Journal of Memory and Language **44,** 548–567 (2001)
doi:10.1006/jmla.2000.2755, available online at http://www.academicpress.com on IDEAL®

Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics

Elizabeth K. Johnson and Peter W. Jusczyk

Departments of Psychology and Cognitive Science, Johns Hopkins University

Fluent speech contains few pauses between adjacent words. Cues such as stress, phonotactic constraints, and the statistical structure of the input aid infants in discovering word boundaries. None of the many available segmentation cues is foolproof. So, we used the headturn preference procedure to investigate infants' integration of multiple cues. We also explored whether infants find speech cues produced by coarticulation useful in word segmentation. Using natural speech syllables, we replicated Saffran, Aslin, et al.'s (1996) study demonstrating that 8-month-olds can segment a continuous stream of speech based on statistical cues alone. Next, we added conflicting segmentation cues. Experiment 2 pitted stress against statistics, whereas Experiment 3 pitted coarticulation against statistics. In both cases, 8-month-olds weighed speech cues more heavily than statistical cues. This observation was verified in Experiment 4, which indicated that greater complexity of the familiarization sequence does not necessarily lead to familiarity effects.

Key Words: coarticulation; word segmentation; stress cues; statistical cues; prosodic stress; transitional probabilities; statistical learning.

Much of the linguistic input that infants receive consists of multiword utterances (van de Weijer, 1998; Woodward & Aslin, 1990). Consequently, to build up a vocabulary and acquire language, infants must first learn to segment words accurately from fluent speech. Unlike printed text, speech contains no reliable spaces or pauses between words (Cole & Jakimik, 1980). The difficulties caused by this lack of clearly marked word boundaries in spoken utterances can be easily appreciated by considering what it is like to listen to someone speaking an unfamiliar foreign language. One has the impression of hearing a continuous flow of speech with few easily discernible word units. This leads to the impression that speakers of these languages speak more rapidly. However, many foreign language speakers report the same difficulties in segmenting English words. It is not until one gains a certain degree of familiarity with a language that the speech seems to slow down and one can begin to segment words. Eventually, parsing the speech stream is possible because of cues to word boundaries that can be exploited in segmentation. However, which cues are most useful for segmenting words depends critically on the sound organization of the language. In this sense, such cues are language specific. In fact, adults will impose a segmentation strategy appropriate for their native language on any speech input, be it a foreign language (Cutler, Mehler, Norris, & Segui, 1986) or an artificial language (Vroomen, Tuomainen, & de Gelder, 1998). Given the complex nature of word boundary cues, word segmentation is a formidable hurdle for infants. If infants fail to learn to parse their speech input correctly, the mapping of meanings to acoustic units will not match the lexicon of the language, and language acquisition will be impeded.

The need to learn about word segmentation cues for acquiring a vocabulary might be obvi-

Support for this research was provided by a research grant from NICHD (15795) and a Senior Scientist Award from NIMH (01490) to PWJ. The authors also thank Doug Whalen and Joyce McDonough for helpful suggestions about coarticulatory cues and Joost van de Weijer, Stew Hulse, Ann Marie Jusczyk, James McQueen, and an anonymous reviewer for comments on previous versions of this manuscript.

Address correspondence and reprint requests to Peter W. Jusczyk, Department of Psychology, Ames Hall, Johns Hopkins University, Baltimore, MD 21218-2686. E-mail: jusczyk@jhu.edu.

0749-596X/01 $35.00

ated, or greatly reduced, if infants could count on receiving a considerable amount of exposure to words occurring in isolation (Bloomfield, 1933; Brent, 1999; Suomi, 1993). However, it seems unlikely that attention to words in isolation is sufficient for infants to parse the input accurately. Most infant-directed utterances contain more than one word (van de Weijer, 1998). Moreover, even when mothers are explicitly instructed to teach their children a new word, they produce words in isolation no more than 20% of the time (Woodward & Aslin, 1990). Furthermore, some function words such as *a* and *the* are extremely unlikely to occur in isolation.

In the latter half of their first year, infants begin to display some capacity to segment words from fluent speech. Jusczyk and Aslin (1995) reported that infants as young as 7.5 months of age can segment monosyllabic words such as *cup* from fluent speech. This result held regardless of whether infants were first exposed to these words as isolated utterances and then tested on passages containing the words or whether they were first exposed to passages containing the words and then tested on isolated versions of the words. In contrast to 7.5-month-olds, 6-month-olds did not display any evidence of segmenting words. Hence, Jusczyk and Aslin concluded that word segmentation abilities of English-learning infants develop between the ages of 6 and 7.5 months of age.

More recent work has focused on identifying the speech cues which infants rely on to identify where one word ends and the next begins. There are several potential sources of information that listeners could use in word segmentation. These include: (1) prosodic markers (Cutler & Norris, 1988; Jusczyk, Houston, & Newsome, 1999; Morgan, 1996); (2) phonotactic constraints, i.e., restrictions on the possible ordering of phones within a word (Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997; Mattys, Jusczyk, Luce, & Morgan, 1999); (3) context-sensitive allophones (Church, 1987; Hockett, 1955; Hohne & Jusczyk, 1994; Jusczyk, Hohne, & Bauman, 1999; Lehiste, 1960); and (4) statistical regularities in the input (Aslin, Saffran, and Newport, 1998; Brent & Cartwright, 1996; Saffran, Aslin, et al. 1996). These cues are probabilistic in nature rather than deterministic. In other words, no single cue is sufficient to accurately segment English words. For instance, Cutler and Carter (1987) reported that approximately 90% of English content words in conversational speech begin with a strong syllable. Given this finding, Cutler and Norris (1988) posited that listeners use a metrical segmentation strategy (MSS) wherein they take each stressed syllable to mark the onset of a new word. This strategy leads to the correct segmentation of words such as *donor* and *pencil*, with the predominant trochaic (strong–weak) stress pattern of English, but to the incorrect segmentation of words with alternative stress patterns, such as *belong* and *across*. To segment words with less common stress patterns (i.e., weak–strong), listeners need to draw on other cues such as the co-occurrence of speech sounds and probabilistic phonotactics (Vitevitch & Luce, 1998).

A prerequisite for infants to use such cues in segmenting words from fluent speech is that they be sensitive to the occurrence of these cues. Studies indicate that, between 6 and 9 months, infants develop sensitivity to two potential word boundary cues—predominant word stress patterns (Jusczyk, Cutler, & Redanz, 1993; Morgan & Saffran, 1995) and phonotactic patterns (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce, and Charles-Luce, 1994).

Because English has a distinctive trochaic structure, and newborns are sensitive to this rhythmic structure (Nazzi, Bertoncini, & Mehler, 1998), a reasonable segmentation strategy for English-learning infants might be to posit the beginning of a new word at each stressed syllable. Using the headturn preference procedure, Jusczyk, Houston, et al. (1999) found that infants as young as 7.5 months segment bisyllabic words with a strong–weak stress pattern (i.e., kingdom). Additional experiments showed that infants had parsed both syllables of the word and were not simply responding to the first stressed syllable (i.e., king). However, infants at this age did not segment weak–strong words (i.e., guitar) from fluent speech. Furthermore, when a weak–strong word was consistently followed by the same unstressed word (i.e., *guitar* followed by *is*), infants treated the stressed syllable and the following unstressed monosyllabic word as a unit (i.e., *taris*). Jusczyk, Houston, et al. concluded that, although 7.5 month olds perceive stressed syllables as markers of word onsets, they are also sensitive to the distribution of syllables within the speech stream. When two syllables consistently co-occur, they are perceived as two parts of a single unit. Not until 10.5 months do infants segment weak–strong words from fluent speech. In this regard, it is interesting that sensitivity to two other potential English word boundary cues, phonotactics (Mattys & Jusczyk, 2001) and context-sensitive allophones (Jusczyk, Hohne, et al., 1999), appears to develop between 7.5 and 10.5 months of age. Thus by 10.5 months, English learners seem to draw on multiple cues to word boundaries. This ability to integrate multiple cues to word boundaries is a necessary skill, since English contains many words with weak–strong stress (e.g., verbose, compute).

Thus, sensitivity to prosodic cues emerges early, followed soon thereafter by sensitivity to other word boundary cues. As noted above, one such cue is the ability to track statistical regularities in the input. Morgan and Saffran (1995) used the conditioned headturn procedure and a click detection paradigm to show that 9-month-olds perform best when presented syllable sequences with both a fixed ordering and a fixed rhythmic pattern. In contrast, 6-month-olds perform equally well with or without fixed ordering patterns. Thus, younger infants are only sensitive to the rhythmic properties of the input. By 8 months, infants are sensitive to statistical properties of the input (Saffran, Aslin, et al., 1996) and by 9 months, they are presumably integrating these two cues.

The study by Saffran, Aslin, et al. (1996) has drawn attention to information that infants can extract from speech on the basis of distributional cues alone. Saffran, Aslin, et al. reasoned that because words can be defined as units of sound which consistently co-occur, noting the likelihood of one syllable following another could provide a reliable strategy for segmenting words. For example, the two-word string *prettybaby* consists of four syllables: *pre*, *ty*, *bay*, and *by*. The first two syllables (*pre* and *ty*) consistently appear together because they form a word. Likewise, the latter two syllables (*bay* and *by*) also tend to occur together. However, the second and third syllables (*ty* and *bay*) occur together relatively rarely. Across a corpus of English, the syllable *ty* follows the syllable *pre* more frequently than the syllable *bay* follows the syllable *ty*, because many different words can follow the word *pretty* (i.e., *pretty flower*), but only a few syllables can follow *pre*. This greater predictability of word internal syllables than syllables spanning word boundaries may be helpful in discovering word boundaries.

Still, noting co-occurrences between syllables will not provide a sufficient cue to accurately segment the speech stream. For instance, if infants were to segment the input simply by noting co-occurrences between syllables, they would be misled to treat commonly occurring syllable pairs, such as *the dog*, as words. Therefore, Saffran, Aslin, et al. (1996) proposed that besides tracking the likelihood of one particular syllable following any other particular syllable, infants also track the baseline frequency of the first syllable in the syllable pair. This parsing strategy can be formalized by a statistical relationship: Transitional Probability, where T.P. = (frequency of Y given X)/(frequency of X). Thus, frequently occurring words such as *the dog* will not be mistaken as a word because *the* also occurs before many other words. Aslin et al. (1998) showed that infants respond to transitional probabilities as opposed to simple co-occurrences between syllables.

The idea of tracking the probability of one phone following another to detect word boundaries is not new (Harris, 1951; Hayes & Clark, 1970). However, Saffran, Aslin, et al. (1996) first showed that statistics are a psychologically plausible means for infants to begin to segment words. They familiarized 8-month-olds with a 2-min stream of an artificial language containing four tri-syllabic nonsense words: *pabiku*, *tibudo*, *golatu*, and *daropi*. No acoustic cues to word boundaries were present in the speech stream. Only the distributional properties of the sequences of syllables provided cues to the location of word boundaries. After familiarization, the infants were tested for their listening preferences to words versus part-words (tri-syllabic sequences composed of the last syllable of a word and the first two syllables of another word; based on the nonsense words mentioned above, *tudaro* is a part-word). The infants looked significantly longer to the part-words, indicating they can segment the speech stream based on statistics alone.

Although a potentially powerful tool for word segmentation, statistical cues alone cannot account for infants' ability to discover word boundaries. Transitional probabilities can be misleading, as illustrated by children's speech errors. Consider this example from Peters (1985). A mother tells her child to "behave," and the child responds, "I am have!" In this case, the high frequency of the syllable "be" as an independent word may have led the child to mis-parse *behave* into two separate words: *be* and *have*. Accurate segmentation of words seems to require an integration of many segmentation cues.

Understanding how infants integrate word boundary cues demands knowledge of the cues they use at particular points in development and how much weight they give to each. One can start to address these issues by examining how infants respond when different cues provide conflicting information about possible word boundaries in speech. Mattys et al. (1999) took a step in this direction when (Experiment 4) they pitted sequences with good prosodic cues and poor phonotactic cues to word boundaries against ones with good phonotactic cues but poor prosodic cues. English-learning 9-month-olds favored the sequences with the good prosodic cues, which suggests that, at this age, they give greater weight to prosodic cues than to phonotactic cues. In the present study, we explore how infants weight statistical cues relative to certain speech cues to word boundaries. In one case, we put statistical cues in conflict with prosodic cues to word boundaries. As indicated above, from 7.5 months of age, English learners can use prosodic cues to segment words (Jusczyk, Houston, et al., 1999). Similarly, Saffran, Aslin, et al. found that 8-month-olds can use statistical cues in the absence of other cues to segment words in fluent speech. Thus, observing how infants respond to conflicts between these cues will be informative of the relative weight they give to each in segmenting words from fluent speech.

Our second comparison contrasts statistical cues against less well-studied types of speech cues, namely, cues produced by coarticulation. Our decision to examine cues produced by coarticulation was based on a consideration of the special role that such cues might play in word segmentation. Although the primary focus in research on word segmentation abilities has been on infants' detection of the onsets of new words in utterances, another component of word segmentation is less frequently discussed. How do learners decide when two successive syllables belong together in the same word? Statistical cues are one possible solution to this problem. Specifically, a high transitional probability for two contiguous syllables is a good indication that they belong together. Indeed, Jusczyk, Houston, et al. (1999) suggested that attention to such co-occurrence relations was the likely reason why 7.5-month-olds tended to segment *taris* from passages containing repeated occurrences of the word sequence *guitar is*. However, other speech cues may link syllables into word units, namely, those produced by coarticulation.

According to Ladefoged (1993), coarticulation can be defined as "the overlapping of adjacent articulations." The acoustic properties of a particular phonetic segment are influenced by the nature of surrounding segments in the utterance. Coarticulation is the primary reason for the difficulty in specifying invariant acoustic properties for phonetic segments. Despite the difficulties that coarticulation raises for identifying phonemes on the basis of acoustic cues, the blending of speech sounds makes production much more efficient by allowing for faster rates of transmission (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Although coarticulation makes the perception of individual phonetic segments more difficult in some ways, the degree or manner of coarticulation between neighboring sound segments may actually aid speech perception in another way. Greater coarticulation within words as opposed to across word boundaries could simplify both segmentation and recognition.

Ladefoged (1993) reported that the amount of coarticulation between two phonemes depends in part on the interval between them. For example, a high degree of lip rounding occurs when the [k] in *coo* is produced, due to the following rounded vowel. However, if [k] is separated from the rounded vowel by another phoneme such as the [l] in *clue*, less lip rounding occurs while the [k] is being produced. If a word boundary is added between the [k] and [l], as in *sack Lou*, even less lip rounding is present in the production of the [k]. Thus, although all fluently produced speech is coarticulated, the degree of coarticulation between neighboring phonemes is influenced by the presence or absence of a word boundary. Of course, in the example above, word boundaries are conflated with syllable boundaries. Syllable boundaries are phonetically marked (Krakow, 1999), but do word boundaries induce additional marking? If *sack Lou* were a single word, would there be more coarticulation between the [k] and [l]?

There are many indications that speech carries acoustic cues to a multi-tiered hierarchy of prosodic boundaries. These signals include changes in F0 (Nespor & Vogel, 1986), varying patterns of glottalization (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996), and final lengthening (Wrightman, Shattuck-Hufnagel, Ostendork, & Price, 1992). These *prosodic signatures* have been shown to include articulatory strengthening of consonants at prosodic junctures. In other words, less overlap of adjacent sound segments occurs in word-final and word-initial consonant articulations belonging to different prosodic domains (Byrd, 1996; Byrd & Saltzman, 1998). A study of artificial grammar learning in adults demonstrates how speech cues produced by coarticulation could be helpful to language learners. Valian and Levitt (1996) found that coarticulation cues aid in recovering syntactic units in utterances by fostering increased internal cohesion among elements of such units.

Word boundaries correlate with syntactic boundaries such as phrases. Therefore, the edges of words will tend to differ from the middle of words. We would like to take this argument one step further and suggest that onsets and offsets of individual words are phonetically cued in the speech signal by differential articulation. This idea is not new, as the following quote from Fujimura (1990, p.232) makes clear: “syllable initial position as well as word and phrase initial position, seem to be generally characterized by more ‘forceful’ articulatory gestures.” Evidence for this view is growing. For example, Pierrehumbert and Talkin (1992) reported that the glottal opening for [h] is greater at the beginning of a phrase than in the middle. Similarly, consonantal articulations have greater magnitudes in word-initial than in word-final position (Cooper, 1991). Also, the magnitude of lingual gestures has been shown to increase as a function of the strength of the prosodic edge along which it falls (Fougeron & Keating, 1996, 1998).

Based on the above evidence, it is highly likely that in the Ladefoged example, the coarticulation between *sack* and *Lou* would be greater if *sack Lou* were one word rather than two. Although we are not the first to suggest that these cues could help in speech segmentation and word recognition (Cole, Jakimik, & Cooper, 1978; Fougeron & Keating, 1997), we are the first to suggest that infants may use this information to help them determine word boundaries. Research on the identification of clause and phrase boundaries (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992) shows not only that infants are sensitive to very subtle acoustic cues to linguistic structure, but also that they use these cues in encoding information (Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000; Soderstrom, Jusczyk, & Kemler Nelson, 2000). If infants are similarly sensitive to coarticulatory cues, then the degree of coarticulation could help in word segmentation. Indeed, by 5 months, infants appear to be sensitive to at least some kinds of coarticulatory relationships in speech (Fowler, Best, & McRoberts, 1990).

To assess the relative contributions of speech and statistical cues in processing, we decided to use natural speech stimuli. In their investigation, Saffran, Aslin, et al. (1996) used synthetic speech. Unfortunately, it is difficult to achieve realistic coarticulation with such stimuli. At best, the algorithms available to Saffran, Aslin, et al.

simply interpolate between the acoustic parameters of adjacent syllables. To ensure that infants respond to naturally produced syllables, as they did to the synthetic syllables in the Saffran, Aslin, et al. study, Experiment 1 was a straight replication of their second experiment. Besides serving as a reference point for our other experiments, Experiment 1 bears on another issue. Saffran, Johnson, Aslin, and Newport (1999) found that infants segmented a tone stream based on statistics alone just as they had done for the speech stream in the original study (Saffran, Aslin, et al., 1996). Saffran et al. (1999) suggested that infants' ability to track the transitional probabilities between syllables may involve a domain-general as opposed to a language-specific learning mechanism. However, if infants in the original study did not perceive the stimuli as real speech, then Saffran et al.'s (1999) study simply proved exactly the same thing: infants are very sensitive to the statistical relationship between nonlinguistic stimuli. By using naturally produced syllables, we can determine whether infants did treat Saffran et al's. synthesized speech as real speech.

Two other experiments were designed to pit statistical cues against particular speech cues. Experiment 2 explores how infants respond to conflicting statistical and prosodic stress cues to word segmentation, whereas Experiment 3 investigates how they respond when statistical cues conflict with speech cues produced by coarticulation. Experiment 4 controls for the possibility that any added complexity to the familiarization stimuli leads to preferences in the test phase for familiar items over novel ones.

EXPERIMENT 1

Saffran, Aslin et al. (1996) exposed infants to a 2-min. stream of synthesized speech containing no cues to word boundaries other than the transitional probabilities between syllables. The continuous stream of speech was constructed by concatenating synthesized consonant–vowel (CV) syllables. Saffran et al. found that infants looked reliably longer toward part-words, indicating that they extracted words defined only by the statistical nature of the speech stream. The present experiment attempted to replicate Saffran et al.'s study using naturally produced syllables concatenated into the same sequences used in their study. If infants respond to naturally produced syllables in the same way as synthetically produced syllables, then infants in the present experiment should also listen significantly longer to the part-words during the test period.

Method

Participants. Sixteen 8-month-old infants from monolingual English-speaking homes were tested (9 males, 7 females; mean age 34 weeks 6 days; age range 32:6 to 37:6). Four other infants did not complete the study due to crying. The parents of all participants gave informed consent for this experiment and the following ones.

Stimuli. As in Saffran, Aslin et al. (1996), a continuous speech stream was created by repeatedly stringing together four tri-syllabic words (Language A: *pabiku, tibudo, golatu, daropi*). The only change from Saffran et al. was the use of naturally produced utterances rather than synthesized speech. Each of the 12 CV syllables used to create the four words was recorded individually, thereby avoiding coarticulation between syllables. These syllables were produced in a monotone (M = 268 Hz; SD = 3.6 Hz) by a female speaker (E.K.J.) from Upstate New York and were closely matched in length (M = .273 secs; SD = .022 s) and amplitude (M = 57 dB; SD = 3 dB). The 12 syllables were concatenated into tri-syllabic words (as above). No pauses occurred between syllables within words.

The four words were concatenated in random order to create five different blocks containing 12 words each (36 syllables per block) with the stipulation that a given word never occurred twice in a row. These five blocks were in turn concatenated to produce an approximately 2 min speech stream during which each word occurred 45 times. No pauses occurred between words. An orthographic representation of the resulting speech stream would be as follows: *pabikudaropitibudopabikugolatu* . . ..

The speech stream was closely matched to Saffran, Aslin et al.'s (1996) on all relevant acoustic parameters such as speaking rate (.273 s per syllable in the present study vs. .278 s per syllable in Saffran, Aslin et al.) and lack of prosodic cues to word boundaries. Their speech stream lasted for 2 min 30 s, whereas ours lasted 2 min 27 s. The only cue to word boundaries in the speech stream was the statistical nature of the language. Since no syllable appears in more than one word, the transitional probability between syllables within a word was always equal to 1. However, the transitional probability between syllables spanning word boundaries was always .33. For instance, the syllable *pa* occurred at the beginning of *pabiku*, and thus always preceded the syllable *bi*. By comparison, the syllable *ku* occurred at the end of a word and could be followed by the first syllable of any of the other three words in the language.

As in Saffran Aslin et al.'s study, a second artificial language was constructed in the same way as the first. The same syllable recordings used in the first language were combined in a new way to form four new words (Language B: *tudaro, pigola, bikuti, budopa*). These four words were concatenated, creating a second speech stream. The statistical structure of both languages was identical. If both languages were learned equally well, we could be assured that performance during the test phase was due to learning in the familiarization period rather than to general biases unrelated to the statistical structure of the language.

The same four test items were used for all infants (*pabiku, tibudo, tudaro,* and *pigola*). This was possible because the words of Language A formed the part-words of Language B and vice versa. In the test period, two items corresponded to words from the language to which the infant had been exposed, whereas the other two items corresponded to part-words. Part-words consisted of the last syllable of one word plus the first two of another (i.e., the words *golatu* and *daropi* were combined to form the part-word test item *tudaro*). Part-words were sequences of syllables that had actually occurred in the familiarization phase, albeit less often than sequences of syllables corresponding to words. This test situation corresponds directly to Experiment 2 in Saffran et al.'s investigation. The same syllable recordings used to form the speech streams were used to form the four test items.

Design. Infants were randomly assigned to hear one of the continuous streams of either Language A or Language B. Immediately following this familiarization, 12 test trials were presented (3 trials for each of the 4 test items, blocked and presented in random order). Half of the trials consisted of repetitions (up to 15 per trial) of part-words; the other half consisted of comparable repetitions of words.

Apparatus. The experiment was conducted in a three-sided test booth constructed out of white pegboard panels. To prevent infants from noticing the experimenter's movements behind the booth, the back of the center panel was lined with white cardboard. A small section of the center panel was left unlined to allow for the monitoring of infants' headturns through the pegboard's preexisting holes. The test booth had a red light and a loudspeaker mounted at eye level on each of the side panels and a green light mounted on the center panel in front of the infant. Directly below the center light, a 5-cm hole accommodated the lens of a video camera used to record each test session. A computer terminal and a response box were located behind the center panel, out of view of the infant. The response box, connected to a Macintosh computer, was equipped with a series of buttons that started and stopped the flashing center and side lights, recorded the direction and duration of head turns, and terminated a trial when the infant looked away for more than 2 s. Data regarding the direction and duration of head turns for each trial were stored in a computer data file. Computer software was responsible for the selection and randomization of the stimuli and for the termination of the test trials. The computer calculated the average listening times for the test items at the completion of each testing session.

All stimuli were recorded in a sound attenuated room using Kay Elemetric's Computerized Speech Lab (CSL). The recordings were made with a Shure microphone and sampled at a rate of 20 kHz via a 16-bit analog-to-digital converter. Digitized versions of the syllables were transferred to a Macintosh Quadra 650 for playback during the experiment. The computer controlled the presentation of the soundfiles during the experiment. The audio output was generated from the digitized waveforms of the samples. A 16-bit D/A converter was used to recreate the audio signal. The output was fed through anti-aliasing filters and a Kenwood audio amplifier (KA 5700) to the two 7-in. Cambridge Soundworks loudspeakers mounted on the side walls of the test booth.

Procedure. We used the same version of the Headturn Preference Procedure (HPP) used by Saffran, Aslin, et al. (1996). Each infant was tested individually while seated in a parent's lap. Both the parent and the experimenter listened to masking music over Peltor Aviation headphones (model 7050) to eliminate bias. The masker consisted of loud instrumental music, which had been recorded with few silent periods. At the beginning of the familiarization phase, the green light on the front panel of the testing booth began to blink. Once the infant had oriented toward the center light, one of the speech streams was presented without interruption from both loudspeakers (one located behind each side panel). During the familiarization, to keep the infants' interest, a blinking red light above one of the two loudspeakers (randomly selected) was lit and extinguished dependent on the infant's looking behavior. When the side light was extinguished, the central light blinked until the infant's gaze returned to center, then one of the side lights began to blink. In the familiarization phase, there was no contingency between lights and sound since the sound played continuously. The test phase followed the familiarization. Each of the 12 test trials began with the blinking center light. When the infant had turned in the direction of the center light, it was extinguished and the red light above the loudspeaker on one of the two side panels began to blink. When the observer judged that the infant had made a headturn of at least 30° toward the blinking sidelight, a button press initiated the presentation of one of the test items from the loudspeaker on the same side. A test item was repeated with a 500 ms ISI until the infant turned away from the blinking light for at least 2 consecutive seconds or until 15 repetitions had occurred. When this criterion was met, the computer extinguished the blinking sidelight, turned off the test stimulus, and turned on the central blinking light to begin another trial.

Results

Mean orientation times to the two types of test items (words and part-words) were calculated for each infant. Thirteen of the 16 infants had longer orientation times for the part-words. A mixed-design 2 (Language: A vs. B) X 2 (Test items: word vs. part-word) ANOVA revealed no significant effect of Language, $F(1, 14) = 2.56$, $p > .10$, or interaction between Language and Test items, $F(1,14) < 1.00$. However, as in Saffran et al. (1996), there was a significant main effect of Test items, $F(1,14) = 8.47$, $p < .02$. As shown in Fig. 1, the latter effect was attributable to longer average orientation times for part-words (8.25 s; $SD = 3.1$ s) than for words (6.49 s; $SD = 2.5$ s).

Discussion

As in Saffran, Aslin, et al. (1996), infants listened significantly longer to the novel part-words, demonstrating their ability to use statistical cues to discover word boundaries in continuous speech. Infants performed nearly identically in this experiment as they did in both the analogous synthesized speech (Saffran,

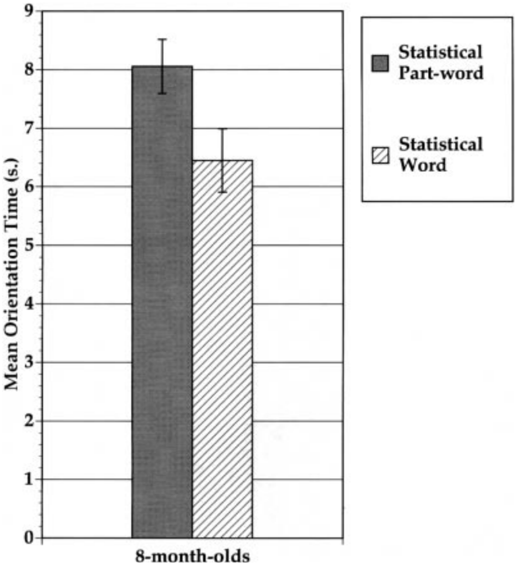

FIG. 1. Mean orientation times in seconds (and standard errors) to items in the test period that had appeared as either statistical words or part-words in the familiarization sequences from Experiment 1.

Aslin, et al., 1996) and the tone segmentation (Saffran et al., 1999) tasks. Regardless of whether the original synthesized stimuli in Saffran, Aslin, et al. (1996) were perceived and processed as speech, the present study provides further evidence that infants employ the same general statistical learning mechanism to parse both speech and nonspeech stimuli. Critically, for the current study, we have replicated Saffran, Aslin, et al.'s results using natural speech. Because infants never hear speech with no cues to word boundaries besides the transitional probabilities between syllables, it is not clear how prominent a role statistical cues actually play in word segmentation. Given that infants can segment our speech stream using statistics alone, we can explore the strength of statistical cues by adding conflicting speech cues.

EXPERIMENT 2

As noted earlier, no segmentation cue is completely foolproof for English. To explore the strength of transitional probabilities relative to speech cues, we added a conflicting speech cue to the familiarization stream. Specifically, stress cues were pitted against statistics. The tendencies of English-learning infants to segment fluent speech at the onsets of stressed syllables are well documented (Echols, Crowhurst, & Childers, 1997; Houston, Jusczyk, Kuijpers, Coolen, & Cutler, 2000; Jusczyk, Houston, et al., 1999; Morgan & Saffran, 1995). However, previous investigations have not directly pitted stress cues against statistical cues to word boundaries. Saffran, Newport, and Aslin (1996) did examine possible interactions between statistical and stress cues in a study with adults. Prosodic cues enhanced adults' abilities to segment an artificial language with no cues to word boundaries other than transitional probabilities between syllables. Particular syllables were lengthened in the familiarization sequence as a cue to stress. Adults performed best when lengthening occurred on word-final syllables, but could segment the speech stream regardless of which syllable in the word had been lengthened. This finding suggests that statistics outweighed prosody for adults. Might the same hold for infants?

In the present experiment, each time a part-word test item appeared in the familiarization sequence, its first syllable was replaced with a stressed version. Because English learning infants have a tendency to perceive stressed syllables as the beginning of a word, stress cues suggest one potential word boundary while statistical cues suggest another. There are three possible outcomes to this experiment, all of which are interesting. If infants orient longer to part-words, as in Experiment 1, then statistics would seem to outweigh stress. If infants' looking times to words and part-words do not differ significantly, then both cues are relatively equally strong. Assuming there are no additional cues to word boundaries present, the conflicting cues cancel each other out, and infants cannot segment the speech stream. The third possible outcome is that infants orient longer toward the words. This last pattern would suggest that stress cues outweigh statistical cues. Past research (Jusczyk, Houston, et al., 1999) leads us to predict that the third outcome will be observed: stress cues will overwhelm the statistical cues.

Method

Participants. Sixteen 8-month-olds from monolingual English-speaking homes were tested (5 males, 11 females; mean age 35 weeks 2 days; range 33:5 days to 36:5). Four additional infants did not complete the study due to crying.

Stimuli. As in Experiment 1, an approximately 2.5-min continuous speech stream was created by repeatedly stringing together four trisyllabic words (Language A: *pabiku, tibudo, golatu, daropi*). The 12 CV syllables used to create these words were produced in the same way and by the same female speaker as in Experiment 1. The 12 syllables were closely matched in average pitch (M = 266 Hz; SD = 5.9), amplitude (M = 61 dB; SD = 1.6), and duration (M = .267 s.; SD = .02). In addition, the same speaker recorded a stressed version of the first syllable of the four test words (i.e., *pabiku, tibudo, tudaro*, and *pigola*). Stressed syllables were higher in average pitch (M = 287 Hz) and amplitude (M = 65 dB), and slightly longer in duration (M = 306). We chose to re-record rather than re-use the 12 unstressed CV syllables from Experiment 1 to minimize voice quality differences between the stressed and the unstressed syllables.

The four words were concatenated to form a speech stream in precisely the same manner as in Experiment 1. However, the speech stream used in Experiment 2 differed from that of Experiment 1 in that each time one of the two part-word test items occurred within the familiarization speech stream, its first syllable was replaced by its stressed counterpart. Orthographically, the speech stream was as follows: *tibudogolaTUdaropitibudodaroPIgolatu* . . . (stressed syllables are capitalized.) Thus, the speech stream contained two conflicting segmentation cues: transitional probabilities between syllables and stress. The former cue suggests the same segmentation seen in Experiment 1, whereas the latter indicates an alternative segmentation (i.e., statistics suggest that *TU* ends a word, whereas stress suggests that *TU* begins a new word). Each of the 2 stressed syllables occurred 15 times during the familiarization period.

As in Experiment 1, a second artificial language was created to counterbalance the first. The same syllable tokens used in the first artificial language were combined in a different way to form four new words (Language B: *tudaro, pigola, bikuti, budopa*). Orthographically, the speech stream was as follows: *bikuTIbudopatudarobudoPAbikuti* . . . The words of Language A formed the part-words of Language B (and vice versa), allowing the same four nonstressed test items to be used for all infants (*pabiku, tibudo, tudaro*, and *pigola*). To provide a strong test of whether stress cues override statistical cues, the four test items were produced from the *unstressed* syllable tokens. The use of unstressed test items actually disfavors the stress segmentation hypothesis because the stress that is present on the first syllable of the part-words in the familiarization sequence is absent from the part-words in the test phase. Hence, infants might find the part-words novel in the test phase simply because they lacked stress.

Procedure, apparatus, and design. The procedure, apparatus, and design were identical to those of Experiment 1. Infants were randomly assigned to hear one of the two speech streams. All infants heard the same test items. Conflicting segmentation cues were presented *only* in the familiarization phase. The first syllable of the part-word test items had been stressed during the familiarization period, but not during the test period.

Results

Mean orientation times to the two types of test items (word and part-words, as defined by the statistical structure or the language) were calculated for each infant (see Fig. 2). Eleven of the sixteen infants had longer orientation times for the statistical words. A 2 (Language: A vs. B) X 2 (Test items: word vs. part-word) ANOVA revealed no significant effect of Language, $F(1, 14) < 1.00$ or interaction between Language and Test items, $F(1, 14) < 1.00$. However, there was a significant main effect of Test items, $F(1, 14) = 7.41$, $p < .02$. In contrast to Experiment 1, the latter effect was attributable to longer average orientation times for words (8.0 s; $SD = 2.5$ s) than for the part-words (6.59 s; $SD = 1.6$ s). The

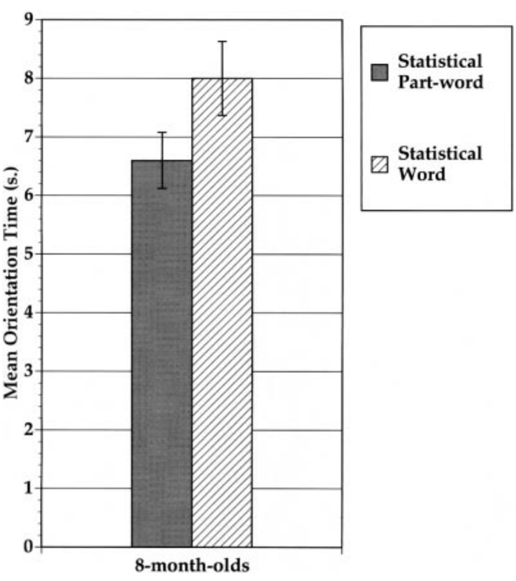

FIG. 2. Mean orientation times in seconds (and standard errors) to items in the test period that had appeared as either statistical words or part-words (with stressed initial syllables) in the familiarization sequences from Experiment 2. Note that all test items were unstressed.

data from the first two experiments were combined and analyzed in another ANOVA to compare the part-word preference found in Experiment 1 with the word preference found in Experiment 2. This analysis revealed only the expected significant interaction between the performance of infants in Experiments 1 and 2 [$F(1, 30) = 16$, $p < .001$].

Discussion

In the present experiment, two competing cues to word segmentation were pitted against each other: stress and statistics. In segmenting the familiarization sequence, the infants relied more heavily on the stress cue to indicate word onsets than on the statistical cue relating to the transitional probabilities of successive syllables. The addition of stress as a conflicting segmentation cue in the familiarization sequence resulted in a reversal of looking time preferences from what was observed in the previous experiment. Thus, in the test phase of the present experiment, infants looked significantly longer toward the words than the part-words. Moreover, the statistical words were treated as novel, despite the fact that the part-words in the test phase did not fully match the ones heard in the familiarization sequence (i.e., the former did not include a stressed first syllable). Consequently, it appears that although statistical cues are sufficient to segment a simple artificial language, 8-month-olds weigh speech cues such as stress more heavily.

Neither the present findings, nor the original Saffran, Aslin, et al. (1996) results reveal the length of the word that infants segment from the speech stream. This is because the test items in these studies were always tri-syllabic patterns. Hence, there is no information available as to whether infants may actually favor shorter or longer strings than the tri-syllabic items that have been used in the present paradigm. However, evidence from another recent investigation indicates that English learners at this age certainly do have the capacity to segment tri-syllabic words from fluent speech (Houston, Santelmann, & Jusczyk, submitted). Although determining the exact length of the words that infants extract from the familiarization stream in the present experiment would be interesting, it is not obvious how it would directly bear on the main issue of the present investigation—whether infants give greater weight to speech or statistical cues. Hence, we shift our attention back to the main issue of the present investigation: how do infants respond when speech and statistical cues to word boundaries conflict?

EXPERIMENT 3

Most discussions of word segmentation by infants have focused on the cues used to identify the onsets of potential words in the speech stream. Yet, because segmentation cues are probabilistic in nature, errors will sometimes occur. For instance, if English learners rely primarily on prosodic stress cues, they will sometimes mis-segment words beginning with weak syllables. An example is the finding of Jusczyk, Houston, et al. (1999) that 7.5-month-olds segmented the sequence *taris* from a speech context in which the word *guitar* was consistently followed by *is*. In contrast, Houston et al. (submitted) found that English learners of the same age did not mis-segment certain three-syllable words with a strong–weak–strong stress pattern, such as *parachute*. Were the latter infants simply imposing a metrical segmentation strategy, they should have segmented *parachute* into two words, *para* and *chute*. These findings raise an interesting issue: What information do infants use to determine whether syllables following a strong syllable belong to the same word or another word? One possibility is that statistical cues, such as transitional probabilities, indicate the ends of words. However, speech cues produced by coarticulation could also potentially indicate whether to attach following syllables to ones associated with onsets of words. To compare the relative importance of speech cues produced by coarticulation and statistical cues, we put these cues in conflict with each other.

We recorded coarticulated versions of the part-word test items, closely matched to the original ones in length, pitch, and amplitude. Each time one of the two part-words occurred in the familiarization sequence, it was replaced by its coarticulated counterpart. If coarticulation causes the three syllables corresponding to a part-word to cohere, then this speech cue will compete with the statistical cue to word boundaries. Once again, the two sets of cues yield conflicting segmentations. If statistical cues outweigh speech cues produced by coarticulation, infants should behave as in Experiment 1—i.e., they should perceive the part-words as novel. If, as was true for stress in Experiment 2, speech cues produced by coarticulation outweigh statistics, infants should perceive the statistical words as novel and listen longer to these than to part-words in the test period.

Method

Participants. Sixteen 8-month-old infants from monolingual English-speaking homes were tested (8 males, 8 females; mean age 34 weeks 5 days; age range 32:6 to 36:1). Six other infants did not complete the study due to crying.

Stimuli. As in the previous experiments, a 2.5-min continuous speech stream was created by repeatedly stringing together four tri-syllabic words (Language A: *pabiku, tibudo, golatu, daropi*). The 12 nonstressed CV syllables from Experiment 2 were used. In addition, we recorded coarticulated versions of the part-word test items of Experiment 1. The coarticulated part-words were closely matched to their non-coarticulated part-word test item counterparts in length (coarticulated part-word $M = .774$ s; $SD = .04$ s vs. noncoarticulated part-word $M = .8$ s; $SD = .04$ s), pitch (coarticulated part-word $M = 273$ Hz; $SD = 2.2$ vs. noncoarticulated part-word $M = 267$ Hz; $SD = 1.2$), and amplitude (coarticulated part-word $M = 61$ dB; $SD = 2.1$ vs. noncoarticulated part-word $M = 61$ dB; $SD = .99$). The coarticulation cue was subtle enough that naïve adults had difficulty identifying the coarticulated tri-syllabic utterances in the speech stream. Unlike Experiment 2, no stress cues appeared in either the familiarization or the test materials.

The four words were concatenated into a continuous stream of speech as in Experiment 1. However, the speech stream used in Experiment 3 differed from that of Experiment 1 in that each time one of the two part-word test items occurred within the familiarization speech stream, it was replaced by its coarticulated counterpart. If we assume that coarticulating a string of syllables makes them more likely to cohere as a word, then this speech stream contained two conflicting cues to word boundaries: transitional probabilities between syllables and speech cues produced by coarticulation. The former cue marks the same segmentation as in Experiment 1, whereas the latter cues yield an alternative segmentation (i.e., statistics suggest that *golatu* is a word, whereas coarticulation suggests that *tudaro* is a word). An orthographic representation of the resulting speech stream would be as follows: *golatudaroitibudodaropigolatu* . . . (coarticulated part-words are underlined). Each of the two coarticulated part-words occurred 15 times during the familiarization sequence.

Again, a second artificial language was constructed like the first. The syllables from the first artificial language were concatenated in a different way to form four new words (Language B: *tudaro, pigola, bikuti, budopa*.) An orthographic version of the stream is as follows: *pigolabikutibudopatudarobudopabikuti*. . . . The underlined coarticulated part-words in Language A were noncoarticulated words in Language B, and vice versa, allowing the same test items to be used for all infants (*pabiku, tibudo, tudaro,* and *pigola*). To provide a strong test of whether speech cues produced by coarticulation override statistical cues, the four test items were produced from the *noncoarticulated* syllable tokens. Again, the choice of noncoarticulated test items disadvantages the speech cues relative to the statistical cues. The coarticulation present for the part-words in the familiarization sequence is absent from the part-words in the test phase. Hence, infants might find the part-words novel in the test phase simply because they lacked coarticulation.

Procedure, apparatus, and design. The procedure and apparatus were identical to Experiment 1. Infants were randomly assigned to one of the two speech streams. All infants heard the same test items. However, unlike in Experiment 1, the familiarization contained conflicting cues. The nonarticulated part-word test items had occurred as a coarticulated trisyllabic utterance during the familiarization.

Results

Mean orientation times to the words and part-words were calculated for each infant (see Fig. 3).

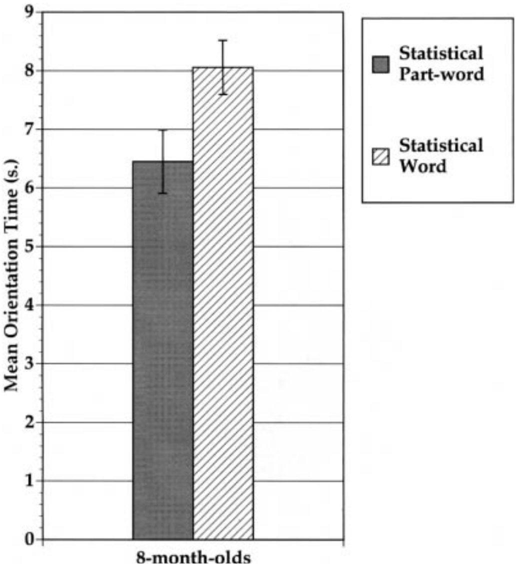

FIG. 3. Mean orientation times in seconds (and standard errors) to items in the test period that had appeared as either statistical words or part-words (with coarticulation present between syllables) in the familiarization sequences from Experiment 3. Note that all test items were noncoarticulated.

Twelve of the 16 infants had longer orientation times for the statistical words. A 2 (Language: A vs. B) X 2 (Test items: word vs. part-word) ANOVA revealed no significant effect of Language, $F(1,14) < 1.00$, or interaction between Language and Test items, $F(1,14) < 1.00$. However, there was a significant main effect of Test items, $F(1,14) = 7.57$, $p < .02$. As in Experiment 2, the latter effect was attributable to longer average orientation times for statistical words (8.06 s; $SD = 2.2$ s) than for the part-words (6.45 s; $SD = 1.8$ s). The data from the Experiments 1 and 3 were combined and analyzed in another ANOVA to compare the part-word preference found in Experiment 1 with the word preference found in Experiment 3. This analysis revealed only the expected significant interaction between the performances of infants in Experiments 1 and 2 [$F(1, 30) = 19$, $p < .001$].

Discussion

In the present experiment, speech cues produced by coarticulation were pitted against statistical cues. The addition of coarticulation as a conflicting segmentation cue reversed the looking time preferences observed in Experiment 1, i.e., infants in the present experiment looked significantly longer toward the words than toward the part-words. Again, this preference emerged despite the fact that the part-word test items lacked the speech cues produced by coarticulation in the familiarization stream.

Thus, coarticulatory cues can outweight statistical cues when infants segment words from speech. Earlier, we suggested that the strength of coarticulation between phonemes within words may be stronger than for phonemes across word boundaries. Certainly, one interpretation of our results is that when coarticulatory cues are available, they help infants to determine which syllables belong together in the same word.

Along with the results of Experiment 2, the present results provide a second instance of a speech-based cue apparently overriding a conflicting statistical cue to word boundaries. However, could there be an alternative explanation for the results obtained in Experiments 2 and 3? Might infants have favored the words over the part-words during the test phase, not because they used the speech cues to segment words in the familiarization stream, but because the addition of stress (Experiment 2) or coarticulation (Experiment 3) to the familiarization stream made it more complex? By this alternative explanation, the greater complexity of the familiarization stream (where stress or coarticulation occurred for some items but not others) made it harder to learn to segment the words, thus leading infants to show a familiarity effect for the statistical words. Although we were skeptical about this alternative account, we conducted Experiment 4 to determine if a more complex familiarization stream inevitably results in infants displaying a familiarity preference for the statistical words.

EXPERIMENT 4

Our claim that speech cues outweigh conflicting statistical cues is based on the reversal of the preferences that infants had for the test items in Experiments 2 and 3, as compared to Experiment 1. As in Saffran, Aslin, et al. (1996), infants in Experiment 1 displayed a novelty effect—they listened significantly longer to the part-words than to the statistical words. By comparison, infants in Experiments 2 and 3 showed a reversed pattern of preferences—they listened significantly longer to the words than to the part-words. Because the test procedure was essentially unchanged, we assumed that the addition of the conflicting speech cues was responsible for the flip in preferences that occurred during the test phase. In contrast, by the alternative account above, the greater complexity of the familiarization stream made it harder to learn, causing infants to display a familiarity effect for the statistical words instead of a novelty effect for the part-words (see Hunter & Ames, 1989, for a general account of novelty and familiarity effects). With the highly varied stimulus materials typically used in our previous studies, infants display familiarity effects for the test items (e.g., Houston et al., in press; Jusczyk & Aslin, 1995; Jusczyk, Hohne, et al., 1999; Jusczyk, Houston, et al., 1999; Mattys & Jusczyk, 2001). However, even with the occasional stressed syllable or coarticulated sequence, the stimuli in the present study are much less varied, lacking, as they do, the range of pitch, amplitude, and duration changes typically present in fluent speech.

To test the alternative account of our findings, we exposed infants to a familiarization stream similar to that of Experiment 3. The familiarization stream contained an equal number of coarticulated items as in Experiment 3, but the location of the sequences coincided with word boundaries as defined by the statistical cues. Thus, rather than providing conflicting information about the locus of a word boundary, the speech cues produced by coarticulation were consonant with the statistical cues. If a more complex familiarization stream is the key factor in determining infants' response patterns, they should show a familiarity effect, just as in Experiment 3. However, if the pattern of responding in Experiments 2 and 3 is attributable to the conflict between speech and statistical cues to word boundaries, then in the absence of such conflicting cues, infants should display a novelty effect as in Experiment 1.

Method

Participants. Twenty 8-month-old infants from monolingual English-speaking homes were tested (12 males, 8 females; mean age 34 weeks 2 days; age range 32:3 to 36:5). Six other infants did not complete the study due to crying.

Stimuli. As in Experiment 3, two 2.5-min continuous speech streams for familiarization were created by repeatedly stringing together four tri-syllabic words (Language A: *pabiku, tibudo, golatu, daropi*; Language B: *tudaro, pigola, bikuti, budopa*). The CV syllables from Experiment 3 were used to create the words. The coarticulated versions of the words, which were only in the familiarization sequence, were also identical to the coarticulated versions of the words in Experiment 3. Owing to the structuring of Languages A and B, the coarticulated items used in Experiment 3 for Language A were used for Language B in the present experiment, and the ones for Language B in Experiment 3 were used in Language A of the present experiment.

The four words were concatenated into a continuous stream of speech as in Experiment 1. However, the speech stream used in Experiment 4 differed from that of Experiment 3 because the coarticulated versions of the test items that were substituted in the stream replaced words, instead of part-words. Hence, the familiarization stream had two redundant cues to word boundaries: transitional probabilities between syllables and speech cues produced by coarticulation (e.g., for Language A, both the coarticulatory and the statistical cues indicate that *golatu* is a word and that *tudaro* is a part-word). Each of the 2 coarticulated part-words occurred 15 times during the familiarization sequence.

The coarticulated words of Language A formed the noncoarticulated words of Language B, and vice versa, allowing the same test items to be used for all infants (*pabiku, tibudo, tudaro,* and *pigola*). As in Experiment 3, the four test items were produced from the *noncoarticulated* syllable tokens. Hence, the procedural difference between Experiments 3 and 4 was whether the coarticulatory cues present in the familiarization sequence were conflicting or redundant with the statistical cues to word boundaries.

Procedure, apparatus, and design. The procedure and apparatus were identical to Experiment 1. Infants were randomly assigned to one of the familiarization streams. All infants heard the same test items. Unlike in Experiment 1, the familiarization contained redundant cues to word boundaries. The noncoarticulated word test items had occurred as a coarticulated tri-syllabic utterance during the familiarization.

Results

Mean orientation times to the words and part-words were calculated for each infant (see Fig. 4). Fifteen of the 20 infants had longer orientation times for the part-words. A 2 (Language: A vs. B) X 2 (Test items: word vs. part-word) ANOVA revealed no significant effect of Language, $F(1, 18) = 1.13$, $p > .30$, or interaction between Language and Test items, $F(1,18) < 1.00$. However, there was a significant main effect of Test items, $F(1,18) = 6.17$, $p < .03$. As in Experiment 1, the latter effect was attributable to longer average orientation times for part-words (7.54 s; $SD = 2.7$ s) than for statistical words (6.00 s; $SD = 1.92$ s). The data from Experiments 3 and 4 were analyzed in another ANOVA to compare the part-word preference

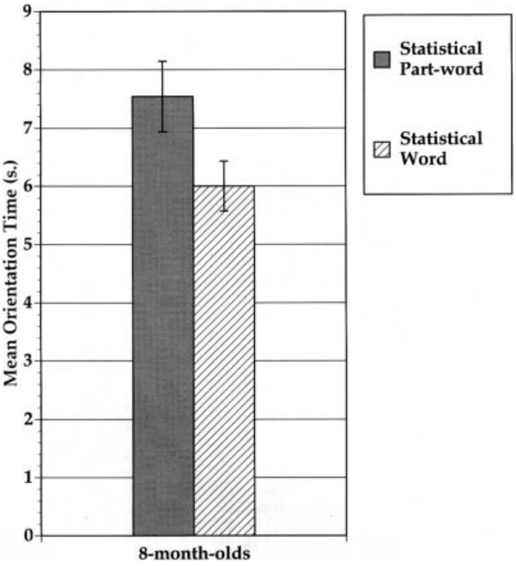

FIG. 4. Mean orientation times in seconds (and standard errors) to items in the test period that had appeared as either statistical words or part-words in the familiarization sequences from Experiment 4 in which coarticulatory cues coincided with statistical cues to word boundaries. Note that all test items were noncoarticulated.

found in Experiment 1 with the word preference found in Experiment 3. This analysis revealed only the expected significant interaction between the performances of infants in Experiments 3 and 4 [$F(1, 34) = 14.02$, $p < .001$].

Discussion

As in Experiment 3, the familiarization sequence contained speech cues produced by coarticulation along with statistical cues to word boundaries. Unlike Experiment 3, in which infants listened significantly longer to the words during the test phase, infants in the present experiment displayed the opposite preference, listening significantly longer to the part-words. The critical difference in the two experiments was whether the speech cues produced by coarticulation conflicted (Experiment 3) or coincided (Experiment 4) with the statistical cues to word boundaries. This difference was sufficient to cause the flip in preferences that was observed. This finding undermines the alternative account of the results found in Experiments 2 and 3; namely, that the greater complexity of the familiarization sequence *per se* caused the flip in the preferences in those experiments. The familiarization sequence in Experiment 4 had complexity equal to that of Experiment 3. There were 15 occurrences of each of two coarticulated tri-syllabic sequences and 150 occurrences of noncoarticulated tri-syllabic sequences in the familiarization streams of both experiments. Yet, infants in the two experiments had opposite preferences. Thus, the results bolster the account originally offered for the results of Experiments 2 and 3: the flip in preferences stems from infants' greater reliance on speech cues than on statistical cues in segmenting the familiarization stream.

GENERAL DISCUSSION

The main goal of the present study was to investigate how 8-month-olds integrate and weight conflicting segmentation cues. We know that infants must use multiple sources of information to segment the speech stream. Specifically, we know that 8-month-olds are sensitive to both prosodic and statistical cues to word boundaries. Furthermore, we suspected that speech cues produced by coarticulation within words might help hold the syllables of a word together, thus facilitating word segmentation.

Our first experiment successfully replicated Saffran, Aslin, et al.'s results using natural speech. These results provide further evidence that infants' ability to track the transitional probabilities between syllables may involve a domain-general as opposed to a language-specific learning mechanism. Our replication with natural speech also allowed us to extend Saffran, Aslin, et al.'s findings by adding speech cues to the speech stream.

Experiments 2 and 3 added a conflicting speech cue to the speech stream. We first pitted stress and then speech cues produced by coarticulation against statistics. In both cases, the presence of a competing speech cue reversed infants' listening preferences for the words and part-words. Our interpretation of these findings was that, for 8-month-olds, speech cues to segmentation carry more weight than statistical cues. Experiment 4 investigated an alternative account, namely that the reversals in preferences in Experiments 2 and 3 were simply due to a more complex familiarization stream, which made it harder for infants to learn the word patterns and thus led to a familiarity effect for the words. The findings from Experiment 4 rule out such an account. The familiarization streams in this experiment were of equal complexity to those of Experiment 3, yet now infants displayed a preference for the part-words over the words. The critical difference responsible for the shift in preference between Experiments 3 and 4 was whether the speech cues produced by coarticulation conflicted (Experiment 3) or coincided (Experiment 4) with word boundary locations indicated by the statistical cues.

In Experiment 2, 8-month-olds relied more heavily on stress cues than transitional probabilities in segmenting words from the speech stream. This result contrasts with the finding that adults can parse an artificial language regardless of whether a prosodic cue added to the speech stream conflicted with the statistical cue (Saffran, Newport, et al., 1996). However, the prosodic cue that Saffran et al. used differed somewhat from the stress cue used here. In English, word-final syllables tend to be lengthened (Klatt, 1975, 1976). Saffran, Newport, et al. compared adults' performance when they lengthened the initial, as opposed to the final, syllable of the words in the familiarization sequence. Even when the initial syllables were lengthened, adults still parsed the speech stream at the points predicted by the transitional probabilities (albeit not as well as when lengthening occurred on final syllables). One possible account for the different outcomes in the two studies is that stress is typically indicated by changes in pitch and amplitude as well as by increases in duration. Combinations of these cues may be more significant indicators of word boundaries than is syllable lengthening by itself. Alternatively, infants and adults may weigh conflicting segmentation cues differently. Although prosodic stress cues are important in word segmentation in languages such as English (Cutler, 1994; Cutler & Butterfield, 1992; Cutler & Norris, 1988) and Dutch (Vroomen & de Gelder, 1995; Vroomen, van Zon, & de Gelder, 1996), a critical role has also been ascribed to lexical knowledge in word recognition by adults (Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994; Norris, McQueen, & Cutler, 1995; Vitevitch & Luce, 1998). Eight-month-olds know very few words and are not likely to use lexical knowledge in segmenting speech. Thus, stress cues may occupy a more prominent role in word segmentation than they do when the lexicon develops. The relative weighting given to particular types of word segmentation cues may change as language is acquired.

In any case, when beginning to segment the sound patterns of words from fluent speech, English learners appear to rely heavily on stress-based cues (Jusczyk, 1997, 1999). Infants are more successful in segmenting targets with trochaic stress than iambic stress (Echols et al., 1997; Jusczyk, Houston, et al., 1999; Morgan & Saffran, 1995). Moreover, 7.5-month-olds use stressed syllables as indicators of word onsets, even when this causes them to mis-segment words from fluent speech (Jusczyk, Houston, et al., 1999). English-learning infants also appear to gain access to stress-based cues at a younger age than to other speech cues such as phonotactics (around 9 months [Mattys et al., 1999]) or context-sensitive allophones (around 10.5 months [Jusczyk, Houston, et al., 1999]). Finally, when stress cues conflict with other word segmentation cues, such as transitional probabilities (the present study) or phonotactic cues (Mattys et al., 1999), infants seem to give stress more weight. One future goal is to gain a fuller appreciation of changes that may occur in the relative weighting and integration of the various cues for segmentation, especially as the lexicon begins to develop and grow.

How do 8-month-olds come to rely on stress to locate the beginnings of words? There is still debate about the origins of the trochaic bias. Some have suggested that the bias may be innate (Allen & Hawkins, 1980), whereas others have claimed that it develops from the linguistic input that infants receive (Jusczyk et al., 1993; Morgan, 1994). There is no sufficient cross-linguistic evidence regarding language acquisition to decide this issue. However, because languages vary greatly in their rhythmic structures and the consistency of their word stress patterns, we suspect that the trochaic bias most probably develops. Were a trochaic bias innate, infants learning languages without a trochaic rhythmic organization, such as French, would have to overcome their initial bias. How might infants learn the predominant rhythm of native language words? One suggestion is that infants acquire the appropriate bias by noting the patterns of words that are most likely to be spoken in isolation (Jusczyk, Houston, et al., 1999). For young infants, this will likely include their own names, or nicknames, and many diminutive terms. In English, names and nicknames often have trochaic patterns (Cutler, McQueen, & Robinson, 1990). Similarly, diminutive versions of adult words in English tend to follow a trochaic pattern (e.g., doggie, mommy, daddy, birdie). Furthermore, previous research shows that infants are very sensitive to the rhythmic properties of language from birth (DeCasper & Spence, 1986; Mehler et al., 1988; Nazzi et al., 1998). Hence, it is not surprising that English learners rely on stress when beginning to segment words.

The results of Experiment 3 are interesting on many levels. First, although transitional probabilities between syllables are sufficient to segment an artificial language, infants seem to give greater weight to subtle speech cues produced by coarticulation. Second, to our knowledge, this study is the first to consider differential coarticulation as a possible cue to word boundaries. Our results justify further exploration of the role of coarticulation in segmentation. Although much research exists on the production of coarticulation by young children (Nittrouer et al., 1993; Repp, 1986; Smith & McLean-Muse, 1986; Turnbough, Hoffman, Danilog, & Absher, 1985), little is known about how infants perceive coarticulation (cf. Fowler et al., 1990). A tight relationship has been observed between prosody, coarticulation, and rate of speech (Dalby, 1984; de Jong, Beckman, & Edwards, 1993; Liberman et al., 1967). In fast speech, not only are segments shortened, but the overlap between adjacent articulations also increases (Browman & Goldstein, 1990; Gay, 1981). Also, the faster the speech, the more reduced the magnitude and number of phonetically marked phrases (Fougeron & Jun, 1998). In child-directed speech, which is typically slower than adult-directed speech, prosodic cues tend to be more robust (Kelly & Martin, 1994). Might coarticulatory cues also be more exaggerated in child-directed speech? If so, then sensitivity to coarticulation could facilitate language acquisition by helping infants to segment and group important units in the input.

Exploring the development of language-specific as opposed to language-general patterns of coarticulation, and their possible roles in language acquisition, would also be useful. Various languages have different patterns of coarticulation that often reflect the phonetic contrasts that are emphasized (see Manuel, 1999, for review). Thus, Boyce (1990) found that English and Turkish speakers differ in how they coarticulate rounded vowels within a word. These differences appear to be due to the vowel harmony patterns seen in Turkish, but not in English. Variations in coarticulation in different languages may draw attention to phonetic distinctions that are meaningful in the native language.

Our aim is not to downplay the significance of statistical cues to segmentation. In many cases, statistical cues will work in concert with speech cues in word segmentation. Moreover, many of the speech cues discussed are probabilistic and could be framed as statistical cues. For instance, phonotactics is simply the probability of one sound following another within the speech stream. Similarly, detecting a predominant stress pattern in isolated words demands some statistical abilities: Infants need to track the distributional frequencies of stressed syllables in such utterances. Therefore, our goal was to investigate how infants weigh and integrate contradictory cues to segmentation. The present results suggest that 8-month-olds rank various speech cues more heavily than a statistical cue regarding the transitional probabilities of certain syllable sequences. The fact that such speech cues have special significance for infants fits with an innately guided learning view of language acquisition (Jusczyk, 1997; Jusczyk & Bertoncini, 1988). Infants may have a bias to selectively attend to certain properties in the acoustic signal, which then guides the direction of learning and enables them to master complex patterns rapidly. By providing infants with a means to begin segmenting speech into coherent units such as words, speech cues such as prosodic stress and those produced by coarticulation might help learners to discover other patterns within such units (e.g., phonotactic sequences and context-sensitive allophones) that serve to refine word segmentation processes. Further studies comparing how infants weight these different types of cues as their lexicons grow and develop should clarify how infants succeed at integrating the various types of word segmentation cues.

REFERENCES

Allen, G., & Hawkins. S. (1980). Phonological rhythm: Definition and development. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: vol. 1: Production* (pp. 227–256), New York: Academic Press.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of probability statistics by 8-month-old infants. *Psychological Science* **9,** 321–324.

Bloomfield, L. (1933). *Language*. New York: Holt.

Boyce, S. E. (1990). Coarticulatory organization for lip rounding in Turkish and English, *Journal of the Acoustical Society of America*, **88,** 2584–2595.

Brent, M. R. (1999). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Science*, **3,** 294–300.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, **61,** 93–125.

Browman, C. P., & L. Goldstein, (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and the physics of speech* (pp. 341–338), Cambridge, UK: Cambridge University Press.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, **24,** 209–244.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, **26,** 173–199.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: Abottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, **33,** 111–153.

Church, K. W. (1987). *Phonological parsing in speech recognition*. Dordrecht: Kluwer Academic.

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 136–163), Hillsdale, NJ: Erlbaum.

Cole, R., Jakimik, J., & Cooper, W. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society for America*, **64,** 44–56.

Cooper, A. M. (1991). Laryngeal and oral gestures in English/p, t, k/. Proceedings of XIIth International Congress of Phonetic Sciences, Aix-en-Provence, Vol. 2, pp. 50–53.

Cutler, A. (1994). Segmentation problems, rhtyhmic solutions. *Lingua*, **92,** 81–104.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, **31,** 218–236.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, **2,** 133–142.

Cutler, A., McQueen, J., & Robinson, K. (1990). Elizabeth and John: Sound patterns of men's and women's names. *Journal of Linguistics*, **26,** 471–482.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25,** 385–400.

Cutler, A., & Norris, D. G.. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experiment. Psychology: Human Perception and Performance*, **14,** 113–121.

Dalby, J. M. (1984). Phonetic structure of fast speech in American English. *Dissertation Abstracts International*, **46-01A,** 138.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborn's perception of speech sounds. *Infant Behavior and Development*, **9,** 133–150.

de Jong, K., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and Speech*, **36,** 197–212.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, **24,** 423–444.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, **36,** 202–225.

Fougeron, C., & Jun, S. (1998). Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics*, **26,** 45–69.

Fougeron, C., and Keating, P. (1996). Variations in velic and lingual articulation depending on prosodic position: Results for two French speakers. *UCLA Working Papers in Phonetics*, **92,** 88–96.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, **101,** 3728–3740.

Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences following stop consonants. *Perception & Psychophysics*, **48,** 559–570.

Friederici, A. D., & Wessels, J. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception & Psychophysics*, **54,** 287–295.

Fujimura, O. (1990). Methods and goals of speech production research. *Language and Speech*, **33,** 195–258.

Gay, T. J. (1981). Mechanisms in the control of speech rate, *Phonetica*, **38,** 148–158.

Harris, Z. S. (1951). *Methods in structural linguistics*. Chicago: University of Chicago Press.

Hayes, J. R., & Clark, H. H. (1970). Experiments in the segmentation of an artificial speech analog. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 221–233), New York: Wiley.

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 (1955). A manual of phonology. Baltimore, MD: Waverly Press.

Hohne, E. A., & Jusczyk, P. W. (1994). Two month-old infants' sensitivity to allophonic differences. *Perception & Psychophysics*, **56,** 613–623.

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review* **7,** 504–509

Houston, D. M., Santelman, L. M., & Jusczyk, P. W. English-learning infants' segmentation of trisyllabic words from fluent speech. Manuscript submitted for publication.

Hunter, M. A., & Ames, E. W. (1989). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (vol. 5, pp. 69–93), Norwood, NJ: Ablex.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W. (1999). How infants begin to extract words from fluent speech. *Trends in Cognitive Science*, **3,** 323–328.

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

Jusczyk, P. W., & Bertoncini, J. (1988). Viewing the development of speech perception as an innately guided learning process. *Language and Speech*, **31,** 217–238.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Preference for the predominant stress patterns of English words. *Child Development*, **64,** 675–687.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, **32,** 402–420.

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,** 252–293.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, **61,** 1465–1476.

Jusczyk, P. W., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, **39,** 159–207.

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.

Kelly, M. H., & Martin, S. (1994). Domain general abilities applied to domain specific tasks: Sensitivity to probabilities in perception, cognition, and language. *Lingua*, **92,** 105–140.

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.

Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics*, **27,** 23–54.

Ladefoged, P. (1993). *A course in phonetics*. Fort Worth, TX: Harcourt Brace.

Lehiste, I. (1960). *An acoustic-phonetic study of internal open juncture*. New York: S. Karger.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. G. (1967). Perception of the speech code. *Psychological Review*, **74,** 431–461.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear & Hearing*, **19,** 1–36.

Manuel, S. (1999). Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts. In W.J. Hardcastle, & N. Hewlett

(Eds.), *Coarticulation: theory, data and techniques* (pp. 179–198), Cambridge, UK: Cambridge University Press.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. Cognition, **78,** 91–121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Word segmentation in infants: How phonotactics and prosody combine. *Cognitive Psychology*, **38,** 465–494.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology,* **18,** 1–86.

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,** 144–178.

Morgan, J. L. (1994). Converging measures of speech segmentation in prelingual infants. *Infant Behavior and Development*, **17,** 387–400.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, **35,** 666–688.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, **66,** 911–936.

Nazzi, T., Bertoncini J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, **24,** 756–766.

Nazzi. T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, **1,** 123–147.

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

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research*, **36,** 959–972.

Norris, D. G. (1994). SHORTLIST: A connectionist model of continuous speech recognition. *Cognition*, **52,** 189–234.

Norris, D. G., McQueen, J., Cutler, A., & Butterfield, S. (1997). The possible word constraint in the segmentation of continuous speech. *Cognitive Psychology*, **34,** 191–243.

Peters, A. M. (1985). Language Segmentation: Operating principles for the perception and analysis of language. In Slobin, D. I. (Ed.), The crosslingvistic study of language acquisition (Vol, 2). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. J. Docherty, & D.R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, segment, prosody* (pp. 90–117), Cambridge, UK: Cambridge University Press.

Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. *Journal of the Acoustical Society of America*, **79,** 1616–1619.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274,** 1926–1928.

Saffran, J. R., Johnson, E. K, Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human adults and infants. *Cognition*, **70,** 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language,* **35,** 606–621.

Smith, B. L., & McLean-Muse, A. (1986). Articulatory movement characteristics of labial consonant productions by children and adults. *Journal of the Acoustical Society of America*, **8,** 1320–1328.

Soderstrom, M., Jusczyk, P. W., & Kemler Nelson, D. G. (2000). Evidence for the use of phrasal packaging by English-learning 9-month-olds. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 708–718), Somerville, MA: Cascadilla.

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

Turnbaugh, K. R., Hoffman, P. R., Danilog, R. G., & Absher, R. (1985). Stop-vowel coarticulation in 3-year-old, 5-year-old, and adult speakers. *Journal of the Acoustical Society of America*, **77,** 1256–1257.

Van de Weijer, J. (1998). *Language input for word discovery*. Unpublished Ph.D., University of Nijmegen, Nijmegen.

Valian, V., & Levitt, A. (1996). Prosody and adults' learning of syntactic structure. *Journal of Memory and Language*, **35,** 497–516.

Vitevitich, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science,* **9,** 325–329.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **21,** 98–108.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, **38,** 133–149.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, **24,** 744–755.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, **91,** 1707–1717.

Woodward, Z. N., & Aslin, R. N. (1990). *Segmentation cues in maternal speech to infants*. Poster presented at the International Conference on Infancy Studies, Montreal, Canada.

(Received October 22, 1999)
(Revision received August 1, 2000; published online March 15, 2001)