J. Child Lang. 23 (1996), 1–30. Copyright © 1996 Cambridge University Press

Infants' sensitivity to word boundaries in fluent speech*

JAMES MYERS

PETER W. JUSCZYK
State University of New York at Buffalo

DEBORAH G. KEMLER NELSON
Swarthmore College

JAN CHARLES-LUCE
State University of New York at Buffalo

AMANDA L. WOODWARD
The University of Chicago

AND

KATHRYN HIRSH-PASEK
Temple University

(*Received* 16 *September* 1994. *Revised* 24 *March* 1995)

ABSTRACT

Infants' sensitivity to word units in fluent speech was examined by inserting 1 sec pauses either at boundaries between successive words (Coincident versions) or between syllables within words (Noncoincident versions). In Experiment 1, 24 11-month-olds listened significantly longer to the Coincident versions. In Experiment 2, 24 four-and-a-half-

---

[*] The research described in this paper was supported by a research grant from N.I.C.H.D. (HD 15795) to P.W.J. and N.I.D.C.D. (DC 00957) to J.C.L. In addition, J.M. was supported by a Training Grant from N.I.D.C.D. (DC 00036). We thank Lori Kennedy, Nan Koenig, Ann Marie Jusczyk, Tracy Schomberg, Alice Turk and Nancy Redanz for their help in testing the infants, and Eric Bylund and Stephen Yeoh for their assistance in programming. We are also grateful to Paul Luce for his help in computing the phonotactic probabilities reported in Experiment 5. Portions of the research were previously described at the Biennial Meeting of the Society for Research in Child Development, held in Kansas City, MO and at the 127th Meeting of the Acoustical Society for America, held in Cambridge, MA. Address for correspondence: Peter W. Jusczyk, Department of Psychology, State University of New York at Buffalo, Park Hall, Buffalo, NY 14260, USA.

and 24 nine-month-olds did not exhibit the preference for the Coincident versions that the 11-month-olds showed. When the stimuli were low-pass filtered in Experiment 3, 24 11-month-olds showed no preference for the Coincident versions, suggesting they rely on more than prosodic cues. New stimulus materials in Experiment 4 indicated that responses by 24 11-month-olds to the Coincident and Noncoindent versions did not depend solely on prior familiarity with the targets. Two groups of 30 11-month-olds tested in Experiment 5 were as sensitive to boundaries for Strong/Weak words as for Weak/Strong words. Taken together, the results suggest that, by 11 months, infants are sensitive to word boundaries in fluent speech, and that this sensitivity depends on more than just prosodic information or prior knowledge of the words.

INTRODUCTION

First language acquisition requires not only the internalization of a grammar, but also the memorization of thousands of lexical items. For this latter task there is nothing for an infant to do but to pay attention to the words, figure out what they mean and how they are pronounced, and commit them to memory. First of all, though, the infant must find them. Adults do not usually speak in isolated words, but in sentences and paragraphs – fluent utterances where the boundaries between words are often hard to find. This problem of segmentation is one of the most fundamental facing an infant acquiring a language, its profound difficulty matched by its enormous importance for the successful acquisition of a grammar. Without words, the rules of syntax will have nothing to manipulate, and the rules of phonology and morphology will have nothing to generalize over. Consequently, infants must become master speech-to-word converters.

As with most engineering marvels developed by evolution, in their skill at this task infants are still far ahead of their artificial competitors. No algorithm for solving the segmentation problem has yet been implemented successfully in a computer (Marcus, 1984; Waibel, 1986). The reason for this is not difficult to see: there are no completely reliable physical cues to the presence of word boundaries in fluent speech. The neat white spaces separating words on the page simply do not exist in spoken language (Liberman & Studert-Kennedy, 1978; Cole & Jakimik, 1980; Klatt, 1980).

Infant-directed speech does make the word segmentation task somewhat easier, in that caregivers often place emphasis on new words (Kaye, 1976; Messer, 1981); exaggerate their prosody or put new words in sentence-final positions (Ryan, 1978; Woodward & Aslin, 1990); speak more slowly (Stern, Spieker, Barnett & Mackain, 1983); and lengthen the duration of certain content words (Garnica, 1977). Nevertheless, infant-directed speech still primarily consists of fluent utterances, even when subjects are explicitly told to teach infants new words (Woodward & Aslin, 1990). Thus to a large degree, infants must still solve the segmentation problem on their own.

Evidence for speech segmentation by infants

The issue of how and when infants begin to segment the speech stream into linguistically relevant units is one that has received considerable attention recently. The primary question investigated has been one associated with the view known as PROSODIC BOOTSTRAPPING (Gleitman & Wanner, 1982; Peters, 1983; Morgan, 1986). This view suggests that learners may derive useful information from the speech signal about the location of linguistically relevant units in the input. Accordingly, a number of studies have focused on whether infants can use suprasegmental cues, such as stress, pitch contour and vowel lengthening, to locate the boundaries of the prosodic constituents associated with clauses and phrases. For example, Hirsh-Pasek, Kemler Nelson, Jusczyk, Wright Cassidy, Druss & Kennedy (1987) examined whether clause-sized constituents are perceptual units for infants. Their technique involved inserting pauses either at clause boundaries or within clauses. The hypothesis was that if infants perceived clauses as units, then they would prefer listening to passages with pauses at the clause boundaries as opposed to ones with pauses in the middle of clausal units. This hypothesis was confirmed by Hirsh-Pasek *et al.* (see also Kemler Nelson, Hirsh-Pasek, Jusczyk & Wright-Cassidy, 1989) for infants as young as seven months of age and subsequently, for four-and-a-half-month-olds (Jusczyk, 1989). Similarly, older infants are also responsive to potential units within clauses, such as those associated with prosodic phrases. However, in contrast to clausal units, infants appear to require more experience with native language input before they respond to subclausal units. Thus, nine-month-olds, but not six-month-olds, show a preference for passages where pauses are inserted at phrase boundaries over passages where pauses are inserted within phrases (Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward & Piwoz, 1992; Gerken, Jusczyk & Mandel, 1994).

Possible bases for learning to segment words from fluent speech

Given that infants are sensitive to the prosodic marking of linguistically relevant units such as clauses and phrases, one can ask whether prosodic information plays some role in the segmentation of constituents even smaller than the phrase. Indeed, there are some indications that prosodic cues could be helpful in word segmentation. For instance, Cutler and her colleagues (e.g. Cutler & Norris, 1988; Cutler & Butterfield, 1992) have suggested that adult English-speakers could use prosodic information to help locate word boundaries in fluent speech. Cutler & Carter (1987) noted that a very high proportion of words in fluent conversational English speech are either stressed monosyllables or begin with a stressed syllable. Cutler & Norris

(1988) suggested that English speakers could make a reasonable first pass at word segmentation by assuming that each strong syllable marks the onset of a new word. This view is supported by a study that shows that listeners are more apt to misparse words with Weak/Strong (WS) stress patterns than ones with Strong/Weak (SW) patterns (Cutler & Butterfield, 1992).

There are indications that infants are at least sensitive to the kind of prosodic cues that could be used to locate word boundaries in English utterances. In particular, Jusczyk, Cutler & Redanz (1993*a*) found that nine-month-old American infants prefer listening to lists of English words with SW, as opposed to WS, stress patterns. Moreover, the preference for the SW words held even when the words were low-pass filtered, suggesting that infants were responding to the prosodic, rather than segmental, properties of these words. Using a different technique, Morgan (1994) investigated the role of prosodic constituents in the perception of strings of nonsense syllables by eight-month-olds. He found that rhythmic stress patterns appeared to be used by the infants in constructing higher-level constituents out of the syllables, as indicated by their differing responses to targets placed either within or between bisyllabic units with a strong/weak stress pattern. Morgan speculated that infants could use the bias toward grouping rhythmically stressed syllables into feet as part of a strategy for finding linguistically significant constituents that are approximately foot-sized, namely words. Nevertheless, English does have many words which do not begin with strong syllables (e.g. *guitar*, *ascent*, *discover*, etc.). Consequently, any word segmentation strategy based exclusively on prosody will sometimes misparse the input – as the data of Cutler & Butterfield (1992) demonstrate.

Of course, prosody is not the only potential source of information that could aid in the segmentation of words from fluent speech. There are other kinds of information that can be used in heuristic strategies for locating word boundaries. To take an example of an unusually reliable cue, in English the phoneme /t/ is almost always aspirated at the beginning of a word, as in *tolerable* and *Toledo*, whereas word-internal instances of /t/ tend to be unaspirated, as in *stage* and *comet*. Church (1987) has suggested that if listeners are sensitive to the contexts in which certain allophones typically appear, they could use this information to facilitate word segmentation. Note that Brown, Cazden & Bellugi (1969) speculated that language learners might use such information to segment the speech stream into the appropriate word units. More recently, Hohne & Jusczyk (1994) found that two-month-olds can at least discriminate the kinds of allophonic differences that could signal the presence or absence of word boundaries.

Although allophonic cues could help infants in locating word boundaries, they could also be misleading in some cases. For instance, the presence of an aspirated /t/ does not always predict a word boundary because an aspirated /t/ can sometimes occur word-internally, as in *atomic*. Similarly, unaspirated

/t/ does not always occur word-internally, since in fluent speech closed class items like *to*, *today*, and *tomorrow* often begin with an unaspirated /t/ (as in *I'll see you tomorrow*). Moreover, an unaspirated /t/ is acoustically very similar to a /d/, which may appear in any position in a word.

Both types of word boundary cues considered thus far (i.e. syllable stress and allophonic constraints) are also limited by an additional factor: they do not correlate with words so much as with prosodic units such as syllables and metrical feet – i.e. constituents consisting in English of a stressed or strong syllable optionally followed by an unstressed or weak syllable (Hayes, 1981; Halle & Vergnaud, 1987). This is clearly true of stress patterns, which by definition mark foot boundaries. Allophonic cues are not necessarily any more helpful in this respect, since most of them also mark syllables and feet, not words (Nespor & Vogel, 1986; Church, 1987). Word-internal aspiration in English, for example, marks the beginning of a metrical foot. This is why an alternation in aspiration co-occurs with the change in stress in word pairs like *atomic* and *atom*.[1]

A different cue that tends to mark words, and not simply prosodic units, is phonotactics (constraints on the possible orders of phonetic segments in native language words). For example, in English, nasal-stop sequences tend to agree in place, even if they appear in different syllables. Thus the sequence /nt/ occurs within words very often, while sequences like /mt/ do not. A listener who discovered a /mt/ sequence within fluent speech might reasonably hypothesize the presence of a word boundary. Furthermore, there is evidence that, by nine months, infants are also sensitive to how phonotactic patterns are distributed in the input. For example, infants at this age are capable of distinguishing not only legal and illegal sequences of phonemes (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud & Jusczyk, 1993*b*) but also forms with phoneme transitions of high probability and those where phoneme transitions are improbable (Jusczyk, Luce & Charles Luce, 1994). This view that language learners may use a distributional analysis of phonotactic information to acquire words from fluent speech is also at the heart of a recent proposal based on a computer simulation model (Brent, Cartwright & Gafos, 1994). Nevertheless, even phonotactic cues cannot be relied on to the exclusion of all else. Thus, although sequences like /mt/ rarely occur word-internally, sequences like /nt/ may well occur across word boundaries purely by accident.

Finally, another proposal about how infants segment speech into words is that they first learn to recognize some words by hearing them spoken in

---

[1] Of course, it is interesting that words such as *atomic* (for which it is possible to interpret the initial syllable as the functor word *a*) are precisely the kinds of words that young children (Holden & MacGinitie, 1972; Tunmer, Bowey & Grieve, 1983; Chaney, 1989), and even adults (Cutler & Butterfield, 1992) are most likely to misperceive. Consider the favourite example of Gleitman & Wanner (1982), 'is it *an adult* or *a nuhdult*?'

isolation and then matching representations of these to the input (Pinker, 1984; Suomi, 1993). In contrast to the earlier strategies considered, this is basically a top-down approach to segmentation. The recognition of one word in a string helps to demarcate the end of a preceding word and the beginning of a succeeding word. As the learner recognizes more words, the amount of unfamiliar information in an utterance will tend to decrease. Moreover, if some unfamiliar patterns recur often enough in different contexts (e.g. *the*), the learner may eventually recognize these as potential lexical items (see Brown *et al.* 1969 for a similar suggestion). However, this approach is also not without its difficulties. First, the acoustic characteristics of a word spoken in sentential context may differ considerably from one spoken in isolation (Klatt, 1980). For example, a word spoken in a sentential context is apt to differ considerably in its overall duration from the same word spoken in isolation. Second, sound patterns associated with monosyllabic words (e.g. *can*) may occur as parts of larger words (e.g. *cancer*, *toucan*, *uncanny*, etc.). Thus, the listener who matches *can* to any of these words ends up misparsing the signal. Consequently, even if some word learning does occur by first recognizing the sound pattern of a particular word in isolation, other routes to learning words must also be available.[2]

The upshot of all of this is that there are many potential sources of information that a listener can draw on for locating word boundaries in fluent speech. However, none of these will fully guarantee a correct segmentation of the input into words. Thus, it is likely that mature listeners draw upon some combination of these sources of information for word segmentation. Furthermore, the utility of these sources depends critically on the nature and organization of the sound structure of a particular language. For example, whether word stress is fixed or not, and if so, which syllable receives stress varies across languages, and so do the phonotactics and the allophonic constraints. Consequently, we anticipate that language learners will require some experience and practice to discover the optimal way to combine these sources for their native language (much as Bates & MacWhinney, 1987, have suggested for other aspects of language acquisition).

When might infants begin to respond to word units in fluent speech?

There is now considerable evidence that infants learn a great deal about the nature and organization of sound patterns in their native language during their first year of life (Werker & Tees, 1984; Kuhl, Williams, Lacerda,

---

[2] The other alternatives considered (i.e. prosodic, allophonic, and phonotactic) also must assume that the infant occasionally has access to words in isolation. For instance, to know that words typically begin with strong syllables, or that some allophone occurs only in initial position, requires exposure to some examples with these patterns. In the case of stress patterns for English words, many items likely to be heard by infants fit this pattern (e.g. *mommy*, *daddy*, *baby*, etc.).

Stevens & Lindblom, 1992; Best, 1993; Jusczyk *et al.* 1993*a*, *b*; Polka & Werker, 1994). Moreover, as noted earlier, there is evidence that infants are sensitive to linguistically relevant units such as clauses at least by four-and-a-half-months of age, and to subclausal units such as prosodic phrases at nine months of age (Jusczyk *et al.* 1992; Gerken *et al.* 1994). There is also evidence that between eight and ten months, infants begin to comprehend at least some words (Huttenlocher, 1974; Benedict, 1979).

These findings raise interesting questions about when infants begin to perceive words as units in fluent speech. One possibility is that infants respond to word boundaries long before they show any evidence of producing or understanding words. This possibility follows from the fact that when infants begin to speak, they use a vocabulary of words extracted from fluent adult speech at some earlier stage of development. For example, it is conceivable that infants make a first pass at segmenting speech into word-like units using some general acoustic characteristics, such as abrupt onsets, rhythmic characteristics, etc. If so, then the ability to detect boundaries of words in fluent speech might develop quite early, just as a host of other basic speech perception capacities do (Eimas, 1982; Aslin, Pisoni & Jusczyk, 1983; Kuhl, 1987). However, word forms do vary considerably across languages. Thus, the segmentation of words in a particular language may take on a language-specific character. Consequently, segmenting words from fluent speech may depend on infants' knowledge of the organization of the sound structure of the language that they are acquiring. Previous investigations regarding infants' responsiveness to the possible units and organization of native language utterances indicate that sensitivity to subclausal units such as phrases is not evident prior to nine-months of age (Jusczyk *et al.* 1992). Indeed, one interpretation of these findings (Kemler Nelson, 1989) is that phrases are recognized somewhat later than clauses, precisely because they are smaller than clauses – implying that infants work from larger to smaller constituents.

In the following studies we addressed these hypotheses by adopting the pause-insertion method of Hirsh-Pasek *et al.* (1987). We reasoned that if infants perceived words as units, they would prefer to attend to passages where words were not interrupted with inserted pauses. Additional experiments examined: (*a*) when sensitivity to word units in fluent speech develops; (*b*) the role of prior familiarity with words; and (*c*) whether sensitivity to word units depends primarily on prosodic information, as sensitivity to larger constituents seems to.

EXPERIMENT 1

In this first experiment, we examined whether 11-month-old English-learning infants display some sensitivity to word structure in fluent speech.

As in previous investigations (e.g. Hirsh-Pasek *et al.* 1987; Kemler Nelson *et al.* 1989), we used a pause insertion technique to explore infants' sensitivity to linguistically relevant units (i.e. words) in fluent speech. Specifically, we inserted a series of 1 sec pauses into matched passages of fluent speech at locations corresponding to boundaries between two successive words or at boundaries between two syllables within a particular word. We hypothesized that if infants are sensitive to word units in fluent speech, they would listen significantly longer to the passages containing pauses at word boundaries.

METHOD

Subjects

Twenty-four infants approximately 11 months of age (mean age: 45 weeks, 5 days; range: 44 weeks, 4 days to 46 weeks 6 days) from monolingual English-speaking homes in the Eugene, Oregon, area were tested. An additional 11 infants were tested but failed to complete the procedure because of failing to look at one of the two sides for at least three trials (six), or because of crying or inattentiveness (five).

Stimuli

We decided not to use spontaneous speech because the task of inserting pauses within words without creating acoustic transients is difficult even with carefully chosen stimuli. The samples for the present experiment were derived from the original recordings made of the 'Storybook' samples prepared for Experiment 3 in Jusczyk *et al.* (1992). These samples involved a woman reading a series of short passages aloud to a two-year-old child. The woman did not know the purpose of the research. She had been instructed to read the passages in a lively manner to the child.

The tape recording of the mother's speech 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. All pauses longer than 400 msec were removed from each sequence by use of an auditory editing program. Two versions of each sequence were prepared by inserting 1 sec pauses at different locations in the utterances. Pauses were inserted only at zero crossings in amplitudes so as not to produce transients in the signal. One version of each sample, the 'Coincident version', was modified by inserting the pauses between words. The other version, the 'Noncoincident version', was modified by inserting the pauses between syllables of the same word.[3] Every effort was made to equate the Coincident and Noncoincident versions of the samples with respect to the stress patterns

---

[3] In some Coincident samples, the pause was inserted at minor phrase boundaries, such as before a prepositional phrase, but there was no indication that this made a difference in infants' responding.

surrounding the pauses. Consequently, if the first pause in a particular Coincident sample came after a stressed syllable and before an unstressed syllable, the comparable pause in the matched Noncoincident sample also came after a stressed syllable and before an unstressed syllable. Equal numbers of pauses were inserted in the Coincident and Noncoincident versions of each sample. Half of the pauses in the Coincident samples occurred just prior to the word targeted for pause insertion in the Noncoincident samples; the other half occurred just after the targeted word.

In every way, except for the placement of the pauses, the Coincident and Noncoincident versions of each sample were identical. Four of the 16 pairs of samples were used in the familiarization trials; the remaining 12 pairs were used for the test trials. Each subject heard a different random ordering of the 12 test items. For the samples presented on the test trials, the mean duration was 27·55 sec (range: 19·93–38·44 sec). An example of the Coincident (A) and Noncoincident (B) versions of a sample (slashes indicate inserted pauses) is the following:

> (A): Some very / big animals live at the zoo. These very / big animals are elephants. (Oh boy!) The smaller baby / elephant is walking with her mother. Both / the mother and the / baby have long trunks instead of noses. The baby / elephant / and her mother are looking for food. Plants, grain and / peanuts are good / food for elephants.
>
> (B): Some very big ani/mals live at the zoo. These very big animals are ele/phants. (Oh boy!) The smaller baby ele/phant is walking with her mother. Both the mo/ther and the baby have long trunks in/stead of noses. The baby ele/phant and her mo/ther are looking for food. Plants, grain and pea/nuts are good food for ele/phants.

Apparatus

The experiment was conducted in a three-sided test booth constructed out of pegboard, with panels of 4 ft by 6 ft on three sides and open at the back. This made it possible for an observer to look through one of the existing holes to monitor the infant's headturns. Most of the pegboard was backed with white cardboard to guard against the possibility that the infant might respond to movements behind the panel. However, there was a small area about 30 cm^2 without backing behind the centre panel to permit the experimenter to observe the infant. Each of the side panels also had a red light and a loudspeaker mounted at the infant's eye level (when seated on the caregiver's lap). A white curtain suspended around the top of the booth shielded the infant's view of the rest of the room. A computer terminal and response box were located behind the centre panel, 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 centre and side lights, recorded the direction and duration of headturns, and terminated a trial when the infant looked away for more than 2 sec. Information about the direction and duration of headturns and the total trial duration were 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 (KA 5700) audio amplifier to the 7-inch loudspeakers mounted on the side walls of the test chamber.

Design and procedure

The procedure used was the version of the Headturn Preference Procedure used by Jusczyk *et al.* (1992). Each infant completed an eight-trial pre-exposure phase (both versions of four pre-trial 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-trail pre-exposure 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. To enhance the reliability of the durational measures, we analysed the data only from those subjects who oriented on at least three trials to each type of version, Coincident and Noncoincident. (For further discussion of this point, see Jusczyk *et al.* 1992.)

Each trial began by blinking the green light to attract the infant's attention to centre. When the infant was oriented at midline, the centre green light was extinguished and the red light above one (familiarization trials) or both (test trials) of the loudspeakers 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 infant made a head-turn of at least 30° in the direction of one of the loudspeakers, the sample appropriate to that side began to play and continued until its completion or until the infant failed to maintain the 30° headturn for 2 consecutive sec (e.g. if the infant turned back to the centre or the other side, looked at the caregiver, the floor or the ceiling). If the infant turned briefly away from the target by 30° in any direction, but for less than 2 sec, and then looked back again, the time spent looking away was not included in the orientation time. Whether the first heard sample was Coincident or Noncoincident and whether it occurred on the right or the left were counterbalanced across infants.

During the familiarization 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 started the excerpt appropriate to that side and also terminated the blinking light on the other side. Across the 12 test trials, no sample was ever repeated.

An experimenter behind the centre panel looked through a peephole and recorded the direction and duration of the infant's headturns using a response box. The experimenter did not know which loudspeakers played the Coincident and Noncoincident versions of the samples. This was possible because the assignment of the versions to the left or right side was determined by the computer and not revealed to the experimenter until the completion of the test session. 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. In addition, both the experimenter and the infant's caregiver listened over headphones to recordings of loud music to prevent them from hearing the speech samples. (See Kemler Nelson, Jusczyk, Mandel, Myers, Turk & Gerken, 1995, for a discussion of the effectiveness of this masking procedure.)

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 was 9·79 sec (S.D. = 6·14 sec) for the Coincident versions and 6·99 sec (S.D. = 4·13 sec) for the Noncoincident versions. Separate analyses with subjects ($t(23) = 2{\cdot}86$, $p < 0{\cdot}01$) and samples ($t(11) = 2{\cdot}20$, $p < 0{\cdot}05$) confirmed that this difference is significant.

The present results suggest that, by 11 months, English-learning infants are sensitive to word units in fluent speech in that infants listened significantly longer to passages in which pauses coincided with word boundaries than to ones in which pauses occurred between syllables within words. These findings parallel earlier ones indicating that English-learning infants are sensitive to markers of clausal and phrasal units (Hirsh-Pasek *et al.* 1987; Kemler Nelson *et al.* 1989; Jusczyk *et al.* 1992; Gerken *et al.* 1994).

One interesting pattern observed in the earlier investigations with clausal and phrasal units was the apparent existence of a developmental trend in sensitivity with respect to a hierarchy of linguistic units in the input. Thus, studies investigating clausal units have found that infants, as young as four-and-a-half months, listen significantly longer to passages with pauses inserted at clausal boundaries than ones with pauses located within clauses

(Jusczyk, 1989). Indeed, the organization afforded by the prosodic marking of clausal units has been shown to confer an advantage in remembering speech information for infants as young as two-months-old (Mandel, Jusczyk & Kemler Nelson, 1994). By comparison, studies with phrasal units suggest that English-learning infants are not sensitive to such units until around nine months (Jusczyk *et al.* 1992). In the present study, 11-month-olds were shown to be sensitive to the presence, in fluent speech, of even smaller units – namely, words. It is tempting to view the present results as further evidence for a developmental trend towards differentiating increasingly smaller units in fluent speech. To explore this possibility further, we conducted the following experiment.

EXPERIMENT 2

Recent studies have documented that between the ages of six and nine months, infants begin to display sensitivity to the organization of sound patterns in their native language. This is evident in how they respond to a number of different properties of spoken language. For example, it has been observed that the ability to discriminate certain non-native speech contrasts begins to decline during this period (Werker & Tees, 1984; Kuhl *et al.* 1992; Best, 1993; Polka & Werker, 1994). In addition, by nine months, infants display signs of distinguishing between native and non-native language words (Jusczyk *et al.* 1993*b*). Moreover, they appear to be sensitive with respect to the frequency with which certain types of syllables (Jusczyk *et al.* 1994) and stress patterns (Jusczyk *et al.* 1993*a*) appear within native language words. These facts suggest that infants are learning a great deal about native language sound patterns during this period. Hence, it would not be entirely surprising if sensitivity to word units in speech were to arise at this time. Yet, there is evidence to suggest that infants between birth and two months are sensitive to the kind of cues that could potentially mark word boundaries in their native language (Christophe, Dupoux, Bertoncini & Mehler, 1994; Hohne & Jusczyk, 1994). Thus, it is conceivable that infants could begin learning about word units in fluent speech at a very early age. To gain a clearer understanding of when sensitivity to word units in fluent speech actually develops, we tested groups of four-and-a-half- and nine-month-olds on the samples used in Experiment 1.

METHOD

Subjects

Twenty-four nine-month-olds (mean age: 39 weeks, 2 days; range: 36 weeks, 2 days to 42 weeks, 2 days) and 24 four-and-a-half-month-olds (mean age: 19 weeks, 5 days; range: 18 weeks, 6 days to 21 weeks, 2 days) from monolingual

English-speaking homes in the Eugene, Oregon, area were tested. An additional six four-and-a-half-month-olds were tested but failed to complete the procedure because of crying (five) and inadequate head control (one). Three additional nine-month-olds were also tested but not included because of crying (one), sibling interference (one), or failure to complete the full set of test trials (one).

Stimuli and apparatus

These were the same as in the previous experiment.

Design and procedure

One change was made in the procedure. This was to eliminate the infant's choice of samples on the familiarization and test trials. This change was made for two reasons. First, in all of our investigations, the duration of orientation time measure has proved to be much more sensitive than any measure based on which side the infant chose for a given trial. Second, eliminating the infant's choice also removed a factor contributing to subject drop-out rates, i.e. the failure to look to one of the two sides on at least three trials. Moreover, it allowed us to improve further the sensitivity of our dependent measure by collecting data on how long infants oriented to each version of a particular sample pair.

The modification of the procedure involved putting the selection of which version of a particular sample (i.e. Coincident or Noncoincident) played on a given trial under computer control. Each infant heard eight pairs of samples (two during the familiarization trials and six during the test trials) in a given test session. Which version of a sample pair (Coincident or Noncoincident) occurred first was randomly determined with the restriction that the two versions of a given sample did not occur on successive trials. Different random orderings of the sample pairs were used for each subject. In addition, half of the subjects heard six of the sample pairs during the test trials, and the other half heard the remaining six pairs. As in Experiment 1, the Coincident versions of the samples all occurred on the same side, and the Noncoincident version on the opposite side. The assignment of Coincident versions to either the left or right side was counterbalanced across subjects.

Each trial began by blinking the green light on the centre panel until the infant had oriented in that direction. Then, the centre light was extinguished and the red light above the loudspeaker on one of the side panels began to flash. When the infant made a head-turn of at least 30° in the direction of the loudspeaker, the next sample appropriate to that side began to play and continued to completion or until the infant failed to maintain the 30° headturn for 2 consecutive sec (e.g. if the infant turned back to the centre or the other side, looked at the caregiver, the floor or the ceiling). If the infant turned briefly away from the target by 30° in any direction, but for less than 2 sec, and then looked back again, the time spent looking away was not included in the orientation time. During familiarization trials, the blinking red light was extinguished as soon as the infant oriented to the side and the list began to play. However, during test trials, the blinking light remained on until the trial ended. Extensive pilot testing convinced us that this was the best way to handle the lights during the procedure. Leaving the flashing light on during the familiarization trials seemed to habituate the infants to the lights, and resulted in very short orientation times during the test trials. Moreover, infants were also less likely to complete the full set of test trials under these circumstances. In all other respects, the procedure was identical to that described for Experiment 1.

RESULTS AND DISCUSSION

As in the previous experiment, the data from the 12 test trials were used to calculate the mean length of orientation to each type of sample. For the four-and-a-half-month-olds, the mean length of orientation across subjects was 9·24 sec (S.D. = 4·26 sec) for the Coincident versions and 9·27 sec (S.D. = 3·55 sec) for the Noncoincident versions. Separate analyses with subjects ($t(23) = 0{\cdot}38$) and samples ($t(11) = 0{\cdot}25$) indicated that orientation times to the two versions of the samples did not differ significantly. The results for the nine-month-olds were similar. The mean length of orientation was 7·80 sec (S.D. = 3·64 sec) for the Coincident versions and 7·89 sec (S.D. = 4·44 sec) for the Noncoincident versions. Once again, separate analyses with subjects ($t\,(23) = 0{\cdot}13$) and samples ($t(11) = 0{\cdot}09$) revealed no evidence of significant differences in orientation times to the two versions of the samples.

Hence, in contrast to the 11-month-olds, there was no indication that either the four-and-a-half- or the nine-month-olds were sensitive to the presence of word units in fluent speech. However, in addition to the difference in age of the subjects in the present study, the procedure had also changed slightly. Is it possible that the procedural change, and not the age difference, was responsible for different patterns of results in the two experiments?

To examine this matter further, we contacted the parents of the nine-month-olds in the present experiment and asked them if they would bring their infants back to the laboratory at 11 months of age. The infants who were tested on one set of six sample pairs at nine months were tested on the other set of six sample pairs at 11 months. Of the original 24 subjects, 15 returned to the lab and successfully completed testing at 11 months. We replaced the other nine subjects with nine new nine-month-olds, who also returned for testing at 11 months. At nine months, the 24 infants in this longitudinal study had a mean orientation time of 7·39 sec (S.D. = 3·63 sec) for the Coincident versions and 6·92 sec (S.D. = 2·82 sec) for the Noncoincident versions. This difference was not significant in the analyses either by subjects ($t(23) = 0{\cdot}67$) or by samples ($t(11) = 0{\cdot}70$). However, for these same infants at 11 months of age, the difference in mean orientation times to the Coincident (8·29 sec; S.D. = 3·9 sec) vs. the Noncoincident versions (5·99 sec; S.D. = 2·92 sec) was significant both by subjects ($t(23) = 4{\cdot}00$, $p < 0{\cdot}001$) and by samples ($t(11) = 3{\cdot}59$, $p < 0{\cdot}005$). To verify this apparent developmental difference, we calculated a difference score at each age level for each infant by subtracting the mean orientation time of the Noncoincident version from that of the Coincident version. A paired *t*-test confirmed that the difference scores for the infants at nine and 11 months of age did differ significantly ($t(23) = 2{\cdot}50$, $p < 0{\cdot}02$).

Importantly, the present results with 11-month-olds replicate those of Experiment 1. Hence, the procedural change seems not to have been the factor responsible for the difference between the infants at nine and 11 months of age. Rather, there is evidence of a developmental trend in the present study. In particular, 11-month-olds are apparently more sensitive to the presence of word units in fluent speech than are nine-month-olds. Of course, we cannot rule out the possibility that younger infants might show evidence of sensitivity to word units in fluent speech under other circumstances or with more sensitive paradigms. However, given the present paradigm – the pause insertion method – it does appear that infants become sensitive to smaller and smaller units as they progress developmentally.

Previous studies using the pause insertion technique (Hirsh-Pasek *et al.* 1987; Jusczyk *et al.* 1992) have suggested that infants are sensitive to prosodic markers of linguistically relevant units such as clauses and phrases. Is it possible that 11-month-olds are also relying on prosodic markers for word units in fluent speech? The next experiment explored this possibility.

EXPERIMENT 3

Previous investigations have provided strong reason to suspect that global prosodic information is important in the perception by infants of linguistic units in fluent speech. For example, Jusczyk *et al.* (1992) found that infants still preferred passages with pauses inserted at phrase boundaries over those with pauses inserted within phrases, even when the passages were low-pass filtered to remove most nonprosodic information. Moreover, Gerken *et al.* (1994) found that when prosodic phrase boundaries did not line up with syntactic phrase boundaries, infants preferred segmentations of fluent speech that corresponded to the prosodic phrase boundaries. Other recent studies suggest that infants respond to the kinds of prosodic cues that could signal word boundaries. For instance, Jusczyk *et al.* (1993*a*) found that, at nine months, American infants prefer listening to words that follow the predominant (SW) stress pattern of English words. Moreover, Morgan (1994) has shown that nine-month-olds are more apt to group syllables perceptually into units following a trochaic (i.e. SW) stress pattern.

To begin to address this issue, we investigated how 11-month-olds respond to low-pass filtered versions of the stimuli used in the first two experiments. Low-pass filtering leaves the basic prosodic information in the samples intact, while removing most other sources of information that could potentially be used for identifying word boundaries (e.g. allophonic, phonetic, and phonotactic cues). If 11-month-olds rely primarily on prosodic information to identify word units in fluent speech, then even with low-pass filtered samples, the infants should still prefer the Coincident to Noncoincident versions.

METHOD

Subjects

Twenty-four infants approximately 11 months of age (mean age: 45 weeks, 1 day; range: 44 weeks to 49 weeks, 4 days) from monolingual English-speaking homes in the Eugene, Oregon, area were tested. Two additional infants failed to complete the procedure due to inattentiveness (one) or experimenter error (one).

Stimuli

The stimuli were identical to the Storybook passages used in the previous experiments, except 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 filtering level, which has been used in previous studies (e.g. Jusczyk *et al.* 1992), 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, design and procedure

These were the same as described for Experiment 2.

RESULTS AND DISCUSSION

As in the previous experiments, 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 was 6·61 sec (S.D. = 3·50 sec) for the Coincident versions and 6·16 sec (S.D. = 2·89 sec) for the Noncoincident versions. Separate analyses with subjects ($t(23) = 0{\cdot}60$) and samples

($t(11) = 1.01$, $p < 0.35$) indicated no significant differences in orientation times to the two versions of the samples. Hence, with these low-pass filtered stimuli, there was no indication that infants listened significantly longer to passages with the uninterrupted words (i.e. the Coincident versions) than they did to passages with interrupted words (i.e. the Noncoincident versions).

In contrast to previous investigations of infants' responsiveness to phrasal units (Jusczyk *et al.* 1992), the present results suggest that infants may need more information than is available in prosody to locate boundaries of words in fluent speech. As noted earlier, there are other potential sources of information to word boundaries in fluent speech, including allophonic and phonotactic properties of the input. It is possible that infants rely on these kinds of cues in preference to, or in combination with, prosodic cues to word boundaries. Moreover, there is another possibility raised earlier, namely, that the 11-month-olds have actually learned some of the target words previously in isolation and responded by matching their representations of these to the test samples. In fact, the test materials did contain words (like *animals*) that might already be familiar to infants. The mean frequency of the target words in the passages was actually relatively high, 182 according to the Kucera & Francis (1967) norms. Furthermore, some of the target words were actually repeated in different sentences (the mean number of repetitions was 2·5). Given that this was so, it seemed wise to replicate our original study with 11-month-olds using new materials which better controlled for possible effects of familiarity and word frequency.

EXPERIMENT 4

Our goal was to determine whether 11-month-olds would respond differentially to pauses either at the boundaries or within words that were unlikely to be already familiar to them. Thus, we created a new set of materials based on some stories for children and selected as target words ones that were not only likely to be unfamiliar to infants, but also that had phonetic characteristics that made it easier to insert pauses between syllables without creating acoustic transients. If the preferences observed for the Coincident passages in the previous experiments were simply based on prior knowledge of some of the target words, then 11-month-olds might not show a preference for Coincident versions of the new materials.

METHOD

Subjects

Twenty-four infants approximately 11 months of age (mean age: 45 weeks, 6 days; range: 44 weeks to 47 weeks, 4 days) from monolingual English-speaking homes in the Buffalo area were tested. Thirteen additional infants failed to complete the procedure for the following reasons: crying or restlessness (nine), experimenter error (two), looking times less than 3 sec to each side (one), and sibling interference (one).

Stimuli

Short passages derived from published stories were re-written for the present study. Of these, six were selected to be used as test samples and two to be used as familiarization samples. Each passage was designed such that a target word (that is, a word that would either contain an inserted pause or would be directly adjacent to an inserted pause) would occur approximately once per clause, resulting in a mean number of target words per passage of 5·4 (range: 5–6). The target words were chosen so that inserting a pause within a word between syllables would not result in transients. Moreover, the target words were positioned in sentences so as to allow the insertion of a pause before or after them that did not coincide with a clause or phrase boundary. In contrast to the target words in the previous experiments, the mean frequency of the targets in the present study was only 67 according to the Kucera & Francis (1967) norms. A *t*-test for independent samples indicated that the difference in mean frequencies for the words in the two experiments was marginally significant ($t(102) = 1{\cdot}98$, $p = 0{\cdot}051$). Furthermore, the mean number of repetitions of the target words in the passages, 1·3, was significantly smaller than for the previous experiments ($t(102) = 2{\cdot}46$, $p < 0{\cdot}02$).

The passages were then recorded by a female talker who was naive to the purpose of the experiment. She was instructed to speak in a lively voice, as if reading to a young child. The passages were digitized and stored on a VAXstation 3176 where they could be manipulated by an auditory editing program. The mean length of the test passages was 29·98 sec, with a range of 24·99 to 34·48 sec.

Coincident samples were then constructed by inserting 1 sec pauses either before or after target words (total number of preword pauses = 15; total number of postword pauses = 17). In one test sample, it was not possible to insert a pause immediately adjacent to one target word without creating transients and without inserting at a clause boundary, so the between-word pause was inserted after the first word following the target word. Noncoincident samples were constructed by inserting 1 sec pauses between syllables within target words. In 18 cases, the pauses separated an SW syllable pattern; in 14 cases, the pauses separated a WS syllable pattern. An example of the Coincident (C) and Noncoincident (D) versions of a sample (slashes indicate inserted pauses) is the following:

(C): The Emperor was slightly / disappointed, because he had always thought that his wife and the Chief Minister were very foolish. But when they admired / the material, and talked about it, even / reaching out their hands to touch it, he concluded / that he must have been wrong about them all the time. If they could see it, he certainly / was not going to announce that he could not.

(D): The Emperor was slightly disa/ppointed, because he had always thought that his wife and the Chief Minister were very foolish. But when they ad/mired the material, and talked about it, even rea/ching out their hands to touch it, he con/cluded that he must have been wrong about them all the time. If they could see it, he cer/tainly was not going to announce that he could not.

Apparatus and design

These were the same as in Experiment 2.

Procedure

One change was made in the procedure. Because our dependent measure is the duration of orientation on a given trial, there was no real need to assign each type of sample (Coincident or Noncoincident) to the same side throughout the experiment. Moreover, randomly assigning the samples to the two sides makes it even less likely that the experimenter could guess the type of sample playing on a given trial. For half of the samples used in testing a given infant, the Noncoincident version was played first, and for the other half, the Coincident version occurred first. A different ordering of the 12 test trials was selected by the computer for each subject. The experimenter was blind to the test order used for a given subject.

RESULTS AND DISCUSSION

As in the previous experiments, 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 was 9·16 sec (S.D. = 2·89 sec) for the Coincident versions and 7·37 sec (S.D. = 3·61 sec) for the Noncoincident versions. The analysis by subjects ($t(23) = 2{\cdot}16$, $p < 0{\cdot}05$) indicated that mean orientation times were significantly longer for the Coincident versions. The analysis by samples, although in the same direction, was not significant ($t(5) = 1{\cdot}41$, $p < 0{\cdot}25$). We attribute the latter finding to the smaller number of sample pairs in the present experiment – here six, as compared to 12 in the earlier experiments. In fact, for five of the six sample pairs, orientation times were longer for the Coincident versions (this compares to 10 of 12 for Experiment 1 and for the data with 11-month-olds in Experiment 2).

The present results basically replicate those of the earlier experiments in which 11-month-olds heard the unfiltered speech. Thus, even when low frequency (and, presumably, unfamiliar) words were used and rarely repeated in the passages, the infants continued to show a significant preference for the passages with the pauses at the word boundaries (i.e. the Coincident ones). This result suggests that infants are doing more than simply responding to manipulations involving target words that they already know. Rather, they seem to be sensitive to the presence of other sources of information in speech that help to mark the locations of units such as words. Accordingly, we decided to explore the role that prosodic information might play in helping infants to locate word units in fluent speech.

EXPERIMENT 5

Cutler & Norris (1988) have suggested that listeners could use information about the presence of strong syllables in identifying the onsets of new words in fluent English speech. Specifically, listeners could achieve a rough parse of speech into word units by assuming that each strong syllable initiates the start of a new word. If English-learning infants are following a strong syllable segmentation strategy, they might be more apt to notice interruptions to words with SW stress patterns than they would interruptions to words with WS patterns. This is because a pause in the middle of an SW word (i.e. the Noncoincident versions) would result in a weak syllable appearing right after the pause. By comparison, a pause inserted in the middle of a WS word (i.e. the Noncoincident versions) would result in a strong syllable appearing right after the pause, and could conceivably even be preferable to placing a pause just before the onset or after the offset of the word (i.e. the Coincident versions).

We revised the passages from Experiment 4 to create two new sets of passages. In one set, the SW set, pauses appeared adjacent to or within words with an SW stress pattern (i.e. the word-internal pause was preceded by a strong syllable and was followed by a weak syllable). In the other set, the WS set, pauses appeared adjacent to or within words with a WS stress pattern (i.e. the word-internal pause was preceded by a weak syllable and followed by a strong syllable). Our aim was to determine whether 11-month-olds responded differently to our inserted pause manipulations with these different types of stress pattern.

METHOD

Subjects

Sixty infants approximately 11 months of age (mean age: 45 weeks, 2 days; range: 42 weeks, 3 days to 48 weeks, 5 days) from monolingual English-speaking homes in the Buffalo area were tested. Seventeen additional infants

(10 SW and 7 WS) were tested but failed to complete the procedure for the following reasons: crying (three SW and two WS), experimenter error or equipment failure (four SW and four WS), sibling interference (one SW). In addition, three subjects (two SW and one WS) who completed the procedure were not included in the statistical analyses because each of them had a mean difference in looking times greater than ± 10·0 sec, more than three standard deviations from the mean.

Stimuli

Twelve new sample pairs (6 SW and 6 WS) were used in the test phase of the experiment, along with four new sample pairs (2 SW and 2 WS) for the practice phase. To construct these samples, the short passages used in Experiment 4 were re-written. Of these, six were used as test samples and two as practice samples. Each passage was adapted into two versions, the SW version and the WS version. In the SW version, all target words contained an SW stress pattern on two of its adjacent syllables. In the WS version, all target words had a WS stress pattern on two of its adjacent syllables. All efforts were made to ensure that the passages were otherwise very similar in content and rhythm. Target words in SW passages were matched as closely as possible for frequency relative to target words in WS passages. The mean frequency of the targets was 89 for the SW passages and 70 for the WS passages. They were also matched for syntactic category and number of syllables.

The 16 passages thus derived (six test SW passages, two practice SW passages, and the corresponding six test WS passages and two practice WS passages) were designed so that a target word would occur approximately once per clause. Once again, the target words were chosen so that inserting a pause within a word between syllables would not result in transients, and they were positioned in sentences so that the insertion of pauses did not coincide with a syntactic clause or phrase boundary.

The passages were recorded in a lively voice by the same female talker used in Experiment 4, who was again naive to the purposes of the experiment. The passages were digitized and stored on a VAXstation 3176 where they could be manipulated by an auditory editing program. The mean length of the passages was 29·96 sec, ranging from 26·7 to 32·7 sec.

Coincident samples were constructed by inserting 1 sec pauses either before or after target words. In the SW passages, there was an equal number (2·5) of preword and postword pauses. The placement of preword and postword pauses in the WS passages was the same. Noncoincident samples were constructed by inserting 1 sec pauses between syllables within target words. In the SW passages, the pauses were always preceded by a stressed syllable and followed by an unstressed syllable. In the WS passages, the pauses were always preceded by an unstressed syllable and followed by a stressed syllable. Examples of the Coincident (E) and Noncoincident (F) versions of an SW pair and comparable Coincident (G) and Noncoincident (H) versions of a WS pair are the following:

> (E): The Emperor was slightly / disgruntled, because he had always thought that his wife and the Chief Minister were very foolish. But when they complimented / the material, and talked about it, even / reaching out their hands to touch it, he concluded that they must have more aptitude / than he had thought. If they could see it, he simply / could not proclaim that to him it was invisible.
>
> (F): The Emperor was slightly disgrun/tled, because he had always thought that his wife and the Chief Minister were very foolish. But when they com/plimented the material, and talked about it, even rea/ching out their hands to touch it, he concluded that they must have more ap/titude than he had thought. If they could see it, he sim/ply could not proclaim that to him it was invisible.
>
> (G): The Emperor was slightly / disgruntled, because he had always thought that his wife and the Chief Minister were very foolish. But when they admired / the material, and talked about it, even / extending their hands to touch it, he concluded / that they must have more aptitude than he had thought. If they could see it, he simply could not / proclaim that to him it was invisible.
>
> (H): The Emperor was slightly dis/gruntled, because he had always thought that his wife and the Chief Minister were very foolish. But when they ad/mired the material, and talked about it, even ex/tending their hands to touch it, he con/cluded that they must have more aptitude than he had thought. If they could see it, he simply could not pro/claim that to him it was invisible.

Apparatus and procedure

These were basically the same as in Experiment 4. However, a small hole was cut into the centre panel of the apparatus, to allow for videotaping the test sessions. The videotapes were used to provide observer reliability checks (there was correlation = 0·94 between the live and videotape observers).

RESULTS AND DISCUSSION

As in the previous experiments, the data from the 12 test trials were used to calculate the mean length of orientation to each type of sample. For infants who heard the samples with SW target words, the mean length of orientation across subjects was 10·36 sec (S.D. = 1·83 sec) for the Coincident versions and

9·04 sec (S.D. = 3·61 sec) for the Noncoincident versions. The analysis by subjects ($t(29) = 2{\cdot}17$, $p < 0{\cdot}05$) indicated that mean orientation times were significantly longer for the Coincident versions of the samples with SW targets. However, the analysis by samples was not significant ($t(5) = 1{\cdot}00$). A similar pattern of results occurred for the samples with the WS targets. The mean length of orientation across subjects was 10·03 sec (S.D. = 1·51 sec) for the Coincident versions and 8·70 sec (S.D. = 1·50 sec) for the Noncoincident versions. The analysis by subjects ($t(29) = 2{\cdot}42$, $p < 0{\cdot}05$) indicated that mean orientation times were significantly longer for the Coincident versions of the samples with SW targets and, once again, the analysis by samples was not significant ($t(5) = 1{\cdot}51$, $p < 0{\cdot}20$).

Thus, the present study provides further confirmation that 11-month-olds are sensitive to the presence of word units in fluent speech. At the same time, the results indicate that 11-month-olds were no less sensitive to the packaging of WS word units in fluent speech than they were to SW word units. Infants consistently listened longer to passages with uninterrupted words than to ones with interrupted words, regardless of whether the words had SW or WS stress patterns. It appears that infants in our study must have relied on more than prosodic sources of information to identify word units in fluent speech.

As noted earlier, one type of information that might be available to infants has to do with the phonotactic properties of the input. Some kinds of phonetic sequences (e.g. 'nt') occur relatively more often within words, whereas other sequences (e.g. 'mt') appear mostly across word boundaries (and seldom between syllables within words). Moreover, there are indications that by nine months, infants show some sensitivity to the distribution of phonotactic sequences in their native language (Friederici & Wessels, 1993; Jusczyk *et al.* 1993*b*). Accordingly, it would be interesting to determine whether phonotactic cues to word boundaries were present in the actual speech samples that we employed in our study.

To pursue this issue, we examined the nature of the phonotactic sequences which were interrupted by the inserted pauses. We calculated biphone frequencies based on Bernstein (1982), a sample of child-directed speech available in MacWhinney (1991). This corpus was used to provide measures of the probability of finding a particular sequence of two consonants either within a word or between two successive words. All probabilities were computed based on log frequency-weighted values (Kucera & Francis, 1967). For each such sequence that was interrupted in either the Coincident or Noncoincident versions of the samples, we devised a measure of how likely a given sequence was to mark a word boundary. To derive this index of word-unit informativeness, the probability of the sequence appearing within words was divided by the probability of the sequence appearing between two words plus the probability of its occurring within a word. The higher the index, the greater the likelihood that the consonant sequence tends to appear in the same word unit. For the SW items, there were 15 such consonant–consonant sequences interrupted by pauses in the Coincident versions, and 23 in the Noncoincident versions. The comparable data for WS items were 20 and 17 in the Coincident and Noncoincident versions, respectively.

The mean values of the phonotactic index for Coincident and Noncoincident versions were respectively 0·232 and 0·523 for the SW materials and respectively 0·140 and 0·763 for the WS materials. The indices were entered into an ANOVA for a 2 (stress pattern: SW or WS) × 2 (passage type: Coincident or Noncoincident) design. There was a significant main effect of Passage Type ($F(1,71) = 22{\cdot}84$, $p < 0{\cdot}0001$) indicating that high probability within-word sequences were more likely to occur in the Noncoincident versions (where they would be disrupted by pauses). Importantly, this confirms that there was phonotactic information relevant to the presence or absence of word boundaries in the samples the infants actually heard. There was no main effect of stress pattern ($F(1,71) < 1{\cdot}00$), but the interaction between stress pattern and passage type was marginally significant ($F(1,71) = 3{\cdot}02$, $p < 0{\cdot}10$). In particular, the WS materials tended to be better marked with respect to phonotactic cues for word boundaries. This latter finding was borne out by separate *post-hoc* tests for the SW and WS materials. In particular, whereas there was evidence that the within-word consonant-to-consonant probabilities were significantly higher for the Noncoincident versus the Coincident versions of the WS materials ($t(35) = 5{\cdot}23$, $p < 0{\cdot}0001$), the same comparison for the SW materials proved to be only marginally significant ($t(36) = 1{\cdot}95$, $p < 0{\cdot}06$).

As the mean indices show, the WS materials were more apt to include higher probability within-word consonant sequences at pause locations in Noncoincident samples and lower probability within-word sequences at pause locations in Coincident samples. Indeed, it is possible that the better phonotactic marking in the WS samples may actually have helped to offset any advantage that the SW materials had with respect to prosodic marking of word boundaries. Perhaps, infants were responsive to both kinds of cues. Thus, the present study is consistent with the view that infants draw on multiple sources of information to locate word units in fluent speech.[4]

---

[4] We also examined the materials in Experiment 4 using the same metric to determine whether there were reliable phonotactic cues to word boundaries present. A *t*-test for independent samples indicated once again that the index of word-unit informativeness was significantly higher at pause locations in Noncoincident versions than in Coincident versions ($t(32) = 2{\cdot}46$, $p < 0{\cdot}02$). The mean values of the phonotactic index for coincident and noncoincident versions were 0·152 and 0·510, respectively.

GENERAL DISCUSSION

Two major findings emerge from the five experiments in the present study. First, there is clear evidence that by the age of 11 months, but apparently not earlier, infants are able to discriminate between passages where inserted pauses coincide with word boundaries and passages where inserted pauses occur between syllables within words. This was a very robust result, having been found with four different sets of passages, with three different versions of the headturn preference procedure, and on five different occasions (i.e. the original Storybook passages used in Experiments 1 and 2, the new passages used in Experiment 4, and the two new versions of the Experiment 4 passages in Experiment 5). This strongly suggests that by this age, infants have some capacity for detecting word units in the speech stream.

Second, there is little evidence in the present studies that prosody played a pivotal role in the way in which infants performed. Specifically, infants showed no preference between Coincident and Noncoincident passages when they were low-pass filtered to remove most nonprosodic information. We also found no evidence that infants preferred Noncoincident passages when in such passages the inserted pauses always preceded stressed syllables, which, according to the Cutler & Norris (1988) hypothesis about word segmentation in English, should have served as a prosodic cue to word onset. However, there was suggestive evidence that any differential advantage for words with the predominant English stress pattern might have been offset in our samples by more reliable phonotactic cues in the WS samples. Perhaps infants are relying on multiple sources of information in the signal, not just prosody, to help locate word boundaries.

These two findings have interesting implications for the interpretation of results found in earlier studies. The first finding, that infants only respond differentially to Coincident and Noncoincident passages by the age of 11 months, is consistent with the developmental trend observed by Hirsh-Pasek *et al.* (1987) and Jusczyk *et al.* (1992). Namely, infants seem to become sensitive to linguistically significant units in an order inverse to the size of the units. Thus infants detect clause-sized units in the pause-insertion paradigm already by four-and-a-half months of age, phrase-sized units only by nine months, while word-sized units only appear to be detected by 11 months.

This trend is consistent with a differentiation view of development such as that advocated by Gibson (1969) and others, where perceptual development occurs through the individual's learning to be sensitive to stimuli at ever finer levels of detail. One interpretation of the trend, however, would lead to the improbable prediction that infants will not show any ability to detect units in fluent speech that are smaller than the word, such as syllables, until after 11 months. This prediction is impossible to test using the pause-insertion paradigm, since inserting artificial pauses within syllables would give rise to transients that could lead infants to reject manipulated passages for purely acoustic reasons. However, evidence from other tasks suggests that infants far younger than 11 months, including newborns, already perceive syllables as units (Bertoncini & Mehler, 1981; Jusczyk, Jusczyk, Kennedy, Schomberg & Koenig, 1995).

The view that infants differentiate increasingly smaller units from the speech stream may also have to be refined in light of some other recent findings. Evidence garnered from another paradigm suggests that the beginnings of word segmentation processes in infants may occur earlier than expected on the basis of our pause-insertion task. In a study using the word-detection task, Jusczyk & Aslin (1995) first familiarized seven-and-a-half-month-olds with two short passages, each of which included a monosyllabic target word that was repeated in the different sentences of the passage. They then presented the infants with trials consisting of a repeated word produced in isolation. Two of the trials contained the target words from the passages; the other two contained novel words. Jusczyk & Aslin (1995) found that infants attended longer to words from the passages than to the novel words. In a follow-up study that is even more germane to the present findings, Newsome & Jusczyk (1995) used the word-detection paradigm and found that seven-and-a-half-month-olds also showed evidence of detecting bisyllabic words with an SW stress pattern in fluent speech contexts. On the face of it, these results suggest that already by seven-and-a-half months, infants have some capacity to recognize words embedded in fluent speech. How can this finding be reconciled with the results of the present investigation?

There are two possible explanations for this apparent discrepancy in the age at which infants appear to be responding to word units in fluent speech. The first is that the word detection paradigm may simply be more sensitive than the pause insertion paradigm used in the present study. Note that the method used in the present study may require that infants have more precise information about the location of word boundaries than does the word-detection method used by Jusczyk & Aslin. Specifying word boundaries with greater precision may require that infants draw on more than just prosodic cues. In particular, to determine fully the location of word boundaries, the infant listener may have to co-ordinate information from a range of different cues (prosodic, phonotactic, allophonic, etc.). Indeed, some investigators have suggested that infants may only begin to develop the capacity to co-ordinate different sources of speech information at around nine months of age (Morgan & Saffran, 1995; Lalonde & Werker, in press).

The second possible reason for the discrepancy is that the two tasks may be appropriate for tapping different phases of the development of word perception. The fact that infants in the present study do not false-alarm on Noncoincident WS passages, where pauses appear at foot boundaries, shows that infants are in fact detecting the boundaries of the morphological and syntactic units known as words and not prosodic units like feet. This pattern contrasts sharply with what was found in the pause-insertion studies of clause-sized and phrase-sized units. The results of Gerken *et al.* (1994) make this unambiguous in the case of phrase-sized units: when one allows the boundaries of syntactic and prosodic phrases to vary independently in the samples, the infants are clearly seen to be responding to prosodic phrases, not syntactic phrases.

The fact that the units being detected in the current study are truly words suggests that what seven-and-a-half-month-olds are responding to in the word-detection paradigm are not exactly words *per se*, but rather prosodic units near to words in size, perhaps feet. Thus, the first step that English-learning infants may take toward word segmentation is to parse the input into something like trochaic feet (i.e. units consistent with the predominant SW stress patterns of English words). If so, we predict that infants might initially have difficulty in segmenting WS words from fluent speech. For example, seven-and-a-half month-olds might be more likely to match *tar* than *guitar* to the occurrence of *guitar* in fluent speech. This is because *guitar* contains a foot boundary beginning with the syllable *tar*. In fact, this result is exactly what Newsome & Jusczyk (1995) found when they examined seven-and-a-half-month-olds' detection of WS words in fluent speech passages.

The picture of the development of word segmentation that emerges is somewhat complex. The findings of Newsome & Jusczyk (1995) and Morgan (in press) suggest that it is indeed appropriate to apply something like the Cutler & Norris (1988) hypothesis to the question of word segmentation by infants, since English-learning infants do seem to rely heavily on stress patterns in making a first pass at determining the location of words in fluent speech. Nevertheless, because English also has many words that do not have syllable-initial stress, fixing the location of word boundaries more accurately may require the use of nonprosodic information as well. It is here that phonotactics and allophonic constraints may contribute to word segmentation – by supplementing and refining a prosodically based initial parse of the input. However, gaining access to phonotactics and allophonic constraints may depend on the prior breakdown of the input into the appropriate-sized processing chunks. Consequently, by delimiting possible units in the input, a trochaic foot-based segmentation may actually help to make the phonotactics and allophonic constraints more accessible to infant listeners. In sum, although prosodic cues may play a leading role in word segmentation, sensitivity to native language phonotactics and allophonic constraints may provide some necessary fine tuning.

The relatively late age at which the pause-insertion task finds evidence for the detection of word boundaries seems to reflect both a genuinely later stage of development along the road to mature word segmentation strategies, as well as a difference in what factors of the signal are crucial for performing well on this particular task. Word segmentation is a profoundly difficult problem, and it should not surprise us to learn that information from multiple sources must ultimately all be co-ordinated in order for infants truly to gain mastery over it.

REFERENCES

Aslin, R. N., Pisoni, D. B. & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In M. M. Haith & J. J. Campos (eds), *Infancy and the biology of development*. New York: Wiley.

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

Benedict, H. (1979). Early lexical development: comprehension and production. *Journal of Child Language* **6**, 183–201.

Bernstein, N. (1982). Acoustic study of mothers' speech to language-learning children: an analysis of vowel articulatory characteristics. Unpublished doctoral dissertation, Boston University.

Bertoncini, J. & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development* **4**, 247–60.

Best, C. T. (1993). Emergence of language-specific constraints in perception of native and non-native speech: a window on early phonological development. In B. de Boysson-Bardies, sec de Schonen, P. Jusczyk, P. McNeilage & J. Morton (eds), *Developmental neurocognition: speech and face processing during the first year of life*. Dordrecht: Kluwer.

Brent, M. R., Cartwright, T. A. & Gafos, A. (in press). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*.

Brown, R., Cazden, C. & Bellugi, U. (1969). The child's grammar from I to III. In J. P. Hill (ed.), *Minnesota Symposium on Child Psychology*, vol. II. Minneapolis: University of Minnesota Press.

Chaney, C. (1989). I pledge a legiance to the flag: Three studies in word segmentation. *Applied Psycholinguistics* **10**, 261–81.

Christophe, A., Dupoux, E., Bertoncini, J. & Mehler, J. (1994). Do infants perceive word boundaries? An empirical approach to the bootstrapping problem for lexical acquisition. *Journal of the Acoustical Society of America* **95**, 1570–80.

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

Cole, R. & Jakimik, J. (1980). How are syllables used to recognize words? *Journal of the Acoustical Society of America* **67**, 965–70.

Cutler, A. (1989). Auditory lexical access: where do we start? In W. Marslen-Wilson (ed.), *Lexical representations and process*. Cambridge, MA: MIT Press.

Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: evidence from juncture misperception. *Journal of Memory and Language* **31**, 218–36.

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* **2**, 133–42.

Cutler, A. & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* **14**, 113–21.

Eimas, P. D. (1982). Speech perception: a view of the initial state and perceptual mechanisms. In J. Mehler, M. Garrett & E. C. T. Walker (eds), *Perspectives on mental representation: experimental and theoretical studies of cognitive processes and capacities*. Hillsdale, NJ: Erlbaum.

Friederici, A. D. & Wessels, J. M. I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception & Psychophysics* **54**, 287–95.

Garnica, O. K. (1977). Some prosodic and paralinguistic features of speech to young children. In C. Snow & C. A. Ferguson (eds), *Talking to children: language input and acquisition*. Cambridge: C.U.P.

Gerken, L. A., Jusczyk, P. W. & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: nine-month-olds' sensitivity to phonological vs. syntactic phrases. *Cognition* **51**, 237–65.

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

Gleitman, L. & Wanner, E. (1982). The state of the state of the art. In E. Wanner & L. Gleitman (eds), *Language acquisition: the state of the art*. Cambridge: C.U.P.

Halle, M. & Vergnaud, J. R. (1987). *An essay on stress*. Cambridge, MA: MIT Press.

Hayes, B. (1981). A metrical theory of stress rules. Ph. D. dissertation, MIT.

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**, 269–86.

Hohne, E. A. & Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception & Psychophysics* **56**, 613–23.

Holden, M. H. & MacGinitie, W. H. (1972). Children's conception of word boundaries in speech and print. *Journal of Educational Psychology* **63**, 551–7.

Huttenlocher, J. (1974). The origins of language comprehension. In R. L. Solso (ed.), *Theories in cognitive psychology*. New York: Wiley.

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, Missouri.

Jusczyk, P. W. & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology* **29**, 1–23.

Jusczyk, P. W., Cutler, A. & Redanz, N. (1993*a*). Preference for the predominant stress patterns of English words. *Child Development* **64**, 675–87.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y. & Jusczyk, A. M. (1993*b*). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* **32**, 402–20.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L., Woodward, A. & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology* **24**, 252–93.

Jusczyk, P. W., Jusczyk, A. M., Kennedy, L. J., Schomberg, T. & Koenig, N. (1995). Young infants' retention of information about bisyllabic utterances. *Journal of Experimental Psychology: Human Perception and Performance* **21**, 822–36.

Jusczyk, P. W., Luce, P. A. & Charles Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*. **33**, 630–645.

Kaye, K. (1976). Infants' effect on their mothers' teaching strategies. In J. C. Glidewell (ed.), *The social context of learning development*. New York: Gardner.

Kemler Nelson, D. G. (1989). Developmental trends in infants' sensitivity to prosodic cues correlated with linguistic units. Paper presented at The biennial meeting of the Society for Research in Child Development, Kansas City, Missouri.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W. & Wright-Cassidy, K. (1989). How prosodic cues in motherese might assist language learning. *Journal of Child Language* **16**, 55–68.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A. & Gerken, L. A. (1995). The Headturn Preference Procedure for testing auditory perception. *Infant Behavior & Development* **18**, 111–16.

Klatt, D. H. (1980). Speech perception: a model of acoustic–phonetic analysis and lexical access. In R. A. Cole (ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.

Kucera, H. & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (eds), *Handbook of infant perception*. New York: Academic Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. (1992). Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science* **255**, 606–8.

Lalonde, C. E. & Werker, J. F. (in press). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior & Development*.

Liberman, A. M. & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Liebowicz & H. L. Teuber (eds), *Handbook of sensory physiology*. Berlin: Springer Verlag.

MacWhinney, B. (1991). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Erlbaum.

Mandel, D. R., Jusczyk, P. W. & Kemler Nelson, D. G. (1995). Does sentential prosody help infants to organize and remember speech information? *Cognition* **53**, 155–80.

Marcus, S. M. (1984). Recognizing speech: on the mapping from sound to word. In H. Bouma & D. G. Bouwhuis (eds), *Attention and performance: control of language processes*. Hillsdale, NJ: Erlbaum.

Messer, D. J. (1981). Identification of names in maternal speech to infants. *Journal of Psycholinguistic Research* **10**, 69–77.

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

—— (1994). Converging measures of speech segmentation in prelingual infants. *Infant Behavior & Development* 17, 387–400.

—— (in press). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*.

Morgan, J. L. & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development* **66**, 911–36.

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

Newsome, M. & Jusczyk, P. W. (1995). Do infants use stress as a cue for segmenting fluent speech? In D. McLaughlin & S. McEwen (eds), *Proceedings of the 19th Boston University Conference on Language Development*. Boston, MA: Cascadilla Press.

Peters, A. (1983). *The units of language acquisition*. Cambridge, C.U.P.

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

Polka, L. & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* **20**, 421–35.

Ryan, M. L. (1978). Contour in context. In R. N. Campbell & P. T. Smith (eds), *Recent advances in the psychology of language*. New York: Plenum.

Stern, D. N., Spieker, S., Barnett, R. K. & Mackain, K. (1983). The prosody of maternal speech: infant age and context-related changes. *Journal of Child Language* **10**, 1–15.

Suomi, K. (1993). An outline of a developmental model of adult phonological organization and behavior. *Journal of Phonetics* **21**, 29–60.

Tunmer, W. E., Bowey, J. A. & Greve, R. (1983). The development of young children's awareness of the word as a unit of spoken language. *Journal of Psycholinguistic Research* **12**, 567–94.

Waibel, A. (1986). Suprasegmentals in very large vocabulary word recognition speech perceptions. In E. C. Schwab & H. C. Nusbaum (eds), *Pattern recognition by humans and machines*. New York: Academic Press.

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.

Woodward, J. Z. & Aslin, R. N. (1990). Segmentation cues in maternal speech to infants. Paper presented at 7th biennial meeting of the International Conference on Infant Studies, Montreal, Quebec, Canada.