COGNITIVE PSYCHOLOGY 24, 252–293 (1992)

Perception of Acoustic Correlates of Major Phrasal Units by Young Infants

PETER W. JUSCZYK

State University of New York at Buffalo; and Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS & EHESS

KATHY HIRSH-PASEK

Temple University

DEBORAH G. KEMLER NELSON

Swarthmore College

LORI J. KENNEDY

University of Oregon

AMANDA WOODWARD

Swarthmore College

AND

JULIE PIWOZ

Haverford College

How might young learners parse speech into linguistically relevant units? Sensitivity to prosodic markers of these segments is one possibility. Seven experiments examined infants' sensitivity to acoustic correlates of phrasal units in English. The results suggest that: (a) 9 month olds, but not 6 month olds, are attuned to cues that differentially mark speech that is artificially segmented at linguistically COINCIDENT as opposed to NONCOINCIDENT boundaries (Experiments 1 and 2); (b) the pattern holds across both subject phrases and predicate phrases and across samples of both Child- and Adult-directed speech (Experiments 3, 4, and 7); and (c) both 9 month olds and adults show the sensitivity even when most phonetic information is removed by low-pass filtering (Experiments 5

This research was supported by NICHD Grant (HD-15795). We greatly appreciate the assistance of Eric Bylund, Tara DeLeon, Elizabeth Hohne, Ann Marie Jusczyk, Denise Mandel, Stacey Jacobs, Nan Koenig, Karen Neumer, Jan Charles-Luce, and Jim Sawusch at various stages of this work. We also thank Ardell and Lindy for their help in producing the speech samples that we used in the experiments.

0010-0285/92 $7.50

and 6). Acoustic analyses suggest that pitch changes and in some cases durational changes are potential cues that infants might be using to make their discriminations. These findings are discussed with respect to their implications for theories of language acquisition.

Hockett (1954) once described language as having a duality of patterning. On one level, languages employ distinctive, ordered sequences of sound; on the other level, they organize semantic units into meaningful strings. The principles by which each level is organized differ from language to language, as does the correspondence between the levels. A crucial task facing the language learner is determining the nature of the correspondence between sound and meaning in his or her language. Part of this problem involves learning how a particular sequence of sounds refers to a particular object, event, or action in the environment. A still more basic problem is how to segment the stream of sounds into pieces that correspond to units on the level of meaning, i.e., words, phrases, and clauses.

In the language acquisition literature, relatively little attention has been given to the issue of how the child arrives at the correct segmentation of speech (for exceptions see Gleitman, Gleitman, Landau, & Wanner, 1988; Gleitman & Wanner, 1982; Hirsh-Pasek, Kemler Nelson, Jusczyk, Wright Cassidy, Druss, & Kennedy, 1987; Morgan, 1986). In part, this is due to the fact that most language learning studies have focused on the child's production of words and sentences in the target language. At this point in the process, the child is assumed to have largely solved the problem of correctly isolating and recognizing words and phrases in the language. Analogously, most theories of language acquisition concentrate on the way in which the child arrives at the right set of grammatical units from a correctly segmented set of sound sequences. The problem of segmentation itself is not addressed. In fact, even such fundamentally different positions concerning language acquisition as the functionalist (e.g., Bates & MacWhinney, 1987) and the learnability (e.g., Pinker, 1984; Osherson, Stob, & Weinstein, 1984; Wexler & Culicover, 1980) approaches start with the assumption that the child is able to isolate segments of speech that correspond to important linguistic units such as the clause or the sentence. These different approaches then contrast at the next level with respect to the way in which they view the child as discovering a meaningful analysis and interpretation of such units.

In principle, there are a number of ways in which the units that are important for syntactic and semantic analyses could be indicated in the sound stream of speech. These potential markers include both segmental and suprasegmental features. With respect to segmental features, it has been noted that different allophones (variants of the same phoneme) are often severely constrained in terms of the position that they can occupy within a syllable (e.g., Church, 1987; Frazier, 1987). In English an aspirated [p] can appear only in the initial position of a syllable and never in the final position of a syllable. So, finding an aspirated [p] in a sequence of sounds can serve as a marker between two different syllables. Similarly, restrictions on the allowable sequences of phonemes that a language permits within a word or syllable (i.e., phonotactic constraints) can provide information about both syllable and word boundaries. Thus, although it is permissible to begin a word in Polish with two successive stop consonants (e.g., "kto" or "dba") such sequences never occur at the onset of a syllable in English. Therefore, the occurrence of such a sequence is indicative of a boundary between successive syllables or words, as in "blacktop" or "red bag."

Suprasegmental cues such as intonation groups, stress patterns, pausing, and durational differences are other potential markers of units in the speech stream. In general, these types of markers tend to correspond to units larger than words and syllables. Thus, Cruttenden (1986; see also Bolinger, 1978; Cooper & Paccia-Cooper, 1980; Selkirk, 1984) points out that there is a tendency for intonation groups to correspond to major syntactic units such as clauses or their important components. Moreover, it has been noted that clause boundaries in English are often marked by changes in fundamental frequency, increases in the duration of syllables preceding the boundary, and by pausing between successive clauses (e.g., Klatt, 1975; Luce & Charles-Luce, 1983; Nakatani & Dukes, 1977). Nevertheless, such acoustic changes that frequently coincide with important syntactic units in speech also occur for nonsyntactic reasons (Nespor & Vogel, 1986; Vaissiere, 1981). For example, lengthening may occur to distinguish a voiced stop consonant from a voiceless one (Klatt, 1976). In addition, intonation may convey stylistic and affective attributes of the talker (Fairbanks & Pronovost, 1939; Lieberman, 1961; Williams & Stevens, 1972). Consequently, although an important factor, syntax is not the sole determinant of the organization of suprasegmental information in a sentence (Beckman & Edwards, 1990; Grosjean & Gee, 1987).

The fact that the speech stream may contain some, if albeit imperfect, acoustic markers of important grammatical units raises questions about the extent to which such markers could provide the infant with a way to segment utterances into their essential components. We contemplate the possibility that these markers are among the cues that the language learner uses to arrive at a segmentation of an utterance that largely corresponds to important grammatical units. This is not to claim that acoustic markers are the sole means by which the language learner arrives at a segmentation of utterances into clauses, phrases, and other relevant units. Nor is it to say that attention to such things as intonation groups, prosodic phrases, and the like will yield a segmentation of speech that corresponds perfectly to units on the level of meaning. Rather, it is to suggest that attention to such acoustic markers may furnish the language learner with a rough categorization of the input that can help delimit the range of alternatives for speech segments that map onto the grammatical units in the target language. Indeed, in much the same way that early speech perception capacities serve to organize the sounds into rough categories that are refined later to correspond to phonemic categories in the native language (Eimas, Miller, & Jusczyk, 1987; Jusczyk, 1986, 1992), a segmentation of speech based on acoustic markers and related cues may provide the guiding cues for the discovery of segments of speech that correspond to grammatical units.

In this context, it is interesting to note that there are indications that speech directed to young children, "motherese," often exaggerates features such as intonational contour, pausing, and duration. Hence, Fernald (1984) has shown that motherese is characterized by much greater swings in pitch range than is speech directed to adults. Could it be the case that the kinds of exaggerated prosodic cues that one finds in motherese (Bernstein-Ratner, 1985; Broen, 1972; Fernald, Taeschner, Dunn, Papousek, Boysson-Bardies, & Fukui, 1989) provide the infant with the necessary information to segment speech into the units required for grammatical analysis? Of course, establishing that such prosodic changes are more salient or more reliable markers of important grammatical units in Child-directed, as opposed to Adult-directed, speech is not sufficient basis for claiming that they guide the infant's segmentation of speech. Rather, in addition to finding such markers, it is necessary to show that the infants are able to respond to their presence in the speech stream.

Previous research in our laboratories has demonstrated that infants do appear to be sensitive to acoustic correlates of one type of grammatical unit, namely, the clause. Hirsh-Pasek et al. (1987) examined whether infants between 6 and 10 months of age preferred to listen to child-directed speech samples that were artificially disrupted by pauses either at clause boundaries or at within-clause locations. Artificial pauses inserted at clause boundaries presumably coincided with changes in other acoustic cues such as fundamental frequency and duration, whereas artificial pauses inserted at a variety of within-clause locations presumably did not correlate with systematic changes in these other types of cues. Hirsh-Pasek et al. found that the infants listened significantly longer to the samples segmented at clause boundaries than they did to the samples segmented at within-clause locations. A follow up study by Kemler Nelson, Hirsh-Pasek, Jusczyk, and Wright Cassidy (1989) demonstrated that the exaggerated prosody of Child-directed speech may play a role in providing the infant with reliable cues to clausal units. In particular, the preference for samples segmented at clause boundaries was greater for

Child-directed speech than for Adult-directed speech. In fact, it did not show at all for Adult-directed speech in that study. Hence, there is some indication that Child-directed speech may be particularly well adapted to presenting information about important grammatical units.

From the point of view of language learning, an ability to isolate important grammatical units may indeed constitute a significant step forward in solving the induction problem of language acquisition. The problem that the child faces is to induce a grammar for a particular language on the basis of the limited set of input strings that the language learner hears. From this limited data base, the child must discover the complex rule set that is used to generate all and only the sentences of the target language. In order to account for the way that the child arrives at the correct induction on the basis of the input, it appears necessary to attribute some basic abilities to the child. The minimal set of these abilities includes some parsing of the environment into objects and events, segmenting the linguistic input into structures like clauses and phrases, and mapping pieces of the environment onto structures in the language (e.g., see Gleitman et al., 1988). Thus, among the component abilities necessary for the solution of the induction problem is the ability to determine the units in the language and the relations that adhere among them. In this sense, Hirsh-Pasek et al.'s finding is intriguing because it suggests that attention to the acoustic structure of speech can provide the infant with important clues for the discovery of structural units, such as clauses.

The next natural step is to ask about the degree to which attention to the sound structure of the language can provide the infant with information about finer units for linguistic analysis such as phrases. Phrases serve as the organizational constituents within grammar. Universally, sentences require at least a noun phrase and a verb phrase. The relations that exist among the various types of phrases define sentence structure, characterize families of sentences, and reveal dependencies that hold across different parts of the sentence. For example, in English, grammatical properties like verb agreement are described as dependencies between noun phrases and verb phrases ("He goes" vs "They go"). In addition, relations that are internal to phrases often provide important information about grammar. Thus, dependencies between word classes often hold within phrase boundaries. In English, phrase structure rules specify that determiners (i.e., "the" or "a") and adjectives precede nouns. In case marking languages, like Spanish, number and gender markers must agree for all items within a phrase (Morgan, Meier, & Newport, 1987).

Given the critical role that phrasal constituents play in the description of grammar, it follows that the ability to detect such units, to differentiate them, and to note their distributional properties would provide the child with an important head start in learning language. Indeed, Pinker (1984) has suggested that the ability to track the distribution of phrases and of constituents within phrases would take the child a long way toward constructing a grammar.

A prerequisite to noting the distributional properties of phrases and their constituents is an ability to detect segments in the speech stream that correspond to phrasal units. Logically, the detection of cues to such units depends both upon the existence of some type of marking of phrases in speech and on the ability of the learner to make use of such marking. With respect to the presence of cues, there are data to indicate that phrasal units may be marked in a number of different ways. First, prosodic correlates of phrase boundaries such as declinations in fundamental frequency and lengthening of vowels in phrase-final words have been noted (e.g., Beckman & Edwards, 1990; Cooper & Paccia-Cooper, 1980; Klatt, 1975, 1976; Lehiste, 1973; Luce & Charles-Luce, 1983; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Streeter, 1978). Moreover, Morgan (1986) has reported that such prosodic markers tend to be exaggerated in Child-directed speech (see also Fisher, 1991; Lederer & Kelly, 1991). Second, the presence of certain morphemes in the input may furnish cues to phrasal units. For instance, "the" generally appears at the beginning of noun phrases (Morgan, 1986). Consequently, there are some potential indicators of phrasal units in the speech stream.

Unfortunately, less is known about the ability of listeners to make use of potential markers of phrasal units. A suggestion that adults are attentive to markers of phrasal units comes from studies demonstrating that listeners are sensitive to speech cues that are correlated with phrasal units in unfamiliar foreign languages (Pilon, 1981; Wakefield, Doughtie, & Yom, 1974). Another indication comes from speech perception studies demonstrating that segmental and prosodic features can be used by listeners as perceptual cues to the location of a major syntactic boundary (Collier & t'Hart, 1975; Lehiste, Olive, & Streeter, 1976; Price et al., under review; Scott, 1982; Scott & Cutler, 1984; Streeter, 1978). Finally, some indication that language learning may be facilitated by the presence of cues like prosody, concord morphology, and the appearance of certain lexical items like function words come from studies of adults learning artificial grammars. Morgan (1986; Morgan et al., 1987) demonstrated that the existence of such cues in the input facilitated adults' acquisition of an artificial language.

Although the results of the studies just described suggest an ability to attend to markers of phrasal units, the fact is that they were obtained with adult subjects who already possessed a native language. Thus, it is difficult to know how much the ability to use such information depends on the existence of sophisticated information processing strategies or on the subjects' previous experience speaking a language. In the case of the

Morgan et al. results, there was some indication that subjects were capable of using cues of a type (concord morphology) that did not appear in their native language (English). On the other hand, Scott and Cutler (1984) found that listeners were unable to use certain segmental cues to syntactic boundaries unless they actually produced these cues in their own utterances. More generally, studies of adults establish the foundation for—but do not actually address—the question of whether the ability to detect speech cues to the internal structure of clauses is present in infants learning a first language.

The current research focused directly on this question: Early in the language acquisition process, are infants sensitive to cues that mark speech segments corresponding to major subclausal units, like subject–noun phrase and main verb phrase? In exploring this question, we also considered a number of related issues that concern the role of experience and the nature of the input that the listener receives.

EXPERIMENT 1

The first question that we addressed was whether infants display any sensitivity to the markers of major phrasal boundaries in English. As noted already, many important syntactic regularities are most easily described by treating the phrase as an important unit of organization. Moreover, some accounts of language acquisition have postulated explicitly that cues to phrasal structure in the linguistic input to the child are critical for inducing the underlying organization of sentences (e.g., Gleitman et al., 1988; Gleitman & Wanner, 1982; Morgan, 1986; Morgan et al., 1987).

In a language like English, it is reasonable to expect that subject–noun and predicate–verb phrases are units that would receive significant marking in the linguistic input. This is because sentence subjects in English function in a number of important ways. For example, they carry nominative case marking, they control agreement in person and number with the verb, they are usually identified as the agent of a transitive verb, and they function as the topic of the sentence (Givon, 1979). Predicate–verb phrases define the architecture of the sentence. The verb phrase, particularly the verb, is responsible for assigning thematic roles in the sentence. In fact, the learning of verbs and verb phrase structure is seen as pivotal to the induction of grammar in current theories of language acquisition (Gleitman, 1990; Pinker, 1989). Finally, there is evidence that, at least under some circumstances, intonation groups do align with the subject–predicate division in sentences (Beckman & Edwards, 1990; Cruttenden, 1986). Consequently, we chose to focus our investigation on the infant's detection of prosodic cues to subject and predicate phrases.

To evaluate the infant's sensitivity to phrasal units we used the same method that we have employed in our previous investigations of speech segmentation into clauses (Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989). Namely, we observed whether infants preferred to listen to speech that was artificially interrupted either at a major phrasal boundary or at another location within the sentence. For this purpose, matched pairs of samples were selected from the spontaneous speech of a mother to her 2-year-old child. These samples were then altered by artificially inserting 1-s pauses at various locations in the samples. One set of the samples had pauses inserted at locations corresponding to the boundary immediately preceding the predicate. The other set included an identical number of pauses but these were inserted at a different location in the sentence, namely, following the main verb (but before the end of the sentence). The rationale behind inserting the pauses was the following. To the extent that infants do perceive some inherent organization of sentences into perceptual subgroups, then an artificial segmentation that coincides with the natural subgroups should be preferable to one that disrupts them. Our previous findings indicated that, by 6 months of age, infants show just such a preference for Child-directed speech that is segmented at clause boundaries. This procedure of pause insertion has also been used successfully in studies examining adults' segmentation of speech samples in unfamiliar languages (Pilon, 1981; Wakefield et al., 1974).

Method

Subjects. The subjects were 24 7- to 10-month-old infants (mean age: 9.5 months; range: 7.3 to 10.4 months) from the suburbs of Philadelphia. All of the infants came from monolingual English-speaking homes. Twelve additional infants were tested but failed to complete the procedure because of crying or inattentiveness (9 Ss) or for failure to look to one of the two sides for at least three trials (3 Ss). The rationale for the latter criterion is described below.

Stimuli. The stimulus tapes were generated from a recording of a mother speaking to her 2-year-old child. The talker did not know about the purpose of the research. She was simply instructed to interact naturally with the child. During the course of the interaction, the mother played with the child and commented on the child's actions and the objects in the room. From this corpus, 16 excerpts were chosen, each one between 7 and 9 clauses in length.

The samples as originally spoken by the mother were digitized at the University of Oregon on a PDP 11/73 computer using a 12-bit A/D converter. An auditory editing program was used to remove all existing gaps in the speech with durations over 400 ms. (Attempts to remove pauses of shorter durations would have resulted in very unnatural sounding samples.) Removing a gap entailed reducing the silent interval to the minimal duration that still ensured that the phonemes on either side of the gap were unperturbed. Depending on the phonemic context, this duration ranged from about 10 to 100 ms. The aim in removing the gaps was to minimize systematic differences in the *durations and numbers* of pauses in the different versions of each sample (see Hirsh-Pasek et al., 1987, Experiment 2 for further discussion).

Two versions of each excerpt were created for use as stimulus materials. The samples all began and ended at a sentence boundary. One version of each sample, the "Coincident version" was modified by inserting 1-s pauses just before the main verb. In the majority of

TABLE 1
Examples of the Alternate Versions of a Spontaneous Speech Sample

Coincident version
What happened? Did you / spill your cereal? Oh well, how about cleaning it up? Did you / want to pick it all up and put it back in your container? Okay, that / looks great. You / want to put it back in your little container here? Would it help if I / move the chair?
Noncoincident version
What happened? Did you spill / your cereal? Oh well, how about cleaning it up? Do you want / to pick it all up and put it back in your container? Okay, that looks / great. You want / to put it back in your little container here? Would it help if I move / the chair?

instances, the pause came directly between the subject–noun and the verb. However, on 39% of the insertions, the pause was placed between the auxiliary and the verb. This was done for several reasons. First of all, there were a number of contractions with the noun and a following auxiliary; attempts to segment the speech in these locations would have resulted in transients in the acoustic signal. Second, the spontaneous samples contained many yes/no questions wherein the auxiliary was placed before the noun phrase. Third, there is still considerable disagreement in the linguistic literature about whether the auxiliary should be properly considered as an element of the verb phrase or as a separate phrase (e.g., see Chomsky, 1981, 1986). For these reasons, we decided that for the present materials, the best placement for the pause was just before the main verb. The other version of each sample, the "noncoincident version," was prepared by inserting a 1-s pause immediately after the main verb and before the end of the sentence.[1] Although in some cases the Noncoincident pauses occurred at a phrase boundary, namely, that of a prepositional phrase, such phrases are considered minor by comparison to the predicate phrase. In inserting pauses into the speech signal, we endeavored to find locations at zero crossings in amplitude so as to avoid obvious acoustic transients.

With the exception of the location of the pause, the Coincident and Noncoincident versions were equated in all respects. Hence, they contained the same number of pauses greater than 400 ms. The mean duration of the stimulus excerpts for the test trials was 22.2 s, with a range between 15.4 and 32.8 s. The number of inserted pauses ranged from 5 to 7. An example of a Coincident and a Noncoincident sample is shown in Table 1.

Two orders of the 16 samples were recorded on tape. Both began with both versions of the same four (preexposure) samples. The remaining 12 (test) samples were randomly ordered differently in the two cases.

Apparatus. The infant was seated on the mother's lap in the center of a three-sided testing booth, with panels 4 ft by 6 ft on three sides and open at the back. A red light and a loudspeaker were mounted at eye level on each of the side panels, 78° to the left and right

[1] If the main verb was followed by a sentence boundary, then that sentence was not artificially segmented in either the Coincident or Noncoincident version. Artificial segmentation was also withheld if the subject–noun was not expressed as in imperatives. In still a few more cases, opportunities for differential segmentation were passed up in order to avoid making the speech samples too choppy overall. Finally, the first clause in a sample was never artificially segmented.

of the infant when facing midline. A blue light, centered on the front panel, could also be flashed. An observer was concealed behind a curtain above the front panel. She signaled the timing and direction of the infant's head orientation to a tape recorder operator and a microcomputer in an adjoining room.

Design and procedure. As in our earlier studies, the procedure was a modified version of the one developed by Fernald (1985). Each infant completed an eight-trial preexposure phase (both versions of four pretrial stimuli) and a 12-trial test phase. Over all trials, the infant consistently heard the Coincident versions of the speech samples through the loudspeaker on one side of the booth and the Noncoincident versions through the loudspeaker on the other side. For half of the infants, the Coincident versions were assigned to the right side and the Noncoincident versions to the left; for the other half, the assignments were reversed. An eight-trial preexposure period was designed to acquaint the infants with the assigned position of each type of version.

Preferences were indexed by monitoring the duration of the infant's headturns to the two types of versions, Coincident and Noncoincident, over the set of 12 test trials. In Fernald's original use of the preference methodology, preferences were measured instead by the frequency of the infants' *direction* of headturn. However, given that only the durational measure revealed systematic effects in our earlier studies (Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989) we elected to focus exclusively on the duration measure in this and all other studies reported in this paper. To enhance the reliability of the durational measures, we analyzed the data only from those subjects who oriented on at least three trials to each type of version, Coincident and Noncoincident. Of course, a by-product of this practice is that the directional measure of preference becomes potentially less discriminating. Accordingly, we report only the results obtained with the durational measure.[2]

Each trial began by blinking the blue light in order to draw the infant's attention to center. When the observer signaled that the infant was oriented at midline, the center blue light was extinguished and the red light above one (preexposure trials) or both (test trials) of the acoustic speakers began to flash. These lights indicated that a speech sample was available on that side or sides, provided that the infant made a headturn of at least 30° in the direction of the corresponding speaker. When the observer detected such a headturn, she signaled the tape recorder operator in the adjoining room who began the speech excerpt appropriate to the direction of orientation. The speech excerpt was continued either to completion (with orientation time recorded as equal to the total duration of the sample) or until the observer indicated that the infant failed to maintain the 30° headturn for at least 2 consecutive s (e.g., if the infant turned back to center or the other side, looked at the mother, the floor, or the ceiling). In this event, timing of the orientation stopped and the sample was terminated at the next occurring inserted pause. A silent moving puppet at center entertained the infant in a short intertrial interval while the tapes were advanced to the next excerpts. Whether the first heard sample was Coincident or Noncoincident and whether it occurred on the right or the left were counterbalanced across infants.

In the preexposure phase, a speech sample was available on only one side per trial. Eight different excerpts were heard, the Coincident and Noncoincident versions of each of four samples. Matched versions were played successively.

On the test trials, matched excerpts were simultaneously available on the right and left. The infant chose which one of the two versions was played according to the direction of the headturn. A headturn not only started the excerpt appropriate to that side but also terminated the blinking light on the other side. Across the 12 test trials, no speech sample was

[2] In fact, in more recent work, we have further modified the procedure in a way that the directional measure is eliminated entirely. Instead of subjects choosing between versions available on either side during a trial the version is predetermined.

ever repeated. Infants were given a short break if they fussed or failed to maintain a headturn on two successive trials. Such prematurely aborted trials were also rerun and were not counted in the data analysis. No infant received more than one break before the experimental session was ended.

The observer, whose signals started trials and indicated the occurrence, direction, and termination of headturns, was blind to whether the Coincident samples were presented to the infant's right or left. Both she and the mother wore headphones over which loud music was played to prevent them from hearing the speech samples. In addition, the assignment of the Coincident samples to either the right or left side was done by the second experimenter (the tape operator) who also was responsible for setting the loudness level of the samples and equating their volume.

Results and Discussion

The data from the 12 test trials were used to calculate the mean length of orientation to each type of sample. The mean length of orientation across subjects for the Coincident samples was 19.9 s as opposed to 16.8 s for the Noncoincident ones. Separate analyses with subjects ($t(23) = 6.39$, $p < .001$) and samples ($t(11) = 2.66$, $p < .05$) as random factors confirmed that this difference is significant. Overall, 23 of the 24 subjects had longer mean fixation times for the Coincident versions. In addition, for 10 of the 12 sample pairs, the mean fixation time was longer for the Coincident versions.

Hence, the results indicate that infants listen longer to strings of English that are segmented at the major phrase boundary, just prior to the main verb of the predicate, as opposed to strings that are segmented within the predicate phrase itself. The pattern of results in the present study is strongly reminiscent of what we observed for speech samples interrupted at either clause boundary or within clause locations (Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989). Those results suggested that infants as young as 6 months are sensitive to acoustic markers of clausal units, preferring a segmentation that leaves the clauses intact. In the present case, the results suggest that the infants are also sensitive to markers of clause-internal units. Thus, the pauses in the Coincident versions of the present study apparently corresponded to a detectable degree with the underlying perceptual organization of clausal subunits. Given our previous findings that infants are sensitive to markings of speech segments corresponding to clausal units, and our current findings that they are also sensitive to markings related to subclausal units (i.e., predicates), the picture that begins to emerge is that infants are attuned to acoustic designates of units within the speech stream that relate to a hierarchy of grammatically relevant categories. To test the generality of the findings of the first study, as well as to obtain some information regarding the age at which sensitivity to subclausal units might arise, we conducted the following experiment.

EXPERIMENT 2

Research on acoustic correlates of clausal units indicates that both 6 and 9 month olds display sensitivity to such units (e.g., Hirsh-Pasek et al., 1987). In Experiment 1, we showed that 9 month olds also display some sensitivity to acoustic correlates of certain clausal subunits. Might this sensitivity also be evident in 6 month olds? Recent cross-linguistic evidence suggests that at least at the level of phonetic contrasts, there is some tendency between 6 and 12 months of age for infants' ability to distinguish foreign language contrasts to diminish while their skills with respect to native language contrasts are maintained (Werker & Lalonde, 1988; Werker & Tees, 1984; but see Best, McRoberts & Sithole, 1988). If this is an indication that infants become more closely attuned to the structure of the native language in this period, then perhaps sensitivity to clausal subunits also develops during this time. To explore this issue, as well as to attempt to replicate the results of the previous experiment, we tested both 6 and 9 month olds on the Coincident and Noncoincident versions of the phrase samples.

Method

Subjects. The subjects were 32 infants from English-speaking homes in the Eugene, Oregon area. The 16 6 month olds had an average age of 6.5 months (range: 6.0 months to 7.3 months), whereas the 16 9 month olds had an average age of 9.3 months (range: 8.8 months to 10.3 months). In order to obtain the 32 subjects, it was necessary to test 26 6 month olds and 22 9 month olds. Subjects were excluded for the following reasons: failure to look to one of the two sides for at least three test trials (eight 6 month olds; four 9 month olds), crying (two 6 month olds; one 9 month old), and experimenter error (one 9 month old).

Stimuli. The same stimuli were used as in Experiment 1, the only difference being that instead of prerecording the stimuli on audio-tape, all stimuli were digitized on a LSI 11/73 computer for on-line, real-time presentation to the infants. The computer was programmed to randomly select the test pairs for each trial. Hence, each subject had a different random ordering of the 12 test pairs.

Apparatus. A testing booth was built with dimensions comparable to those described in Experiment 1. The booth was constructed out of peg board, which made it possible for the observer to look through one of the existing holes to monitor the infant's headturns. For this purpose, a small section of the peg board in the center was free of cardboard backing. The remainder of the peg board was backed with white cardboard to guard against the possibility that the infant might respond to any movements behind the panel. A white curtain suspended around the top of the booth shielded the infant's view of the rest of the room. The only other changes were the substitution of a green light for the blue one in the center panel and the presence of a computer terminal and response box behind the screen and out of view of the infant. The response box, which was connected to the computer, was equipped with a series of buttons that started and stopped the flashing center and side lights, recorded the direction and duration of a headturn, and terminated a trial when the infant looked away for more than 2 s. Information about the direction and duration of headturns and the total trial duration was stored in a data file on the computer.

The audio output for the experiment was generated from the digitized waveforms of the samples stored on the computer in an adjacent laboratory room. A 12-bit D/A converter fed the output through anti-aliasing filters and a Kenwood audio amplifier (Model KA 5700) to the 7-inch loudspeakers mounted on the side walls of the test chamber.

Procedure. Only slight changes were made in the procedure. First, the use of the computer made it possible to run the procedure with only a single experimenter. The experimenter sat holding the response box and initiated trials only when the infant was facing forward. The experimenter was blind to whether the Coincident samples were on the left or the right. This was possible because the assignment of the Coincident samples to the left or right side was determined for each subject by the computer and was not revealed to the experimenter until after the completion of the test session. The loudness levels for the samples were set by a second assistant, who was not involved in the observations, at 72 ± 2 dB (C) SPL using a Quest (Model 215) sound level meter. As in the previous experiment, both the experimenter and the child's parent listened to recorded music over headphones to prevent them from hearing the samples.

A second change in the procedure was also a consequence of using the computer. To get a more accurate estimate of looking, the time that the infant actually spent orienting to the loudspeaker on a given trial was recorded. This was accomplished by having the experimenter press a button only when the infant was looking to one side. Thus, if the infant turned away briefly from the target by 30° in any direction, but for less than 2 s, and then looked back again, the time spent looking away was not included. Hence, measures of orientation time in this experiment are expected to be shorter than those reported in the previous one. As before, only if the infant turned away for more than 2 s was the trial terminated but unlike the previous procedure, the computer stopped the speech sample right at this point, rather than at the next inserted pause. In addition to the measure of orientation time, the total trial duration used in Experiment 1 was recorded to facilitate comparisons between the experiments.

Results and Discussion

Analyses of the data indicated that the same pattern of results was evident for both the orientation time and total trial duration measures. To conserve space and also because we believe that it is a better index of the infant's attention to the samples, we will present only the orientation time data. The mean duration orientation times for the Coincident and Noncoincident versions for each age group are shown in Table 2. The data for the 9-month-old subjects replicate the results of Experiment 1 in that the infants displayed significantly longer orientation times for the Coincident

TABLE 2
Mean Length Orientation Times for Spontaneous Speech Samples in Experiment 2

	Coincident	Noncoincident
	Cross-sectional comparison	
6-month olds	6.37 s	6.22 s
9-month olds	8.83 s	7.09 s
	Longitudinal comparison	
6-month olds	7.03 s	6.49 s
9-month olds	9.88 s	7.33 s

versions ($t(15) = 2.03, p < .05$, one-tailed). However, for the 6-month-old infants, the difference in orientation times between the Coincident and Noncoincident versions was not significant ($t(15) = 0.19$). These same tendencies were manifested in the analyses for the samples. Thus, for the 9 month olds there was a significant difference in favor of the Coincident versions ($t(11) = 4.82, p < .001$), whereas no difference was evident for the 6 month olds ($t(11) = 0.05$). Overall, 13 of the 16 9 month olds and 11 of the 16 6 month olds oriented longer to the Coincident versions. In addition, for all 12 of the sample pairs, the 9 month olds oriented longer to the Coincident versions, whereas the 6 month olds had longer orientation times to the Coincident versions for only 8 of the 12 sample pairs.

The initial data analysis suggested a difference in the way in which the 6 and 9 month olds reacted to the different versions of the samples. To determine whether this was the case, we conducted planned comparisons based on an ANOVA for a 2 (age) by 2 (sample version) mixed design. The expected interaction was only marginally significant ($F(1,30) = 2.88$, $p = .10$). Inspection of the data indicated that there was considerable variability in the younger age group, so we conducted an additional analysis using a nonparametric procedure, the Mann–Whitney U test. This analysis based on the rank-order of difference scores (mean looking time to Coincident samples minus mean looking time to the Noncoincident samples), indicated that the contrast between the two age groups was significant ($U = 76, p = .03$).

In an effort to obtain more information about the suspected developmental trend, we decided to conduct a longitudinal investigation by arranging to test our original 6-month-old group at 9 months of age. Twelve of the original subjects successfully completed a second test session at 9 months of age. Two of the original subjects were not available, and the other two infants looked only to one side during the second test session and thus supplied no useful data. These four subjects were replaced by testing four new subjects at both 6 and 9 months of age. With the new subjects added, the younger group averaged 6.5 months of age and the older group 9.6 months of age. The data were analyzed to determine whether there were any preferences manifest for the Coincident samples at either age. Not surprisingly, at 6 months, the infants gave no evidence of a significant preference ($t(15) = 0.80$). However, at 9 months, there was a significant preference for the Coincident versions ($t(15) = 2.91$, $p < .02$). The analyses with samples as a random factor showed the same tendency with the difference between the Coincident and Noncoincident versions significant at 9 months of age ($t(11) = 3.63, p < .01$), but not at 6 months ($t(11) = 1.16, p > .10$). Still, paired t tests used to evaluate changes in performance across the two age levels did not reveal statistically reliable results at a conventional level of significance ($t(15) = 1.57$,

$p < .10$ with subjects as a random factor and $t(15) = 1.56$, $p < .10$ with samples as a random factor). Consequently, the data are, once again, only suggestive of a developmental trend.

The results obtained thus far provide strong evidence that 9-month-old infants systematically distinguish between the Coincident and the Noncoincident versions. This pattern was obtained both in Experiment 1 and in two instances in the present experiment. In contrast, there is no evidence to this point of a similar preference on the part of 6-month-old infants. Yet, 6 month olds in our studies of clause-relevant sensitivity did show the preferences robustly. Thus, the findings begin to hint of an interesting developmental trend for the analysis of cues corresponding to the level of subclausal segmentation. However, given the weakness of statistical evidence for a developmental difference in the current study, we await further evidence before concluding that 6 month olds have no preference (or more correctly, a lesser preference) for utterances segmented at a phrase boundary than 9 month olds. Perhaps such a preference exists and would show itself under different circumstances. We will reconsider this issue later in light of the results of subsequent experiments.

Experiment 3 focuses on a different set of issues. The findings of Experiments 1 and 2 are limited by the fact that all the Coincident versions of speech were interrupted before the main predicate verb and all the Noncoincident versions were interrupted within the predicate–verb phrase. Is the preference for the intact phrase peculiar to the predicate or does it generalize as well to cues corresponding to the subject–noun phrase, the other major subclausal unit? Moreover, given the restriction on the earlier stimuli, one also must entertain an artifactual account of the preferences observed so far. Note that in the first two experiments, Coincident versions were always interrupted earlier in the string than Noncoincident versions, a logical necessity that follows from breaking the samples as we did, i.e., breaking the Noncoincident versions within the predicate phrase. Perhaps later pauses simply have a higher probability of terminating infants' attention. Both of these concerns can be met simultaneously by testing pairs of speech samples in which the Noncoincident version is interrupted within the subject–noun phrase. This is what we proceeded to do in Experiment 3.

EXPERIMENT 3

The natural speech materials used in the earlier experiments were not suitable for this study. As is generally true of natural motherese, our materials tended to contain many sentences with very short subject–noun phrases, often only a single word. Thus, artificial pauses could not be inserted within them. For Experiment 3, materials with long subject–noun phrases were specially created. Pairs in which the Noncoincident versions were interrupted in the subject–noun phrase were prepared as well as pairs, analogous to those used previously, in which the Noncoincident version was interrupted in the predicate verb phrase. If sensitivity to speech markers correlated with subclausal units is general across subject–noun phrases and predicate–verb phrases and if our previous results are not based in the temporal artifact, then we would expect to see the preference emerge for both kinds of pairs.

In fact, there is even some reason to suppose that the preferences would be stronger with the new materials. There are suggestions that intonational marking of subject phrases as distinct from predicate phrases is more common when the subject phrases are long (e.g., Cruttenden, 1986). In addition, there is some evidence from a study with 7 year olds that children are better able to use prosodic markers when subject phrases are long (Read & Schreiber, 1982). Consequently, if infants in the previous studies were responding to acoustic correlates of phrasal units, then one would expect that the effect of using utterances with long noun phrases would be, if anything, to enhance the preference for the Coincident versions.

Method

Subjects. The subjects were 24 7- to-10-month-old infants (mean age: 9.2 months; range: 7.5 to 10 months) from the Philadelphia suburbs. Five additional subjects were tested but failed to complete the procedure due to fussiness, inattentiveness, or failure to look to one of the two sides for at least three trials. All of the infants came from monolingual English-speaking homes.

Stimuli. In seeking materials for the present experiment, we recorded several additional talkers interacting with a young child under several different conditions. In no case were we able to obtain a sufficient number of sequences that included the kinds of long noun phrase sequences that we sought. For this reason, we decided to create our own materials using a series of drawings selected from children's books and an accompanying text, with content appropriate for a young child, which we wrote. Associated with each picture was a series of sentences that described events and characters portrayed in the pictures. All sentences were declarative and had subject–noun phrases of four words or more. In all, there were 16 different scenes and accompanying texts. A woman was brought into the laboratory with a 2-year-old child. She was given the Storybook materials to study and practice. When she was ready, we recorded her while she read the story to the child in an animated way. She was not informed of what the recording was to be used for.

The tape recording was used to prepare the stimulus materials for the experiment. Each of the 16 sequences was digitized and stored on a PDP 11/73 computer using a 12-bit A/D converter. Once again, all pauses longer than 400 ms were removed from each sequence by use of an auditory editing program. Two versions were prepared for each sequence by inserting 1-s pauses into the utterances at different locations. Once again pauses were inserted only at zero crossings in amplitudes so as not to produce transients in the signal. For the Coincident versions, the pauses were inserted just after the subject–noun phrase and before any auxiliaries. This represented a slight change from the previous experiments wherein the pause occurred just before the main verb. The use of the prepared "storybook" materials permitted us to avoid some of the problems (e.g., high occurrence of yes–no questions, contractions involving the noun, and a following auxiliary, etc.) that we would have had with a placement just after the subject phrase. The new placement allowed us to more consistently isolate the subject–noun and predicate–verb phrases in the Coincident versions. The Noncoincident versions were prepared by inserting the pause in other locations in the sentence. For half of the samples, the pause came just after the main verb (as in the previous experiments); for the other half, the pause occurred at a location within the subject–noun phrase. It followed that for half of the pairings the Noncoincident sample included a pause that occurred earlier, and for the other half later, than that for the Coincident version. This permitted us to separate out any possible preferences for early vs late placement of pauses from any preferences that might be associated with the phrasal units. In every way, except for the placement of the pauses, the Coincident and Noncoincident versions of each sample were identical. For the samples presented on the test trials, the mean duration was 26.07 s with a range between 21.35 and 33.25 s. Examples of the Coincident and Noncoincident versions of two samples are shown in Table 3.

Two orders of the 16 samples were recorded. Both began with the same four (preexposure) samples, but the remaining 12 (test) samples were randomly ordered differently in the two versions.

Apparatus. The same test equipment was used as described for Experiment 1.

Procedure. The same procedure was used as described for Experiment 1.

TABLE 3

Examples of the Alternate Versions of Two Storybook Samples

Coincident version
The little boy at the piano / is having a birthday party. All of his friends / like to sing. The happy little boy / loves to play music for his friends. The little boy's parents / gave him the piano for his birthday. The boy and his friends / are having a good time.
Predicate-interrupted noncoincident version
The little boy at the piano is having / a birthday party. All of his friends like / to sing. The happy little boy loves / to play music for his friends. The little boy's parents gave / him the piano for his birthday. The boy and his friends are having / a good time.
Coincident version
Many different kinds of animals / live in the zoo. The dangerous wild animals / stay in cages. Some of the animals / are friendly and like to be petted. Chimpanzees, gorillas, and apes / are in the monkey house. Many people at the zoo / like the monkeys best of all. The playful little monkeys / climb in their cages. The crowd in the monkey house / laughs at the silly monkeys.
Subject-interrupted noncoincident version
Many different kinds / of animals live in the zoo. The dangerous / wild animals stay in cages. Some / of the animals are friendly and like to be petted. Chimpanzees, / gorillas, and apes are in the monkey house. Many people / at the zoo like the monkeys best of all. The playful / little monkeys climb in their cages. The crowd in / the monkey house laughs at the silly monkeys.

Results and Discussion

The data from the 12 test trials were used to compute the mean length orientation for each type of sample. Across subjects, the mean length orientation for all the Coincident versions was 16.23 s as opposed to 10.15 s for all the Noncoincident ones. Analyses with subjects as the random factor ($t(23) = 12.33$, $p < .001$) and with samples as the random factor ($t(11) = 10.87$, $p < .001$) confirmed that this difference is significant. In fact, all 24 subjects looked longer to the Coincident than to the Noncoincident versions. In addition, for 12 out of 12 samples, the Coincident version had the longer orientation time. Most important, as shown in Table 4, a breakdown of the data according to whether the pauses in the Noncoincident versions occurred within the subject–noun phrase or within the predicate–verb phrase indicated no significant differences between these two locations ($F(1,10) = 1.68$ for the main effect of pair type and $F(1,10) < 1.00$ for the interaction between pair type and Coincident vs Noncoincident versions). Hence, there is no indication that the findings are explicable on the grounds of a preference for early versus late pauses. Also, there is every evidence that the effect generalizes to both subject–noun and predicate–verb phrases.

The present results provide strong support for the contention that infants in this age range are sensitive to acoustic correlates of major phrasal units. Thus, the results of the previous experiments were replicated with an entirely new set of materials under circumstances that also eliminated a possible alternative account based on the temporal location of the pauses. The fact that the infants continued to listen longer to the Coincident samples supports the view that they are responding to acoustic correlates of phrasal units.

Earlier we noted that utterances that include long subject–noun phrases may be better marked by intonational cues than ones that include short subject–noun phrases. For this reason, one might expect that the stimuli in the present experiment would include more salient markers than the ones in the previous experiments. Nevertheless, another difference between the stimulus sets is that the present stimuli were read, whereas those in the previous experiments were selected from spontaneous speech. Previous work comparing the prosodic organization of spontane-

TABLE 4
Mean Length Orientation Times for Storybook Samples in Experiment 3

Type of sample pair	Coincident	Noncoincident
Predicate-interrupted	16.42 s	10.12 s
Subject-interrupted	15.60 s	9.87 s

ous and read prose suggests that read prose tends to have less variability in fundamental frequency than does spontaneous speech (Remez, Rubin, & Ball, 1985; Remez, Rubin, & Nygaard, 1986). In this respect, the intonational cues may actually have been less distinctive with the read samples than with the spontaneous speech samples. The data suggest that any reductions in distinctiveness that may have come about due to reading were more than offset by the presence of the longer subject–noun phrases. Thus, if anything, the preference for the Coincident over the Noncoincident versions was more marked here than in Experiment 1 (under fully comparable conditions). In the present study, 100% of the subjects displayed a preference for the Coincident versions as compared to 79% in Experiment 1.

In one of our previous investigations (Kemler Nelson et al., 1989), we found that infants were sensitive to acoustic correlates of clausal units in Child-directed speech but not in Adult-directed speech. The most plausible interpretation attributes this difference to the presence of exaggerated and perhaps more reliable prosodic cues that are typically observed for Child-directed speech (Fernald & Simon, 1984; Fernald et al., 1989; Greiser & Kuhl, 1988). We argued that the exaggerated prosodic marking might provide a better indicator of speech segments corresponding to clausal units in the input. Might the same be true for phrasal units?

EXPERIMENT 4

The existence of the Storybook materials afforded us with the possibility of presenting the same content in both the Child-directed and Adult-directed speech conditions. By employing the same talker to record both the Child-directed and Adult-directed samples, we could control for individual characteristics associated with the talker's voice as well as for the content of the materials. To provide a thorough test of any differences that might occur for Child-directed and Adult-directed speech, we decided to test a group of infants on the Child-directed Storybook samples in addition to a group on the Adult-directed Storybook samples. Hence, the present experiment also provides an opportunity to replicate and extend the findings of Experiment 3.

Method

Subjects. The subjects were 48 infants from English-speaking homes in the Eugene, Oregon area. The infants had an average age of 9.3 months (range: 8.8 months to 10.3 months). In order to obtain the 48 subjects, it was necessary to test 78. Subjects were excluded for the following reasons: failure to look to one of the two sides for at least three test trials (18 Ss), crying (11 Ss), and experimenter error (1 S).

Stimuli. For the Child-directed samples, the same stimuli were employed as in Experiment 3. For the Adult-directed samples, we recorded the same talker who had produced the Child-directed samples. She returned to the laboratory and read the Storybook materials as naturally as possible to an undergraduate student. Once again, she did not know what the materials were to be used for. The mean duration of the Adult-directed samples was 22.07 s; the range in durations was 18.12 to 25.04 s. Thus, the overall durations of the Adult-directed samples were on average about 4-s shorter than were those of the Child-directed samples (26.07 s) covering the same material. A sample-by-sample comparison for the Adult- and Child-directed materials indicated that in each instance, the Adult-directed sample was shorter than the comparable Child-directed one. A paired t test confirmed this difference to be significant ($t(11) = 4.89$, $p < .001$). Thus, the speaking rate for the Adult-directed samples was faster than for the Child-directed samples in line with what has been observed in previous studies (e.g., Broen, 1972; Remick, 1971).

The tape recordings of both the Child-directed and Adult-directed samples were digitized and stored on the LSI 11/73 at the University or Oregon using a 12 bit A/D converter. The Coincident and Noncoincident versions of the Child-directed and Adult-directed samples were produced by introducing 1-s pauses in the same locations as described for Experiment 3. Again, care was taken to insert the pauses at zero crossings in amplitude. Thus, the Child-directed and Adult-directed Coincident samples had pauses in exactly the same locations. Similarly, the Child-directed and Adult-directed Noncoincident versions were matched in the location of the pauses. Within pairs, the Coincident and Noncoincident versions of each sample were identical in all respects except for the location of the pauses. Finally, as in Experiment 2, on-line, real-time presentation of the stimuli was provided by the computer which was also programmed to select randomly the test pairs for each trial. Hence, each subject had a different random ordering of the 12 test pairs.

Apparatus. The same test equipment was used as described in Experiment 2.

Design and procedure. Half of the infants were assigned randomly to the Child-directed condition and the other half to the Adult-directed condition. In all other respects, the procedures were identical to those described for Experiment 2.

Results and Discussion

The mean length orientation times for the Coincident and Noncoincident versions under the Child-directed and Adult-directed speech conditions are shown in Table 5. In both conditions, infants displayed significantly longer orientations for the Coincident versions ($t(23) = 4.74$, $p < .0001$ for the Child-directed condition and $t(23) = 2.13$, $p < .05$ for the Adult-directed condition). In addition, the same tendencies were manifested in the analyses with samples as a random factor ($t(11) = 4.31$, $p < .01$ for the Child-directed condition and $t(11) = 2.86$, $p < .02$ for the Adult-directed condition). Overall, 21 of 24 subjects in the Child-directed condition and 16 of 24 subjects in the Adult-directed condition had longer orientation times for the Coincident versions. In addition, the Coincident

TABLE 5
Mean Length Orientation Times for Child-directed and Adult-directed Storybook Samples in Experiment 4

	Coincident	Noncoincident
Child-directed	9.20 s	6.55 s
Adult-directed	7.69 s	6.61 s

version had longer orientation times in 12 of the 12 sample pairs for the Child-directed condition and in 10 of the 12 pairs in the Adult-directed condition.

Although infants in both the Child-directed and Adult-directed conditions displayed significant preferences for the Coincident versions, the effects appeared to be stronger for the Child-directed samples. To investigate this, difference scores were computed for each subject by subtracting the mean length orientation times for the Noncoincident versions from those for the Coincident versions. The resulting scores were compared for the two conditions and indicated only a marginally significant difference in favor of the Child-directed condition ($t(22) = 1.93$, $p < .10$). An analogous analysis conducted with samples as a random factor did not produce evidence of a significant difference ($t(10) = 1.14$, $p < .20$).

The preference data for the Child-directed condition completely replicate those reported for Experiment 3. They indicate that 9-month-old infants are sensitive to acoustic markers of speech segments that relate to phrasal units. Moreover, the results of the Adult-directed speech condition suggest that these samples, too, contain sufficient information to allow infants to detect acoustic markers of major phrasal units. This latter finding is a bit surprising given previous work suggesting that infants are sensitive to acoustic correlates of clausal units in Child-directed speech, but not in Adult-directed speech (Kemler Nelson et al., 1989). However, the overall trend here was in the same direction (i.e., the cues appear to be more readily available in Child-directed speech). In addition, the use of texts that consistently contained long subject–noun phrases (and maybe even the use of speech produced by reading content appropriate to children) may have provided more reliable markers of the units than is ordinarily available in spontaneous Adult-directed speech. Perhaps, with such regular marking, the additional boost provided by the exaggeration of prosodic features in Child-directed speech is not absolutely necessary. Indeed, there is some indication that matched Child-directed and Adult-directed samples provided many of the same cues. A comparison between the Child-directed and Adult-directed groups involving the orientation time differences (Coincident–Noncoincident) for the 12 sample pairs revealed a significant correlation ($r(11) = .52$, $p < .02$). Thus, there is some indication that the two groups were reacting to many of the same aspects of the stimuli.

The consistent finding that emerges over the first four experiments is that by 9 months of age infants are sensitive to acoustic markers that provide information about the mapping of speech to major phrasal units. Given that they are in the beginning stages of language acquisition, this sensitivity may be an important component in the processes relevant to determining the regularities that hold among sentences in the native language. But once the native language has been acquired and other cues (e.g., semantic or syntactic ones) are available to the listener, does the sensitivity to acoustic markers that relate phrasal boundaries continue? There is some suggestive evidence from studies of adults listening to foreign language utterances that they are sensitive to some acoustic markers of grammatically relevant units (e.g. Pilon, 1981; Wakefield et al., 1974). Similarly, there are indications from some perceptual studies that segmental and prosodic cues may be called upon in disambiguating syntactic boundaries in the native language (e.g., Collier & t'Hart, 1975; Lehiste, Olive, & Streeter, 1976; Price et al., 1991; Scott, 1982; Scott & Cutler, 1984). Findings such as these encouraged us to explore the way in which adults responded to the kinds of materials that we used with infants.

EXPERIMENT 5

Testing adults on the Coincident and Noncoincident versions of the samples should contribute to our understanding of the infant data in a number of ways. First, although we have proceeded on the assumption that infants show a preference for the Coincident samples because the segmentations therein are perceived to be more natural than for the Noncoincident versions, we have not been able to put this question to them directly. An indication that adults explicitly judge the Coincident samples to be perceptually "more natural" than the Noncoincident ones makes our interpretation of the infant results more plausible. Second, by examining adults' perception of both the Child-directed and Adult-directed materials, we may follow up on the suggestion that the perceptual effects with infants may be stronger with the Child-directed samples. Does the Child-directed speech provide clearer perceptual markers of the units?

One problem that is posed by testing adults on the samples which did not arise with the infants has to do with their prior knowledge of the syntax and semantics of English. Were we to play the utterances used in the previous experiments to adults, then they might show a preference on grounds that have to do with linguistic content only. In fact, the speech perception literature provides a number of examples where what was first thought to be a perceptually based judgment, like assigning degrees of stress to syllables in sentences, was shown to be based more on semantic and syntactic considerations than on purely perceptual ones (e.g., Lieberman, 1963, 1967). For this reason, it was necessary to ensure that our adult subjects would have access only to whatever information is available in the prosody. To achieve this, we low-pass filtered our original samples at a frequency level that made the identification of individual words impossible.

Method

Subjects. The subjects were 16 college-aged students enrolled in an Introductory Psychology class at the University of Oregon. The students participated in the experiment as a means of fulfilling a course requirement. The students were all native speakers of English and had no known hearing deficits. One additional subject was tested but dropped from the study for failing to answer on too many trials.

Stimuli. The same stimulus materials used in Experiment 4 were employed in the present study. The one change that occurred was that the audio output of the samples was passed through a Krohn-Hite filter with the low-pass cutoff set to 400 Hz with an attenuation slope of 48 dB per octave. This filter level which has been used in previous studies (e.g., Fernald, 1989; Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988) is sufficient to eliminate almost all of the distinctive phonetic information from the samples while leaving intact prosodic features such as intonation, stress and rhythm.

Apparatus. All subjects were tested in a sound-insulated room equipped with six test cubicles. A response box was presented at each cubicle. Each box was connected to the PDP 11/73 computer and permitted the on-line registration of subjects' responses to the sample pairs. For purposes of the experiment, one of the buttons was marked as response "1" and another as response "2." Each booth was equipped with a set of matched and calibrated TDH-39 headphones over which the test pairs were heard. The stimulus materials were stored in digital form on the computer and reconverted by a 12 bit D/A converter and low-pass filtered for presentation to the subjects.

Design and procedure. Subjects were tested in small groups of three to six subjects. Half of the subjects heard the Child-directed materials first and the other half heard the Adult-directed speech samples first. A short break intervened and then the remaining set of samples was played. The subjects were instructed that they would be listening to some speech-like samples. They were told that on each trial they would hear two samples that were very similar, but not identical. Their task was to choose the one that sounded most natural, i.e., sounded more as if it preserved the melodies and/or rhythm of normal speech. The subjects were also encouraged to listen for the melodies and not words because the stimuli were altered so that no words could be identified. On each trial, subjects were presented with the Coincident and matched Noncoincident version of each sample in a randomly determined order. Responses were indicated by pressing "1" if the first sample sounded more natural and "2" if the second one sounded more natural. The presentation of the trials was paced to the responses of the slowest subject. A warning light came on 1 s prior to the onset of the next trial to alert the subjects. For each type of material (i.e., Child-directed or Adult-directed samples), the subjects made judgments about all 16 pairs (i.e., the four preexposure pairs as well as the 12 test pairs that were used in the infant experiments). The whole procedure took about a half hour to complete.

Results and Discussion

A preliminary analysis of the data indicated that there were no significant effects attributable to test order, i.e., whether the Child-directed or Adult-directed speech samples were heard first ($F(1,14) < 1.00$). Accordingly, the data were collapsed across test order for the rest of the data analysis. For the Child-directed samples, the Coincident version was chosen as the more natural sounding on 69.1% of the trials. For the Adult-directed samples, the Coincident version was selected on 72.6% of the trials. In both instances, responding was significantly greater than the

50% level expected by chance (i.e. $t(15) = 4.71$, $p < .001$ for the Adult-directed condition; $t(15) = 3.26$, $p < .01$ for the Child-directed). A paired t test indicated that the difference between the Child-directed and Adult-directed conditions was not significant ($t(15) = 0.75$). Finally, further analysis of the data indicated that there were no significant differences according to whether the pauses in the Noncoincident versions occurred within the subject noun or predicate verb phrases ($t(15) = 0.67$). The present results indicate that adults judged the Coincident versions of the samples to be "more natural" than the Noncoincident versions. This suggests that the division of the utterances into parts at the major phrasal boundary between subject and predicate constitutes a more natural perceptual partitioning of the utterance for them than the alternative positions within the verb or subject noun phrases. Thus, at least for the present set of materials, adults are attentive to acoustic correlates of major phrasal units. These results with low-pass filtered utterances are in line with previous findings using foreign language strings (Pilon, 1981; Wakefield et al., 1974).

What might be the acoustic cues to which both infants and adults are sensitive? As noted earlier, studies of fluent speech in adults have turned up evidence that boundaries between important grammatical units such as clauses and phrases are marked by changes in prosodic structure such as drops in fundamental frequency, increases in syllable duration, pausing, etc. (e.g., Cooper & Paccia-Cooper, 1980; Grosjean & Gee, 1987; Klatt, 1975; Martin, 1970; Nakatani & Dukes, 1977; Nakatani & Schaffer, 1978; Price et al., 1991). There are also a number of recent indications in the language acquisition literature (e.g., Fisher, 1991; Lederer & Kelly, 1991; Morgan, 1986) that phrasal units may be marked by fundamental frequency changes and increases in syllable duration. Moreover, in a recent study of infants' perception of musical phrasal units, which parallels the present study, Krumhansl and Jusczyk (1990) found that infants preferences for musical samples correlated highly with the presence of certain acoustic changes in the vicinity of phrasal boundaries. Specifically, the pitch of the melodic line tended to drop and the duration of the last note increased prior to musical phrase boundaries in the Coincident samples used by Krumhansl and Jusczyk. In order to provide a better indication of whether significant changes in prosody occur in the vicinity of phrasal boundaries in the samples used in the present study, we decided to undertake some acoustic analyses of the Storybook samples.

For purposes of analysis, we transferred the digitized versions of both the Child-directed and Adult-directed Storybook samples to a VAXstation Model 3176 computer. A specially developed waveform editor was used to visually display the samples. Following Krumhansl and Jusczyk (1990), we measured the average pitch of the vowels of the last three syllables prior to and the first two syllables following each pause for both the Coincident and Noncoincident versions of the samples. Average pitch was determined by counting the individual pitch pulses from the center of each vowel. Whenever possible, we sampled at least 50 ms from each vowel. The average values for each of the five syllables for the Coincident and Noncoincident versions are presented in Figs. 1 (Child-directed samples) and 2 (Adult-directed samples). Inspection of the figures suggests that the patterns for the Child-directed and Adult-directed samples were much the same, although the overall pitch height is higher for the Child-directed samples. In both cases, the Coincident versions of the samples give indications of a pitch drop in the vicinity of the pauses. This declination in pitch does not appear in the case of the Noncoincident samples.

To confirm these observations, the average pitch values for each syllable position of each sample were entered into an ANOVA for a 2 (Type: Child- vs Adult-directed) by 2 (Version: Coincident vs Noncoincident) by

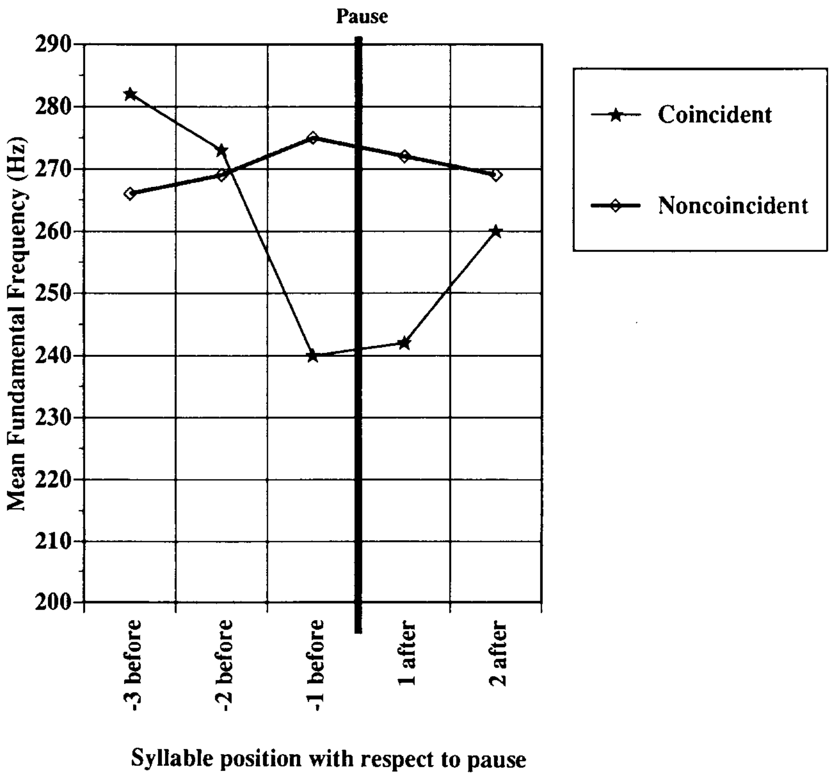

FIG. 1. The mean fundamental frequency for each of the last three syllables prior to and the first two syllables after each pause for the Coincident and Noncoincident versions of the Child-directed Storybook samples.

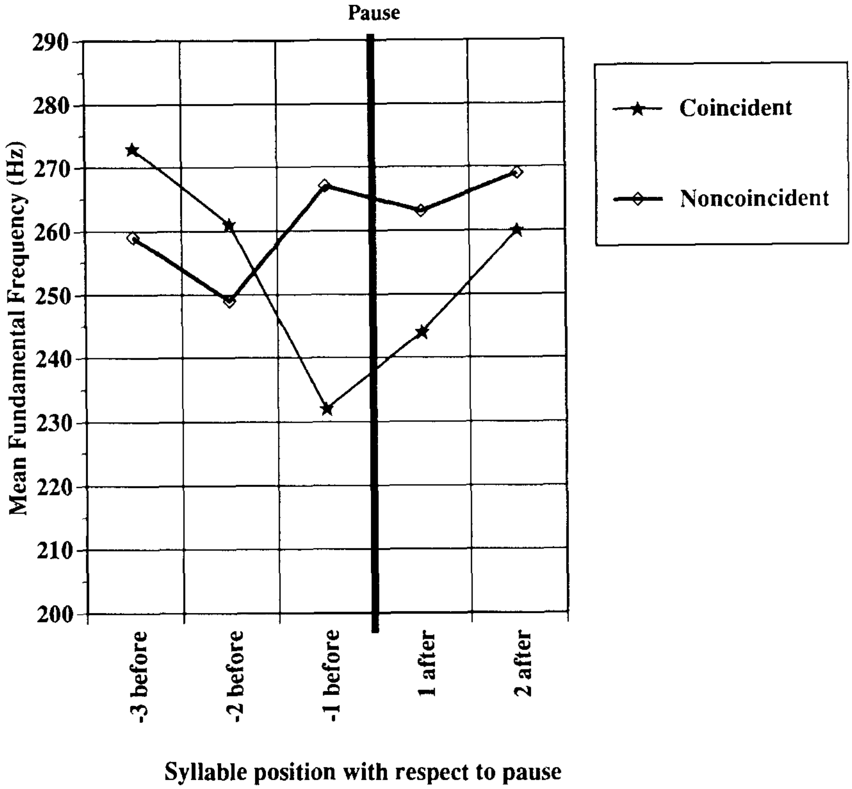

FIG. 2. The mean fundamental frequency for each of the last three syllables prior to and the first two syllables after each pause for the Coincident and Noncoincident versions of the Adult-directed Storybook samples.

5 (Syllable Position) design. The main effects of Type ($F(1,220) = 9.22$, $p < .005$), Version ($F(1,220) = 5.24$, $p < .025$) and Syllable Position ($F(4,220) = 2.97$, $p < .025$) were all significant. However, the only significant interaction was the one between Version and Syllable Position ($F(4,220) = 6.695$, $p < .001$). Thus, these results bear out the observations from the figures. In addition, we analyzed the data to determine where there was evidence of significant pitch differences among the various syllable positions. For this purpose, we used Tukey's test with Cicchetti's adjustment (Cicchetti, 1972). A statistically significant result at the $p < .05$ level, required a difference of 24.55 Hz between any two syllable positions. Differences of this magnitude were found only for the Coincident versions of the samples and occurred between the third and first syllables prior to the break and between the second and first syllables prior to the break. These results confirm that there is a significant drop in

pitch at the syllable immediately prior to the break. The only other reliable difference was between the third syllable prior to the break and the first syllable after the break.

Unfortunately, the samples were not constructed in such a way as to permit a very informative analysis of syllable durations. To do this, we would either have had to construct the samples so as to permit measurements of identical items at boundary and nonboundary locations within the phrases (e.g., Morgan, 1986; Price et al., 1991), or have had access to computational algorithms which normalize vowel durations across different phonetic contexts. Nevertheless, even with these limitations, we decided to examine whether there were any consistent durational differences for the syllable immediately prior to the breaks in the Coincident and Noncoincident versions of the samples. Using the waveform editor, we measured the duration of the last syllable just prior to each break in our samples. For the Child-directed samples, the average duration of this syllable was 271 ms for the Coincident versions and 241 ms for the Noncoincident versions. Similarly, for the Adult-directed samples, the average duration was 289 ms for the Coincident versions and 234 ms for the Noncoincident versions. To determine the statistical reliability of these observations, the average durations of the last syllable prior to the break for each sample were entered into an ANOVA for a 2 (Type: Child- vs Adult-directed) by 2 (Version: Coincident vs Noncoincident) design. There was a highly significant main effect of Version ($F(1,44) = 11.90$, $p < .001$), but neither the main effect of Type ($F(1,44) = 0.07$) nor the interaction of Type × Version ($F(1,44) = 0.67$) was significant. Paired t tests confirmed that the longer durations for the Coincident versions were statistically significant for each type of material (Child-directed: $t(11) = 2.21$, $p < .05$; Adult-directed: $t(11) = 3.34$, $p < .01$). Thus, even with the rather crude comparison used here, there is some indication that the duration of the last syllable prior to the break is significantly longer for the Coincident than for the Noncoincident samples.

The acoustic analyses of fundamental frequency and duration that we performed indicate that the Child-directed and Adult-directed samples are more or less comparable in the degree to which phrasal boundaries are marked. The findings from adults in the present experiment are fully consistent with this pattern. Specifically, there was no indication that the Child-directed samples provided adults with more salient cues to the phrasal units than did the Adult-directed samples. In both cases, there were indications of significant changes in pitch and duration prior to the breaks in the Coincident samples. The absence of significant differences between the Adult- and Child-directed versions is interesting in view of previous reports which find consistent differences in favor of Child-directed speech (e.g., Berstein Ratner, 1985; Broen, 1972; Fernald &

Simon, 1984; Garnica, 1977). We suspect that the nature of the spoken materials may have had something to do with this. The passages used here were based on stories intended for children. Even though a faster speaking rate was used in the presence of an adult listener, the materials may lend themselves to some exaggeration of prosodic characteristics. Certainly, the average pitch of the Adult-directed speech in our materials (254 Hz, as compared to 265 Hz for Child-directed speech) is somewhat higher than what is reported in previous comparisons of spontaneous speech to adults and children (e.g., Fernald et al., 1989).

What, then, can we make of the earlier hints that Child-directed speech was apparently more effective than Adult-directed speech for the infants? Of course, there may be other cues for which we have not performed an acoustic analysis that may be differentially available in the Child-directed and Adult-directed samples. If so, perhaps infants are sensitive to these. Another possibility is simply that the Child-directed speech samples, as long as they sound reasonably natural, do a better job of attracting infants' attention than do the Adult-directed ones. Fernald (1984) suggested this as one of the possible functions of Child-directed speech. Moreover, she and other researchers have shown (e.g., Cooper & Aslin, 1990; Fernald, 1985; Fernald & Kuhl, 1987; Werker & McLeod, 1989), that infants are attracted to the acoustic properties of Child-directed as compared to Adult-directed speech. In Experiment 4, the longer listening times shown by the infants presented with samples that were both Child-directed and naturally segmented (i.e., Coincident) is consistent with this view.

To gain a better understanding of the extent to which the acoustic markers that we identified for the Storybook samples are also present in conversational speech, we decided to analyze the Spontaneous speech samples used in Experiments 1 and 2. Following the procedures used with the Storybook samples, we measured the average pitch of the vowels of the last three syllables prior to and the first two syllables following each pause for both the Coincident and Noncoincident versions of the samples. These average pitch values are presented in Fig. 3. Once again, for the Coincident versions, there appears to be a significant drop in pitch between the third and first syllables prior to the pause. To confirm this tendency, the average pitch values for each Syllable Position of each sample were entered into an ANOVA for a 2 (Version: Coincident vs Noncoincident) × 5 (Syllable Position) design. Neither the main effect of Version ($F(1,110) = 0.28$) nor of Syllable Position ($F(4,110) = 1.37$, $p > .20$) was significant. However, the interaction between these two variables ($F(4,110) = 2.418$, $p = .053$) was very close to the $p < .05$ level of statistical significance. We examined the data to determine whether there was evidence of significant differences across the various syllable positions. A statistically significant result at the $p < .05$ level by a Tukey's test

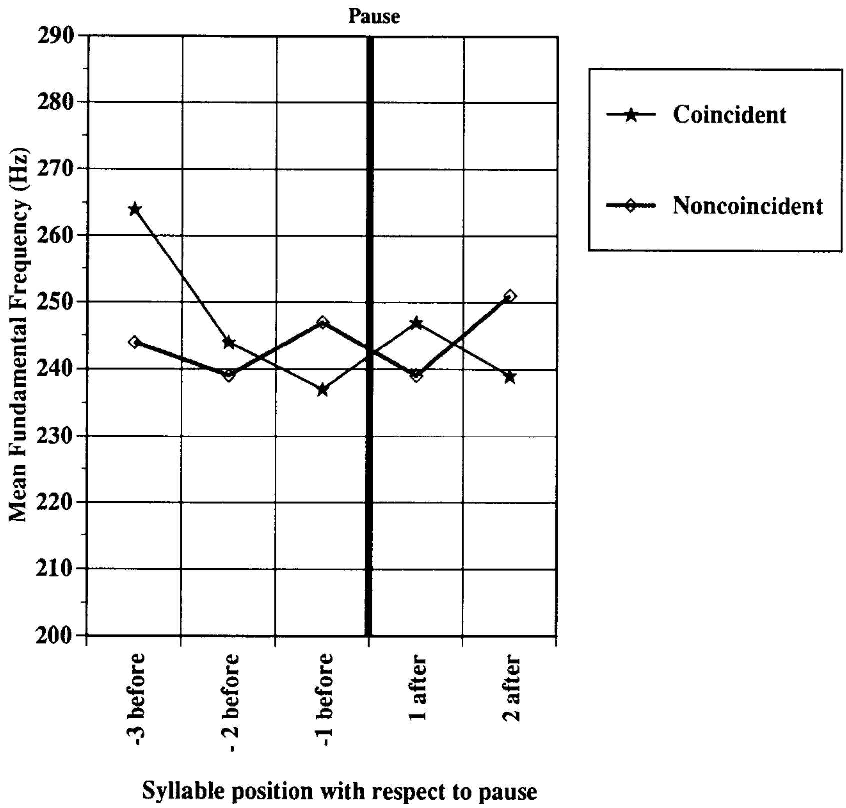

FIG. 3. The mean fundamental frequency for each of the last three syllables prior to and the first two syllables after each pause for the Coincident and Noncoincident versions of the Spontaneous speech samples.

with Cicchetti's adjustment required a difference of 26.56 Hz between any two syllable positions. Only one comparison exceeded this magnitude and it occurred between the third and first syllables prior to the break for the Coincident versions of the samples. Thus, for the Spontaneous samples, as for the Storybook samples, there is a significant drop in pitch between the third and first syllables prior to the break for the Coincident versions.

We also measured the durations of the syllable immediately prior to the pause for the Coincident and Noncoincident versions of the samples. The average duration of this syllable position was 156 ms for the Coincident versions and 177 ms for the Noncoincident versions. A paired t test using the average duration of this syllable position for the Coincident and Noncoincident versions of each sample indicated no significant difference on this measure ($t(11) = 1.62, p > .10$). Thus, there was no indication of the kind of durational differences for the last syllable prior to the break in the

Spontaneous samples that were observed for the Storybook samples. We do not think that it is prudent to make too much of this discrepancy. For one thing, in the Coincident versions of the Spontaneous samples, the last syllable prior to the break was often a pronoun or auxiliary of the verb, whereas, in the Storybook samples, it was often a noun. Moreover, as noted earlier, this durational measure is less than optimal. It provides information about the absolute durations in a syllable position, whereas there is reason to believe that perceived lengthening is most likely based on the relative length of an item—i.e., compared to its duration in other sentential positions (e.g., Morgan, 1986; Price et al., 1991).

In summary, the preliminary sorts of acoustic analyses that we were able to conduct on our materials do suggest the presence of acoustic markers of phrasal units. In particular, there are clear pitch drops prior to phrase boundary breaks for the Coincident versions of both the Storybook and Spontaneous speech samples. In addition, for the Storybook materials, there is some indication that the average duration of the syllable immediately prior to a pause is longer in the Coincident versions than in the Noncoincident versions of the samples. Thus, we now have evidence that there are potential acoustic markers of phrasal units in our materials, and that infants listen longer to samples that are consistent with these markers. Can we make the inference that they are responding to cues of this nature, rather than to other kinds of marking of phrasal units?

EXPERIMENT 6

We have been making the assumption that the critical information to which infants are responding occurs in the prosody of the utterances and we have identified some such prosodic units to which they might be sensitive. However, as we noted earlier, there is other information in the speech signal which could play a role in marking important syntactic units in the utterance. For example, the presence and placement of certain morphemes could also provide information relevant to marking phrasal units (e.g., Morgan et al., 1987). Of course, information of this sort is available along with potential prosodic markers in the materials that we have presented to infants. How can we determine whether the information in the prosody is sufficient to have produced longer listening times to the Coincident samples? One possibility is to follow the strategy that we used with our adult subjects to ensure that they did not rely on their knowledge of syntax and semantics in judging the samples. In other words, low-pass filtering our samples will separate the prosody from the kind of information that allows for the identification of morphemic markers in the input. Accordingly, we decided to test two new groups of infants on low-pass filtered versions of both the Spontaneous speech samples (Experiments 1 and 2) and the Child-directed Storybook speech samples.

Method

Subjects. The subjects were 48 9-month-old infants form English-speaking homes in the Eugene, Oregon area. The infants had an average age of 9.2 months (range: 8.6 months to 10 months). In order to obtain the 48 subjects, it was necessary to test 57. Subjects were excluded for the following reasons: failure to look to one of the two sides for at least three trials (6 Ss) and crying (3 Ss).

Stimuli. The stimulus materials were the spontaneous speech samples (see Experiments 1 and 2) and the Child-directed Storybook samples (see Experiments 3 and 4). These samples were already digitized and stored on the PDP 11/73 computer. They were played out over a 12 bit D/A converter. The audio signal was passed through a Krohn-Hite filter with the low-pass cutoff set to 400 Hz.

Apparatus. The same test equipment was used as described in Experiment 2.

Design and procedure. Half of the infants were assigned randomly to the Spontaneous speech samples and the other half to the Child-directed Storybook samples. In all other respects, the test procedure followed was identical to that outlined for Experiment 2.

Results and Discussion

The mean length orientation times for the Filtered Coincident and Noncoincident versions of the Spontaneous speech and Child-directed Storybook Samples are shown in Table 6. The mean orientation times of each subject to the Coincident and Noncoincident versions of the samples were submitted to ANOVAs for a 2 (Materials: Spontaneous/Storybook) by 2 (Version: Coincident/Noncoincident) mixed design. With subjects as a random factor, this analysis yielded a significant main effect of Version ($F(1,92) = 10.16, p < .002$). Neither the main effect of Materials ($F(1,92) = 1.53, p = .219$) nor the interaction of Materials × Version ($F(1,92) = 1.77, p = .186$) was significant. The same analysis conducted with samples as a random factor yielded exactly the same pattern of results—a significant main effect of Version ($F(1,44) = 8.62, p < .005$), and no significant main effect of either Materials ($F(1,44) = 1.06, p = .310$) or the interaction of Materials × Version ($F(1,44) = 1.69, p = .203$). Thus, the present results show that even with low-pass filtering, the preference

TABLE 6
Results of *Filtered* Phrase Study with 9 Month Olds

	Mean orientation times	
	Coincident	Noncoincident
Spontaneous speech	7.61 s	6.57 s
Child-directed storybook	9.05 s	6.51 s

for the Coincident versions of the samples still remains. Hence, the preferences that were observed for the Coincident versions of the unfiltered samples cannot be attributed solely to the availability of information about the syntactic and semantic features of the samples. The prosodic information available in the low-pass filtered samples is a sufficient basis for establishing the infants' preferences for the Coincident versions.

For purposes of comparison with the earlier experiments, we also did separate analyses for each type of materials. For the Child-directed Storybook Samples, the individual analyses followed the pattern of the overall analysis. In particular, the preference for the Coincident over the Noncoincident versions was confirmed in separate analyses with subjects ($t(23) = 4.25$, $p < .001$) and samples ($t(11) = 3.66$, $p < .01$) as random factors. In all, 19 of 24 subjects had longer orientation times for the Coincident versions, and for 10 of the 12 sample pairs, the longer orientation times occurred for the Coincident versions. Although in the same direction, the results of the individual analyses for the filtered Spontaneous speech samples were weaker. In particular, whereas the analysis with samples as a random factor indicated a significant preference for the Coincident versions ($t(11) = 2.21$, $p < .05$), the analysis with subjects as a random factor did not reach statistical significance ($t(23) = 1.47$, $p = .156$). Overall, 17 of 24 subjects had longer orientation times for the Coincident versions. In addition, the Coincident version received longer orientation times for 9 of the 12 sample pairs.

The fact that the results with the filtered stimuli were more robust for the Storybook samples than for the Spontaneous samples suggests that they may be more effective in providing information about phrasal units.[3] Indeed, the results of the acoustic analyses that we conducted on the samples are certainly consistent with this view. There was evidence for the Storybook samples of both pitch and durational cues to phrasal units, but for the Spontaneous samples, only the pitch changes were reliable. This raises an interesting question about the developmental differences that we observed in Experiment 2 with the Spontaneous speech samples. Recall that 6 month olds also did not display a significant preference for the Coincident versions. We interpreted this as an indication that the capacity to detect acoustic markers of phrasal units had not yet developed at this age. However, given the present indications that the Storybook materials may provide a more favorable presentation of acoustic markers for phrasal units, then perhaps 6 month olds would display a sensitivity

[3] Interestingly enough, in a separate study we carried out with adult subjects listening to low-pass filtered versions of the Spontaneous speech samples, judgments identifying the Coincident version as being more "natural" did not differ significantly from chance.

for these units with the Storybook materials. This was the impetus for conducting the next experiment.

EXPERIMENT 7

The results of Experiment 2 provided some suggestion that the sensitivity to acoustic markers correlated with phrasal units may develop between 6 and 9 months of age. However, we could not rule out the possibility that the younger infants are sensitive to them but just did not show it with the speech materials that we employed. For example, our spontaneous samples did contain a number of yes–no questions and may not have provided as consistent or salient marking of the phrasal units as do the storybook samples. The 9 month old may simply be better able to cope with any such inconsistencies than the 6 month old. Consequently, the 6 month old might first exhibit a sensitivity to acoustic markers of phrasal units with materials that offer a more consistent marking. Thus, we decided to test a group of 6-month-old infants using the Child-directed Storybook materials.

Method

Subjects. The subjects were 16 6-month-old infants from English-speaking homes in the Eugene, Oregon area. The infants had an average age of 6.0 months (range: 5.5 months to 6.7 months). In order to obtain the 16 subjects, it was necessary to test 27. Subjects were excluded for the following reasons: failure to look to one of the two sides for at least three trials (6 Ss) and crying (5 Ss).

Stimuli. The Child-directed Storybook samples used in Experiment 3 and 4 were employed in this study.

Apparatus. The same test equipment was used as described in Experiment 2.

Procedure. The procedure followed was identical to that outlined for Experiment 2.

Results and Discussion

The mean duration orientation times were 6.80 s for the Coincident versions and 6.34 s for the Noncoincident versions. The difference between the two groups was not significant ($t(15) = 0.58$, $p > .30$). The analysis conducted with samples as a random factor also indicated no significant difference between the Coincident and Noncoincident versions ($t(11) = 0.78$, $p > .20$). Overall, only seven of the 16 subjects had longer orientation times for the Coincident versions. Similarly, an examination of the individual samples showed that in six cases the Coincident version had the longer orientation time, whereas in the other six cases, the Noncoincident versions received longer orientation. These results contrast strongly with those for the 9-month-old infants in Experiment 4 where for 21 of the 24 subjects and for all 12 of the samples, the Coincident version had longer orientation times. When the data from the 6 month olds of the current study were systematically compared to the data of the 9 month olds in the Child-directed condition of Experiment 4, an ANOVA with subjects as the random variable, revealed the critical interaction between Age and Coincident/Noncoincident type of sample ($F(1,38) = 4.79$, $p < .03$). A similar analysis, this time with samples as a random factor, also indicated a significant difference in favor of the 9-month-old group ($t(11) = 2.11$, $p < .05$, one-tailed).

So, the present results provide no indication that 6-month-old infants are sensitive to cues relevant to major phrasal units and, combined with Experiment 4, suggest that the sensitivity increases over the second-half year of life. These results basically replicate those that we observed in Experiment 2 with the spontaneous speech samples. The implication is that the earlier failure was not simply the result of an inability of the 6 month olds to cope with a possibly less optimal marking of the phrasal units. In contrast, infants of the same age, tested in the same setting, do show consistent preferences for segmentations that correspond to *clausal units* (Hirsh-Pasek et al., 1987). Thus, the lack of consistent preferences in the present study cannot be attributed to an overall inability of 6 month olds to perform the task itself. Rather, the present results reinforce the view that the capacity to attend to acoustic markers of phrasal units may not be fully developed in 6 month olds.

GENERAL DISCUSSION

The most basic finding that emerges from the present series of studies is that 9-month-old infants are sensitive to acoustic markers that correspond to major phrasal units. The robustness of this finding is given by the fact that it occurred with two different sets of materials, Spontaneous speech and Storybook samples, and for five different groups of infants across Experiments 1–4. Moreover, the effects were significant across sample pairs as well as across subjects. In each instance, infants displayed preferences for samples segmented at a major phrasal boundary as opposed to locations elsewhere within either the subject–noun or predicate–verb phrases. Finally, the same pattern of results occurred in Experiment 6, in which stimuli were low-pass filtered to eliminate cues to linguistic content. These latter findings provide support for the view that infants are responding to information available in the prosody of the utterances.

A second important finding is the strong suggestion of a developmental change in sensitivity to these acoustic correlates between 6 and 9 months of age. In contrast to the 9 month olds, none of the three groups of 6 month olds tested on the spontaneous speech and Storybook materials ever showed significant evidence of a preference for the samples that were segmented at the major phrase boundary. Moreover, the direct comparisons of the two age groups offer further indication that they differed in their responses to the sample pairs. Although the age group comparisons were sometimes only marginally significant, they were consistently in the direction of stronger preferences on the part of the 9-month-old infants.

The fact that 9 month olds are sensitive to acoustic correlates of major phrasal units has potentially important implications for language acquisition. As noted at the outset, many critical grammatical properties and dependencies are defined at the level of phrases. Mastery of a language requires discovering just which properties and dependencies hold among the utterances in that particular language. A necessary step in this process would seem to be the ability to segment the input into the relevant units. Such segmentation would provide the opportunity to keep track of the distributional properties of the major phrasal units within the clause. The kinds of behaviors manifested by infants in the present study imply that the spans of speech that coincide with major phrasal units may have a certain coherence for the 9 month old, in the sense that an artificially imposed segmentation that disrupts these spans is less preferred to one that preserves their integrity. We believe that sensitivity to acoustic markers of important grammatical units can play a role in bootstrapping language acquisition. Indeed, an ability to detect acoustic markers of phrasal units could be a component of the kind of parameter setting process that Chomsky (1981) has suggested occurs in acquiring a native language (see Mazuka, 1991, for interesting suggestions about how the "branching direction" parameter could be set in this way).

Acoustic analyses of the Storybook materials in the present study indicate the presence of declinations in fundamental frequency and longer syllable durations in the vicinity of the boundary between subject and predicate phrases. As noted earlier, there have been other suggestions that lengthening of segmental durations occurs in the vicinity of the phrasal boundaries in Child-directed speech (e.g., Fisher, 1991; Lederer & Kelly, 1991). The presence of such acoustic markers in the input may help direct the child to the kind of information necessary to set the parameters correctly. Nevertheless, we want to make it absolutely clear that we are not claiming that sensitivity to acoustic markers is the sole or even the most important way of bootstrapping language acquisition. As others before us have pointed out (e.g. Bates & MacWhinney, 1987; Gleitman & Wanner, 1982; Morgan, 1986; Pinker, 1984; Slobin, 1985) there are many other types of cues that could serve the same purpose. Thus, Morgan et al. (1987) found that providing learners with appropriate prosodic markers led to more efficient learning of an artificial language by adults than when such cues were not available. However, they also showed that the presence of other kinds of markers such as articles and concord morphology could produce similar kinds of advantages. Cues such as these might also be important for infants in acquiring a native language. It would be interesting to explore if and when infants become sensitive to the kind of marking afforded by the presence of articles and concord morphology.

There is another caution that we would like to add in extrapolating from our results. It has to do with the fact that correspondence between prosodic units and grammatical units is far from perfect. It has been noted often that syntactic structure is not the sole determinant of prosodic structure (e.g., Beckman & Edwards, 1990; Cruttenden, 1986; Nespor & Vogel, 1986; Remez et al., 1986). The stylistic or affective attributes of the talker also influence the prosodic structure of the utterance (Fairbanks & Pronovost, 1939; Williams & Stevens, 1972). Thus, as Grosjean and Gee (1987, p. 141) have noted, "phonological phrases are not always constituents in syntactic structure." The extent to which the prosodic packaging in Child-directed speech may be more consistently linked to syntactic structure than is the case with Adult-directed speech is not well known at present. Regardless, our point is simply that the prosodic packaging may provide a rough estimate of where the syntactic units may lie in a language like English. It may provide the type of perceptual precategorization that allows the infant to divide up the input in a way that makes finding the syntactic categories more likely. This would not be the sole example of such precategorization in the realm of infant speech perception. The available data suggest a similar process for forerunners of the phonetic categories that are associated with a particular native language (e.g., see Aslin, Pisoni, & Jusczyk, 1983).

The present study has some interesting implications for understanding the development of infant speech perception processes. First of all, in contrast to previous work investigating perception of clausal units (e.g., Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989), there was no clear indication that 6 month olds are sensitive to acoustic correlates of phrasal units. However, by 9 months of age, this sensitivity appears to be well established. Thus, the pattern here, whereby the larger units are detected first and then sensitivity develops to subunits within these, is consistent with a differentiation view of development such as that espoused by Gibson (1969) among others. A greater familiarity with the prosodic structure of the native language may put the 9 month old in a better position to detect the cues to the segments correlated with phrasal units. Indeed, in contrast to the situation for clausal units where the markers may be more consistent from language to language, it may be that experience with a specific language is necessary for the detection of speech segments related to phrasal units to occur. Languages vary in the way that information about phrasal units is presented. English is an example of a language where word order plays a critical role in designating syntactic relations. In other languages such as Polish, word order can be free because the critical syntactic relations are conveyed by means of case endings. Because alternative forms of marking for phrasal units may exist for different languages, familiarity with the language may be necessary in order to arrive at the optimal markers for that language.

The notion that there is a developmental trend between 6 and 9 months of age in the direction of attention to specific features of the native language is intriguing in light of the findings of Werker and Tees (1984) with respect to sensitivity to foreign language phonetic contrasts. Specifically, they found that around this period of time, the infant shows a diminished ability to discriminate phonetic contrasts that do not occur in the native language. More recent research which we have conducted indicates that a similar trend occurs at an even earlier age for infants' sensitivity to markers of clausal units in a foreign language (Jusczyk, 1989; Jusczyk, Kemler Nelson, Hirsh-Pasek, & Schomberg, in preparation). Thus, 4½-month-old American infants show significant preferences for uninterrupted Polish clauses, but 6-month-old American infants do not. Both trends may stem from the same source, namely, a tendency to focus more exclusively on the sound patterns associated with the native language. We note as well, that there is evidence that speech production processes also begin to conform more closely to native language characteristics during this period (e.g., Boysson-Bardies, Halle, Sagart, and Durand, 1989; Boysson-Bardies & Vihman, 1991). Hence, speech perception capacities may be undergoing some sort of reorganization to focus more precisely on sources of information that do play an important role in signaling structural distinctions in the language and away from those acoustic cues that are not central to making such distinctions.

Certainly, the demands imposed by the kind of on-line speech processing that adults engage in are quite different in nature from those associated with actually acquiring a native language. Thus, there is no reason to assume that the cues that prove most useful for discovering phrasal units during language acquisition are those that will work best with the other sources of information that are called upon in fluent speech recognition. Hence, a fruitful area of research is to gain a much more precise picture of the kinds of cues that give rise to perceptual units both for infants and adults. The analyses of some of the materials in the present study suggest that declinations in fundamental frequency and changes in syllable durations are potential sources of information that infants could use in detecting major phrasal boundaries. Other studies have also noted increases in syllable durations just prior to such boundaries in input both to adults (e.g., Cooper & Paccia-Cooper, 1980; Klatt, 1975, 1976; Lehiste et al., 1976) and to infants (Fisher, 1991; Lederer & Kelly, 1991). Investigations about the way in which systematic manipulations of these cues, either singly or together, affect segmentation processes will help clarify their role both in fluent speech recognition and in language acquisition.

A further issue to be investigated concerns sensitivity to possible acoustic cues for minor phrase boundaries. In the present research, our focus was directed to cues that correlate with major phrasal units like the subject–noun and predicate–verb phrases. Might the language learner also be sensitive to some sort of prosodic marking corresponding to minor phrasal units such as prepositional and adjectival phrases? Language learning seems to require learning the subcategorization frames associated with the choice of a particular verb. These frames range across the phrasal types. Thus, some kind of marking of these units, prosodic or otherwise, may help to facilitate language learning. Recent reports by Fisher (1991) and Lederer and Kelly (1991) suggest that such marking may be present in Child-directed speech.

Another question that deserves consideration is the extent to which the kinds of effects that we observed here hold across different languages. Research under way in our laboratories is being directed at the issue of whether infants from English-speaking homes are sensitive to acoustic correlates of phrasal and clausal units in other languages. However, a related issue concerns the extent to which infants from other language groups will show similar kinds of effects with their own native languages to the ones reported here. Such information is vital to understanding the potential role that sensitivity to acoustic correlates of grammatical units may play in language acquisition. Direct evidence concerning the perception of cues to clausal or phrasal units by foreign infants is not presently available. However, recent work by Mehler et al. (1988) indicates that newborn infants are able to discriminate utterances in their native language from those in a foreign language. This attentiveness on the part of newborns to the sound properties of native language utterances fits well with the view that sensitivity to cues in the acoustic input may be one of the important routes that infants take in discovering the set of grammatical relations that hold for utterances in their language.

In summary, the present study indicates that, by 9 months of age, American infants are sensitive to acoustic correlates of major phrasal units that are available in the prosody of English utterances. Acoustic analyses of some of the stimulus materials suggest that there are potential prosodic markers in the input in the form of declinations in fundamental frequency and in longer durations of the syllable immediately preceding a major phrasal boundary. Sensitivity to these and other potential acoustic markers of phrasal units could assist infants in language acquisition by segmenting fluent speech into meaningful linguistic units. Of course, 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.

REFERENCES

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In M. Haith & J. Campos (Eds.), *Handbook of child psychology, vol. 2, Infancy and developmental psychobiology* (pp. 573–687). Wiley: New York.

Bates, E., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Erlbaum.

Beckman, M., & Edwards, J. (1990). Lengthening and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 152–178). Cambridge: Cambridge University Press.

Bernstein-Ratner, N. (1985, November). *Cues which mark clause-boundaries in mother–child speech.* Paper presented at ASHA Convention.

Best, C. T., McRoberts, G. W., & Sithole, N. (1988). Examination of the perceptual reorganization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance,* **14**, 245–260.

Bolinger, D. L. (1978). Intonation across languages. In J. P. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language, vol 2: Phonology*. Stanford: Stanford University Press.

Boysson-Bardies, B. de, Halle, P., Sagart, L., & Durand, C. (1989). A cross-linguistic investigation of vowel formants in babbling. *Journal of Child Language,* **16**, 1–17.

Boysson-Bardies, B. de, & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language,* **67**, 297–319.

Broen, P. (1972). The verbal environment of the language learning child. *ASHA Monography,* **17**.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1986). Barriers. *Linguistic Inquiry Monograph Series*. Cambridge, MA: MIT Press.

Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition,* **25**, 53–69.

Cicchetti, D. V. (1972). Extension of multiple-range tests to interaction tables in the analysis of variance: A rapid approximate solution. *Psychological Bulletin,* **77**, 405–408.

Collier, R., & t'Hart, J. (1975). The role of intonation in speech perception. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 107–121). Heidelberg: Springer-Verlag.

Cooper, R. P. & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development,* **61**, 1584–1595.

Cooper, W. E. & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics,* **18**, 341–347.

Eimas, P. D., Miller, J. L., & Jusczyk, P. W. (1987). On infant speech perception and the acquisition of language. In S. Harnad (Ed.) *Categorical perception: The groundwork of cognition*. Cambridge: Cambridge University Press.

Fairbanks, G. & Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotions. *Speech monographs,* **6**, 87–194.

Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, & R. Golinkoff (Eds.), *The origins and growth of communication* (pp. 5–29). Norwood, NJ: Ablex.

Fernald, A. (1985). Four-month-olds prefer to listen to motherese. *Infant Behavior and Development,* **8,** 181–195.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development,* **60,** 1497–1510.

Fernald, A., & Kuhl. P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development,* **10,** 279–293.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology,* **20,** 104–113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. de, & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language,* **16,** 477–501.

Fisher, C. (1991). Prosodic cues to phrase structure in infant directed speech. *Papers & Reports on Child Language Development,* **30,** Stanford, CA.

Frazier, L. (1987). Structure in auditory word recognition. *Cognition,* **25,** 157–187.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton.

Givon, T. (1979). *On understanding grammar.* New York: Academic Press.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition,* **1,** 1–55.

Gleitman, L. R., Gleitman, H., Landau, B., & Wanner, E. (1988). Where the learning begins: Initial representations for language learning. In F. Newmeyer (Ed.), *The Cambridge linguistic survey.* Cambridge: Cambridge University Press.

Gleitman, L. R., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge: Cambridge University Press.

Greiser, D., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology,* **24,** 14–20.

Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition,* **25,** 135–155.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition,* **26,** 269–286.

Hockett, C. A. (1954). Two models of grammatical description. *Word,* **10,** 210–231.

Jusczyk, P. W. (1985). On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion* (pp. 199–229). Hillsdale, NJ: Erlbaum.

Jusczyk, P. W. (1986). Toward a model of the development of speech perception. In J. S. Perkell & D. H. Klatt (Eds.) *Invariance and variability in speech processes.* Hillsdale, NJ: Erlbaum.

Jusczyk, P. W. (1989). *Perception of cues to clausal units in native and non-native languages.* Paper presented at the biennial meeting of the Society for Research in Child Development, Kansas City, April.

Jusczyk, P. W. (1992). Developing phonological categories form the speech signal. In C. A. Ferguson, C. Stoel-Gammon, & L. Menn (Eds.) *Phonological development: Models, research, implications.* Parkton, MD: York Press.

Jusczyk, P. W., Kemler Nelson, D. G., Hirsh-Pasek , K., & Schomberg, T. (in preparation). Perception of acoustic correlates to clausal units in a foreign language by American infants.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, **16**, 53–68.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics*, **3**, 129–140.

Klatt, D. H. (1976). Linguistic uses of segment duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208–1221.

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science*, **1**, 70–73.

Lederer, A., & Kelly, M. (1991). Prosodic correlates to the adjunct/complement distinction in motherese. *Papers & Reports on Child Language Development*, **30**, Stanford, CA.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, **7**, 107–122.

Lehiste, I., Olive, J. P., & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, **60**, 1199–1202.

Lieberman, P. (1961). Perturbations in vocal pitch. *Journal of the Acoustical Society of America*, **33**, 597–603.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, **6**, 172–179.

Lieberman, P. (1967). *Intonation, perception and language*. Cambridge, MA: MIT Press.

Luce, P. A., & Charles-Luce, J. (1983, May). *Contextual effects on the consonant/vowel ratio in speech production*. Paper presented at the 105th meeting of the Acoustical Society of America, Cincinnati.

Mazuka, R. (1991). How can a grammatical parameter be set before the first word. *Papers & Reports on Child Language Development*, **30**, Stanford, CA.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, **29**, 143–178.

Morgan, J. L. (1986). *From simple input to complex grammar*. Cambridge, MA: MIT Press.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of *Cognitive Psychology*, **19**, 498–550.

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, **62**, 714–719.

Nakatani, L. H., & Schaffer, J. A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, **63**, 234–245.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Osherson, D. N., Stob, M., & Weinstein, S. (1984). Learning theory and natural language. *Cognition*, **17**, 1–28.

Pilon, R. (1981). Segmentation of speech in a foreign language. *Journal of Psycholinguistic Research*, **10**, 113–121.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, **90**, 2956–2970.

Read, C., & Schreiber, P. (1982). Why short subjects are hard to find. In E. Wanner & L. R. Gleitman (Eds.) *Language acquisition: The state of the art*. Cambridge: Cambridge University Press.

Remez, R. E., Rubin, P. E., & Ball, S. A. (1985). Sentence intonation in spontaneous utterances and fluently spoken test. *Journal of the Acoustical Society of America*, **77**, S38.

Remez, R. E., Rubin, P. E., & Nygaard, L. (1986). On spontaneous speech and fluently spoken text: Production differences and perceptual distinctions. *Journal of the Acoustical Society of America,* **79,** S26.

Remick, H. W. (1971). *The maternal environment of language acquisition.* Unpublished doctoral dissertation, University of California, Davis.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America,* **71,** 996–1007.

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior,* **23,** 450–466.

Slobin, D. I. (1985). Why study language crosslinguistically? In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 1, pp. 3–24). Hillsdale, NJ: Erlbaum.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America,* **64,** 1582–1592.

Vaissiere, J. (1981). Speech recognition programs as models of speech perception. In T. Myers, J. Laver, & J. Anderson (Eds.) *The cognitive representation of speech.* Amsterdam: North Holland.

Wakefield, J. R., Doughtie, E. B., & Yom, L. (1974). Identification of structural components of an unknown language. *Journal of Psycholinguistic Research,* **3,** 262–269.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology,* **24,** 672–683.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology,* **43,** 230–246.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,* **7,** 49–63.

Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition.* Cambridge, MA: MIT Press.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America,* **52,** 233–248.

(Accepted November 14, 1991)