczyk, Cutler, and Redanz (1993) have shown that infants prefer to listen to isolated words exemplifying the strong–weak pattern, a finding consistent with the metrical segmentation strategy hypothesis, as are the results of the present studies. It remains to be seen, however, whether stimuli exemplifying an iambic (short–long) rhythmic pattern would yield results similar to those obtained here.

Such a metrical strategy can only be the beginning of segmentation. Whether a sequence of sounds is actually a word depends crucially on its distributional properties: A sequence of sounds can constitute a word only if it occurs in the same form (or at least in a predictable form) in a variety of contexts. Many researchers, beginning with Snow (1972) have observed that adults often repeat themselves while speaking to young children, using the same words in different contexts. Several (e.g., Cooper & Aslin, 1989; Peters, 1985) have suggested that such variation may assist infants in identifying words. This hypothesis, however, rests on the assumption that infants can detect distributional properties of the input speech stream. The results of the present studies corroborate those of Goodsub et al. (1993) in showing that, at least by 8 months of age, infants can detect and exploit distributional patterns in input speech.

A spate of recent results shows that infants’ abilities for using various cues that may guide speech segmentation vary with age. Jusczyk et al. (1992) found that 9-month-old infants prefer to listen to uninterrupted phrases, whereas 6-month-old infants do not. In previous work, Jusczyk (1989) reported that 4-month-old
infants prefer to listen to uninterrupted foreign language clauses, whereas 6-month-olds do not. Jusczyk et al. (1992) compared these results to those of Werker and Tees (1984), who found diminishing capacities to discriminate foreign language segmental contrasts at 9 to 10 months of age, and suggested that speech perception capacities may be reorganized in the second half of the first year to direct attention more closely to features of the native language.

Results from the variable trochaic conditions in the present studies also hint at age-related changes: Younger infants, but not older infants, in these conditions performed in a manner comparable to infants in the invariant trochaic conditions. This trend suggests that younger infants may rely more heavily on prosodic cues alone, whereas older infants may rely more heavily on the combination of prosodic and distributional cues. Such a finding would be compatible with Jusczyk et al.'s (1992) explanation. Perhaps younger infants simply need greater exposure to distributional evidence than do older infants, to compensate for their less-focused attentional capacities. Another explanation, involving more central capacities, is also possible: Younger infants may lack the ability to integrate information from different sources in segmenting speech. On this view, emergence of infants’ ability to recognize words and related linguistic units might result from a newly won capacity to integrate prosodic and segmental information in the speech stream. Clearly, investigating age-related changes in infant’s speech segmentation abilities is a fertile area for additional research.

By providing means for triangulating on the organization of infants’ representations of speech, the techniques introduced here remove a major impediment to further study of infant speech segmentation. Many important questions concerning when and how linguistically relevant units begin to be represented remain to be addressed. The infant’s representation of such units forms a crucial part of the foundation of language acquisition; finding answers to these questions is therefore fundamental to our understanding of how language acquisition begins.

REFERENCES


Infant Speech Segmentation


22 May 1993; Revised 27 August 1993