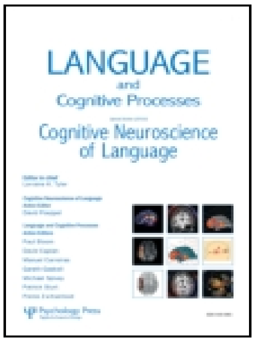

Language and Cognitive Processes

ISSN: 0169-0965 (Print) 1464-0732 (Online) Journal homepage: www.tandfonline.com/journals/plcp20

English-learning infants’ segmentation of trisyllabic words from fluent speech

Derek Houston, Lynn Santelmann & Peter Jusczyk

To cite this article: Derek Houston, Lynn Santelmann & Peter Jusczyk (2004) English-learning infants’ segmentation of trisyllabic words from fluent speech, Language and Cognitive Processes, 19:1, 97-136, DOI: 10.1080/01690960344000143

To link to this article: https://doi.org/10.1080/01690960344000143

LANGUAGE AND COGNITIVE PROCESSES, 2004, *19* (1), 97–136

English-learning infants' segmentation of trisyllabic words from fluent speech

Derek M. Houston
Department of Otolaryngology – Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, USA

Lynn M. Santelmann
Department of Applied Linguistics, Portland State University, Portland, OR, USA

Peter W. Jusczyk
Departments of Psychology and Cognitive Science, Johns Hopkins University, Baltimore, USA

Recent investigations with English-learning infants have shown that the rhythmic properties of English influence how infants segment fluent speech. In particular, 7.5-month-old infants have been shown to extract bisyllabic words that conform to the predominant strong/weak stress pattern of English but not weak/strong words. The present series of studies examined whether English-learning 7.5-month-olds' segmentation abilities are limited to extracting strong/weak bisyllables or whether they are able to segment longer strings, such as trisyllables (strong/weak/strong). The results indicated that infants can segment trisyllabic words from fluent speech but only when the first syllable receives primary stress (e.g., *cantaloupe*). When primary

---

Correspondence should be addressed to Derek M. Houston, Department of Otolaryngology – Head and Neck Surgery, 699 West Dr. / RR044, Indiana University School of Medicine, Indianapolis, IN 46202, USA. Email: dmhousto@indiana.edu

This work was supported by a Research Grant (HD-15795) from NICHD and a Research Scientist Award (MH-01490) from NIMH to PWJ. We thank Mara Goodman and Karla Jusczyk for assistance in recording the stimulus material and Ann Marie Jusczyk, Amy Gambon-Dennis and Eileen Crowley for their help in recruiting and testing infants. We are also grateful to Katie Nuckel for valuable help in acoustic analyses, and to Laura Dilley and Stefanie Shattuck-Hufnagel for their invaluable assistance with pitch accent and rhythmic analyses of the samples. Finally, Ann Marie Jusczyk and Joost van de Weijer made helpful comments on earlier versions of the present manuscript.

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960344000143

stress falls on the last syllable (e.g., *cavalier*), infants segment only the final stressed syllables. Overall, the findings suggest that 7.5-month-old English-learning infants are able to segment longer strings, and that they use stress as a major cue in segmentation. English-learners appear to equate strong syllables with word onsets only when the strong syllable receives primary stress.

A fundamental aspect of understanding language is recognising the words contained in utterances. Listeners must identify sequences of sounds in the input that correspond to representations of spoken words in the mental lexicon. However, speech segmentation is made more difficult by the fact that speech is continuous and often lacking in clear boundaries between words (Cole & Jakimik, 1978, 1980; Klatt, 1979, 1989). Thus, recognising words in the speech stream requires the ability to correctly determine word boundaries without the aid of reliable acoustic cues, such as pauses (Klatt, 1989).

Mature listeners can segment words from fluent speech with ease. Several types of linguistic knowledge contribute to adults' word segmentation abilities. In particular, listeners' familiarity with the typical properties of words is thought to aid the selection of potential word candidates (Cutler, 1994; Jusczyk, & Hohne, 1997; Norris, McQueen, Cutler, & Butterfield, 1997). Also, because the listener's goal is to identify sequences of sounds in fluent speech that match stored representations, the mental lexicon itself may play an important role in segmentation (Brent, 1999; Brent & Cartwright, 1996; Suomi, 1993). Indeed, the notion that one's lexical knowledge plays an important role in word segmentation has led to speculation that children may not segment words from fluent speech until they formed some word representations (Pinker, 1984; Suomi, 1993).

However, recent findings suggest that, between 6 and 7.5 months, infants develop the ability to segment some words from fluent speech. For example, Jusczyk and Aslin (1995) found that when they were familiarised with single syllable words presented in isolation, 7.5-month-olds, but not 6-month-olds, oriented significantly longer to passages containing repeated instances of these words than to comparable passages without these words. These findings suggest that by 7.5 months, infants can segment some words from fluent speech.[1]

---

[1] In another experiment, Jusczyk and Aslin (1995) presented infants with passages first and then tested them on recognition of the words presented in isolation. Infants again showed longer looking times to the familiar words than to the control. The results reported by Jusczyk and Aslin, as well as those by Jusczyk, Houston, and Newsome (1999b), consistently show the same patterns regardless of whether infants are familiarised with words in isolation and tested on passages or vice versa.

The fact that infants segment fluent speech before they have learned many words is not surprising, considering that infants must learn most words from the context of fluent speech. Words are not usually uttered in isolation, even to infants. Indeed, about 90–95% of utterances addressed to infants (excluding vocatives, fillers, and social expressions) are in the form of fluent speech (Morgan, 1996; van de Weijer, 1998). Very rarely do abstract nouns, verbs, or prepositions occur in isolation (van de Weijer, 1998). In addition, some function words, such as ''of'' virtually never do (van de Weijer, 1998). Even when caregivers are explicitly encouraged to teach their infants words, they only present the words in isolated contexts about 20% of the time (Woodward & Aslin, 1990). Thus, language learners must possess or develop the ability to segment words from fluent speech.

The ability to segment words from fluent speech depends on the listener's awareness of and facility with a number of different phonological properties of a language. In particular, sensitivities to three types of language-specific properties have been found to play roles in early speech segmentation abilities: the organisation of speech sounds, the prosodic properties of words and the distributional or statistical properties of sounds and words within the language. Infants are sensitive to many of the individual cues that may aid in segmentation (Friederici & Wessels, 1993; Hohne & Jusczyk, 1994; Jusczyk, Cutler, & Redanz, 1993a; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993b; Jusczyk, Luce, & Charles-Luce, 1994). However, only recently have investigators begun to explore whether infants actually use such information in word segmentation.

Infants need to acquire a number of different types of information in order to segment speech. Because the organisation of speech sounds into phonemes differs from language to language, infants need to acquire information about how the phonemes of their language are organised in order to segment words. Knowledge of the sound system can aid in segmentation. For example, Jusczyk, Hohne, and Bauman (1999a) found that 10.5-month-old, but not 9-month-old, English-learners are sensitive to allophonic cues to word boundaries. In particular, 10.5-month-olds were able to use allophonic information to distinguish between ''nitrates'' and ''night rates'' in fluent speech contexts. Similarly, Mattys, Jusczyk, Luce, and Morgan (1999) demonstrated that 9-month-olds are sensitive to how phonotactic patterns typically align with word boundaries. Moreover, Mattys and Jusczyk (2001) found that 9-month-olds were better able to segment words when good phonotactic cues to word boundaries were present than when such cues were absent.

Another potential source of information for segmenting words from fluent speech relates to the distributional (or statistical) properties of the input. Listeners may implicitly or explicitly notice regularities in the co-occurrence of some speech sound sequences to infer potential word boundaries. For example, if speech sounds X (e.g., /ba/) and Y (e.g., /tl/) occur much more often together than apart, listeners may infer that X + Y (e.g., ‘‘bottle’’) forms a cohesive unit. Likewise, X + Y may be perceived as more cohesive if it occurs across a variety of contexts (e.g., ‘‘big bottle,’’ ‘‘little bottle,’’ ‘‘fill the bottle with milk’’) than if the context is fixed. In fact, Goodsitt, Morgan, and Kuhl (1993) found that 7-month-olds were more likely to treat bisyllables as cohesive if they were previously presented in a variable context than if the context was fixed. More recently, Saffran, Aslin, and Newport (1996) showed that 8-month-olds could rely on such distributional regularities to segment words from fluent speech.

Prosody is another source of information that can be exploited for segmenting words from fluent speech. Infants are sensitive to prosodic patterns from a very early age. Neonates display sensitivity both to the prosodic patterns of their native language (Mehler et al., 1988), and to rhythmic differences between languages (Dehaene-Lambertz & Houston, 1998; Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Jusczyk, & Johnson, 2000). Early sensitivity to prosodic patterns makes prosody a strong candidate to play an important role in the earliest stages of segmentation. Infants begin to show sensitivity to native prosodic patterns of words between 6 and 9 months. For example, Jusczyk et al (1993b) found that 6-month-old, English-learning infants listen significantly longer to English than to Norwegian words, which differ markedly in their prosodic characteristics. Furthermore, evidence shows that English-learning infants become sensitive to some of the rhythmic properties of English words by 9 months. For example, one characteristic of English stress is that most content words in English conversational speech (about 90%) begin with a strong syllable, defined as a syllable with an unreduced vowel (Cutler & Carter, 1987). Jusczyk et al. (1993a) found that English-learning 9-month-olds, but not 6-month-olds, listen longer to lists of bisyllabic words with the predominant stress pattern of English, strong/weak (e.g., ‘‘pliant’’, ‘‘donor’’), than to weak/strong words (e.g., ‘‘abut’’, ‘‘condone’’). The pattern of findings indicates that, between 6 and 9 months, infants develop sensitivity to language-specific prosodic properties, useful in segmenting words.

Sensitivity to the predominant stress pattern of English words appears to influence infants’ segmentation of speech as well. For instance, Echols, Crowhurst, and Childers (1997) found that English-learning 9-month-olds were better able to recognise strong/weak than weak/strong bisyllables contained within longer sequences of strong/weak/strong syllables. Similarly, Morgan and Saffran (1995) found that 9-month-olds were more likely to treat strong/weak bisyllables as cohesive units than they were to treat weak/strong bisyllables. Finally, in a series of experiments using English words and passages, Jusczyk, Houston, and Newsome (1999b) demonstrated that 7.5-month-olds segment strong/weak but not weak/strong words from the context of fluent speech. Jusczyk et al. (1999b) also observed that when distributional context was held constant, 7.5-month-olds segmented strong/weak units from fluent speech even when these units crossed word boundaries. Thus, when infants were familiarised with passages containing weak/strong words always followed by the same function word (e.g., the target word *guitar* was always followed by *is*), infants demonstrated recognition of the pseudowords, such as '*taris*'.

These findings led Jusczyk et al. (1999b) to characterise English-learning infants as following a Metrical Segmentation Strategy (MSS). The MSS proposes that English speakers initially posit a word boundary at strong syllables. Cutler and colleagues had proposed that English-speaking adults rely on the MSS as a first pass strategy in segmenting words from fluent speech (Cutler, 1990; Cutler & Butterfield, 1992; Cutler, McQueen, Baayen, & Drexler, 1994; Cutler & Norris, 1988). The MSS derives from the discovery that the vast majority of words in English begin with a strong syllable (Cutler & Carter, 1987) and from evidence that adult English listeners are apt to treat strong syllables as word onsets (Cutler & Butterfield, 1992; Cutler & Norris, 1988; McQueen, Norris, & Cutler, 1994). This strategy allows for correct segmentation for the majority of English words, and is simple because for English it requires the listener to decide only if syllables contain a full or reduced vowel. The MSS may serve as a good initial strategy for speech segmentation for English-learning infants because they can rely on relatively few sources of information for segmentation: rhythmic and distributional information. The MSS would enable English-learning infants to segment most content words from the continuous speech in their environment. Also, by chunking fluent speech into smaller units, infants would position themselves to discover other sources of information useful for word segmentation, such as phonotactic and allophonic cues. For example, extracting a word like *taste* from fluent speech exposes two allophonic variants of /t/, the initial one aspirated, the second a non-aspirated stop. By analysing such chunks, English-learning infants may correctly infer that [t^h] usually marks a word onset. Similarly, attention to the fact that some types of phonotactic sequences are much more apt to appear within such chunks (e.g., [ft]) than others (e.g., [fh]) could also provide potential word segmentation cues.

While previous research has suggested that infants are able to segment single and bisyllabic words with initial strong syllables, and that they are sensitive to a variety of cues that might aid in segmentation, little is known about how the infants segment longer words. Segmentation of longer words is of particular interest because the Metrical Segmentation Strategy makes different predictions than strategies based on distributional and allophonic cues. The MSS suggests that an infant will posit a new word every time they encounter a strong syllable (an unreduced vowel). Thus, it predicts that if infants use the MSS only they will mis-segment long words such as *submarine*, into two different words, one for each strong syllable: *subma* [sʌbmə] and *rine* [ɹin]. On the other hand, strategies that include distributional and/or phonetic cues would predict that the infant may be able to segment these longer words as single units, if they have sufficient working memory. For example, infants may open a processing window at strong syllables and then rely on other information, such as distributional cues, to indicate where the word ends. An additional possibility is that infants are sensitive to the subtle cues that mark primary and secondary stress, and that they may be more likely to posit word onsets at strong syllables that carry the primary stress of the word than at strong syllables that carry secondary stress.

The current set of studies undertakes to begin to explore whether young, English-learning infants are able to segment longer words, and if so whether they use a strict MSS for longer words, or whether they appear to use additional cues in segmenting these longer strings. To understand better the developing word-segmentation abilities of English-learning infants, it is important to determine whether they posit word boundaries at the onset of each strong syllable in fluent speech. In particular, these studies explore whether infants follow a strict MSS, and thereby are likely to mis-segment long words, or whether they attend to a combination of cues, even at young ages, that allow them to reliably segment long words from fluent speech.

The present study explores these issues by testing English-learning 7.5-month-olds' ability to segment strong/weak/strong words from fluent speech. In the first set of experiments (Expts 1–3), infants are tested with words whose initial syllables are strong and carry the primary stress, and whose final syllables are also strong, but carry secondary stress (*ˈbaritone [ˈbeɹəˌton], ˈvestibule [ˈvɛstəˌbjul]ˈ, cantaloupe [ˈk^{h}æntəˌlop], ˈparachute [ˈp^{h}eɹəˌʃut]*). If English-learning infants always segment fluent speech at strong syllables, then they should segment the strong/weak/strong words into two separate units (e.g., *bari [beɹə]*, and *tone [ton]*) and not as cohesive, trisyllabic units. On the other hand, English-learning infants may be able to group longer sequences of syllables when they consistently follow a strong syllable, especially if that strong syllable also carries the primary stress. In this scenario, infants may segment a word like *baritone* as a whole word if it reoccurs in fluent speech. They may initially notice the first strong syllable because it is stressed and acoustically salient and then notice that the second two syllables consistently follow it, leading to a representation of the whole word.

In a second set of experiments (Expts. 4–6), we test English-learning 7.5-month-olds' ability to segment a different set of strong/weak/strong words from fluent speech. This second set of words differs from the first in that primary stress occurs on the final rather than the initial syllable, while the initial syllables carry secondary stress (*lemoˈnade [ˌlɛməˈned], magaˈzine [ˌmægəˈzin], cavaˈlier [ˌkʰævəˈlir], jamboˈree [ˌdʒæmbəˈɹi]*). Contrasting the segmentation of trisyllabic words with different lexical stress patterns allows us to evaluate whether the degree of lexical stress plays any role beyond that of strong syllables in English-learning infants' word segmentation strategies. For example, if infants are more likely to perceive word onsets at strong syllables that carry primary stress, then infants may extract only the final syllables from stress-final words in fluent speech.

EXPERIMENT 1

The first experiment tests English-learning 7.5-month-olds' ability to segment strong/weak/strong words with primary stress on the initial syllable. According to the MSS hypothesis, English listeners are apt to assume that all strong syllables are word onsets until some other information informs them otherwise. Jusczyk et al.'s (1999b) findings provide evidence that English-learning infants may initially rely heavily on the locus of strong syllables in segmenting words. They appeared to interpret strong syllables as only the onset of words in two syllable words, regardless of stress patterns. If this strategy carries over to trisyllabic words, then they may segment strong/weak/strong words into two units: a bisyllabic strong/weak word and a monosyllabic word with a strong syllable. However, infants at this age can also use distributional/statistical information to segment sequences of syllables from fluent speech (Jusczyk et al., 1999b; Saffran et al., 1996). If infants notice the co-occurrence of three-syllable patterns in fluent speech, then they may be able to segment a word like *baritone* from fluent speech as a cohesive unit rather than as *bari* + *tone.*

These studies investigate infants' segmentation of trisyllabic words from fluent speech with a modified version of the Headturn Preference Procedure (Kemler Nelson et al., 1995). Previous implementations of this procedure (Jusczyk & Aslin, 1995; Jusczyk et al., 1999b) have found longer looking times to passages containing familiarised words than passages containing other target words to be an indication that infants can segment and recognise the familiarised words in the context of fluent speech. Hence, we expect that if the infants can segment strong/weak/strong words from fluent speech, they will show longer looking times to passages containing the strong/weak/strong words they are familiarised with than to passages without these target words.

Method

Participants. Twenty-eight infants (11 female, 17 male) were recruited from monolingual, English-speaking homes in the Buffalo, NY area. The infants were approximately 7.5-months old, with a mean age of 33 weeks, 1 day (range: 31 weeks, 0 days to 35 weeks, 5 days). Another 21 infants were tested but excluded for the following reasons: crying (13), restlessness (4), parental interference (2), equipment failure (1), and experimenter error (1).

Stimuli. A female American-English talker from Western New York recorded the stimuli for the experiment. A 6-sentence passage was constructed for each of four target words (*baritone, cantaloupe, parachute, vestibule*). The target word occurred once in each sentence in variable sentence positions (see Table 1). The talker was encouraged to speak in a lively voice, as if addressing a small child. The recordings were made directly on the computer using a Shure microphone in a sound-attenuated room. The talker recorded the passages several times until they were acceptable. One important criterion in selecting the passages was that the target words be produced with a stressed initial syllable. The passages were digitised on a VAX Station Model 3176 at a sampling rate of 20 kHz via a 16-bit analogue-to-digital converter. The average duration of the passages was 18.94 s (ranging from 18.81 s for the *parachute* passage to 19.05 s for the *baritone* passage).

TABLE 1
Passages with stress-initial strong/weak/strong target words

Cantaloupe passage
The grocery store took away its old cantaloupe. Your cantaloupe is already at home. That red cantaloupe looks extremely strange. The dark cantaloupe tastes delicious. Give the children the plain cantaloupe. They think that her cantaloupe is edible.

Baritone passage
The dark baritone is very annoying. She plays that old baritone very skilfully. We need to get a plain baritone for the concert. The director really likes your baritone. Her baritone looks complicated. They composed a symphony for a red baritone.

Vestibule passage
Your vestibule has lots of decorations. Her vestibule can fit only one. They are renovating that old vestibule. The small, dark vestibule terrifies me. We are considering a plain vestibule. My new red vestibule is very comfortable.

Parachute passage
Our lieutenant has an old parachute. He gave the plain parachute to the beginner. Your parachute is colourful and new. The red parachute belongs to Timothy. The leader brought a dark parachute. Her parachute looks fantastic.

Acoustic analyses to determine the duration, pitch peak (F0), and vowel amplitude of each syllable were performed on the target words in the sentences. These values are shown in Table 2. ANOVAs were conducted to evaluate possible differences among the syllables in the target words. The mean duration of the second, weak syllables (103 ms) was significantly shorter than both the initial (239 ms), $F(1, 20) = 125.81$, $p < .001$, and the final (273 ms), $F(1, 20) = 195.36$, $p < .001$, strong syllables. The final strong syllables were significantly longer than the initial syllables, $F(1, 20) = 7.62$, $p < .001$. The pitch peak means were lower for the weak syllables (220 Hz) than the initial (243 Hz), $F(1, 20) = 9.04$, $p < .01$, and the final (253 Hz), $F(1, 20) = 18.39$, $p < .001$, strong syllables. However, the mean pitch peak difference between the initial and final strong syllables was not significant, $F(1, 20) = 1.64$, $p = .21$. The mean vowel amplitudes were significantly lower in the weak syllables (57 dB) than in the initial (62 dB), $F(1, 20) = 83.89$, $p < .001$, and final (58 dB), $F(1, 20) = 7.51$, $p < .01$, strong syllables. Furthermore, the vowel amplitudes were significantly greater in the initial than in the final strong syllables, $F(1, 20) = 7.62$, $p < .001$.

Taken together, the analyses of the target words revealed that the strong syllables were longer in duration, higher pitched, and louder than the weak syllables. These differences indicate that the strong syllables were clearly more stressed than the weak. However, the differences between the strong syllables were less consistent. The initial, primary-stressed syllables were shorter in duration, had about equal mean pitch peaks, and had vowels of higher amplitude than the secondary-stressed final syllables. Overall, the acoustic measurements do not clearly show differences in stress levels between the strong syllables. This finding can be attributed to the fact that

TABLE 2
Mean acoustic values for target words produced in passages (Experiments 1–3)

	Mean duration (ms)			*Mean pitch peak (Hz)*			*Mean vowel amplitude (dB)*		
	Syll 1	*Syll 2*	*Syll 3*	*Syll 1*	*Syll 2*	*Syll 3*	*Syll 1*	*Syll 2*	*Syll 3*
baritone [ˈbeɹəˌton]	207	99	310	232	214	280	64	57	59
cantaloupe [ˈk^{h}æntəˌlop]	256	107	200	259	234	244	60	57	59
parachute [ˈp^{h}eɹəˌʃut]	221	74	304	230	205	256	61	57	56
vestibule [ˈvɛstəˌbjul]	272	130	275	252	228	232	65	56	59

the acoustic markers of stress are influenced by several other factors besides stress. For example, syllable duration and pitch are influenced by word and utterance position among other things. It is well known that these factors render the assessment of syllable stress level, using acoustic measurements alone, problematic (Lehiste, 1970). In addition, differences in degree of stress, although perceptible, are generally quite subtle with respect to the main acoustic correlates of stress—pitch, duration, and loudness (Mattys, 2000). Thus, although the target words were produced with, and perceived to have, the intended stress-initial pattern, the acoustic analyses do not always reflect this.

The same female talker recorded the isolated versions of the target words. For each target word (*baritone, cantaloupe, parachute, vestibule*), the talker repeated the item with some variation 15 times in a row, in a lively voice, as if naming the object for an infant. The lists were then digitised on the computer in the same way as the passages. The average duration of the lists was 19.22 s (ranging from 19.16 s for the *cantaloupe* list to 19.26 s for the *baritone* list). The talker was instructed to stress the first syllable and reduce the second syllable, producing: (*ˈbaritone [ˈbeɹəˌton], ˈvestibule [ˈvɛstəˌbjul], ˈcantaloupe [ˈkʰæntəˌlop], ˈparachute [ˈpʰeɹəˌʃut]*). Two independent listeners agreed that these words followed this stress pattern.

Acoustic analyses, equivalent to those for the words in the passages, were conducted on the isolated words in the lists (see Table 3). Once again, the mean duration of the second, weak syllables (136 ms) were significantly shorter than both the initial (245 ms), $F(1, 56) = 589.48$, $p < .001$, and the final (383 ms), $F(1, 56) = 3015.28$, $p < .001$, strong syllables. The final strong syllables were significantly longer than the initial syllables, $F(1, 56) = 938.35$, $p < .001$. The pitch peak means of the initial strong syllables (428 Hz) were higher than the weak syllables (321 Hz), $F(1, 56) = 25.23$, $p < .001$, which were, in turn, significantly higher than the final strong syllables (273 Hz), $F(1, 56) = 4.94$, $p < .03$. Similarly, the mean vowel amplitudes were highest for the initial strong syllables (68 dB). They were significantly louder than the weak syllables (62 dB), $F(1, 56) = 99.88$, $p < .001$, which were significantly louder than the final strong syllables (60 dB), $F(1, 56) = 16.43$, $p < .001$. Thus, the stress of the first syllable is reflected in its relatively higher pitch and greater amplitude than the other two syllables. The secondary stress of the final strong syllable seems to have been carried in its much longer duration than the other two syllables.

Digitised versions of the passages and the lists were transferred to a Macintosh Centris 650 computer for playback during the experiment.

Apparatus. The experiment was conducted in a three-sided test booth constructed of pegboard, with 4 ft × 6 ft panels on three sides and open

TABLE 3
Mean acoustic values for target words produced in isolation (Experiments 1–3)

	Mean duration (ms)			*Mean pitch peak (Hz)*			*Mean vowel amplitude (dB)*		
	Syll 1	*Syll 2*	*Syll 3*	*Syll 1*	*Syll 2*	*Syll 3*	*Syll 1*	*Syll 2*	*Syll 3*
baritone [ˈbeɹəˌton]	182	146	416	414	370	272	68	64	59
cantaloupe [ˈkʰæntəˌlop]	276	123	274	454	293	260	67	65	62
parachute [ˈpʰeɹəˌʃut]	170	133	451	421	288	270	69	63	60
vestibule [ˈvɛstəˈbjul]	353	142	391	433	332	292	69	58	58
bari [ˈbeɹə]	238	280		444	415		65	65	
canta [ˈkʰæntə]	337	295		456	293		69	63	
para [ˈpʰeɹə]	251	302		446	426		67	65	
vesti [ˈvɛstə]	405	292		461	304		67	60	
tone [ton]			599			479			65
loupe [lop]			516			451			67
chute [ʃut]			590			455			69
bule [bjul]			508			455			65

at the back. The pegboard allows the experimenter to observe the infant through one of the holes. Except for a small section for viewing the infant, the remainder of the pegboard was backed with white cardboard to guard against the possibility that the infant might respond to movements behind the panel. 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 centre panel. A white curtain suspended around the top of the booth shielded the infants' view of the rest of the room. A Macintosh Centris 650 computer and response box were located behind the centre panel, out of view of the infants. 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 s. Information about the direction and duration of headturns and the total trial duration were stored in a data file on the computer.

Procedure. The Headturn Preference Procedure, as modified by Jusczyk and Aslin (1995), was used to test the infants. Each infant sat on the lap of a caregiver who was seated on a chair in the centre of the test booth. At the beginning of each trial, the centre light flashed until the infant oriented to the centre. Then, the centre light was turned off and one of the side lights began flashing. When the infant oriented at least 30° in the direction of the light, the speech stimulus was presented from a loudspeaker located behind the flashing light. The location of the loudspeaker from which the words were emitted was varied from trial to trial, with a different random order used for each infant. The stimulus continued until the infant looked away for 2 s or until the end of the trial. The amount of time the infant oriented to the stimulus side while the stimulus was playing was recorded for each trial.

The experiment had two phases. During the Familiarisation phase, each infant was presented with two words in citation form. One word was presented per trial in alternating order until at least 30 seconds of orientation time was accumulated for each word. Half of the subjects heard *baritone* and *cantaloupe,* and the other half heard *parachute* and *vestibule.* During the Test phase, all four passages were presented once in each of four blocks. The order of the passages within each block was randomised. For each subject, two of the passages contained the target words presented during familiarisation. An average orientation time difference between the passages containing the familiarised targets and the control passages is taken as an indication that the infants differentiated the two types of passages, presumably because they recognised the familiarised words in the passages (Kemler Nelson et al., 1995).

An observer hidden behind the centre panel looked through a peephole and recorded the direction and duration of the infants' headturns using a response box. The observer was not informed of which items served as familiarisation words for a given infant. The loudness level for the samples was set at 72 ± 2 dB (C) SPL, as referenced by a Quest (Model 215) sound level meter. During the experiment, both the observer and the caregiver wore foam earplugs and listened to masking music over tight-fitting closed headphones (SONY MDR-V600) so were unaware of which particular stimulus was presented at any given time. Caregivers and observers reported that with this masker they were unaware of either the location or the nature of the stimulus on the trial.

Reliability. The looking times of the infants were computed online by the observer. In order to assess the reliability of our online measures, a second person re-coded the looking responses of several of the infants. For each experiment, the videotape recordings of six participants were randomly selected for re-coding. Across the experiments, there was a very high correlation ($r = .992$) between the two codings. On 88% of the trials, the difference in coding was less than 0.50 s. The differences were greater than 1 s on only 4% of the trials. The average looking time was coded as 6.63 s by the original coder and 6.56 s by the second coder. This difference was not statically significant, $F(1, 70) = 3.53$, $p = .064$, and was very small (0.07 s). The difference between the coders' measurements was in the same direction for both types of passages. For the passages containing the familiarised words, the average looking time was coded as 7.08 s by the original coder and 7.00 s by the second coder. For the control passages, the average looking times were 6.17 s and 6.13 s by the original and second coders, respectively. There was no significant interaction between coding differences and (familiar vs. unfamiliar passages) condition, $F(1, 70) < 1$, suggesting that there was no live-observer bias to code the looking times as being longer for one passage condition or the other.

The original coder consistently measured slightly longer looking times than the second coder. There are several possible reasons for this small but consistent difference. One possible explanation is that the speed of the videotape play out may not have been exactly the same as the original live event. Another possible reason is that the eye movements may be more difficult to see on the videotape than live. As a result, the second coder may have been slightly slower to code the initial orientations to the lights. The results of the reliability measurements are given in the Results and Discussion section of each experiment.

Results and discussion

In all experiments reported here as well as other previous experiments using this procedure (Houston & Jusczyk, 2000; Jusczyk & Aslin, 1995; Jusczyk et al., 1999b), orientation times tend to decrease across blocks. However, no consistent interactions have been found between this block effect and differences in orientation time to the different passages. Hence, our analyses in all experiments focus on the mean listening times to the four different passages, calculated for each infant across the four blocks of trials. The average orientation times to the passages containing the familiarised words and to the control passages were computed for each infant. Nineteen of the 28 infants had longer average orientation times to the passages containing the familiarised words. The average orientation times were 7.65 s (SD = 2.53 s) for the passages containing the familiarised words and 6.77 s (SD = 1.56 s) for the control passages.[2] A paired *t* test indicated that this difference was significant, $t(27) = 2.25$, $p < .04$; 95% CI: $.08 < .88 < 1.68$. The results indicate that 7.5-month-olds were able to detect the strong/weak/strong words they were familiarised with in the context of fluent speech.

The present findings suggest that 7.5-month-old English-learning infants can segment words from fluent speech that are longer than a trochaic foot (i.e., strong/weak/strong words). In other words, infants were apparently able to extract repeating sound patterns in fluent speech and match them with what they stored in memory during the familiarisation phase.[3] However, these results do not preclude the possibility that infants were simply responding to the initial trochaic foot (e.g., *bari*) or the final syllable (e.g., *tone*) rather than the entire unit *(baritone)* in the passages. The MSS predicts that infants would segment strong syllables as the beginnings of a unit. If this is so, then, it is possible that 7.5-month-olds may have segmented only the initial trochaic feet or final strong syllables and then recognised the partial equivalence between these segments and the words they were familiarised with. On the other hand, it is possible that infants are attending not only to the strong syllables in the speech stream as the MSS predicts, but also to the distributional information contained in the speech stream. If infants are attending to both distribution and stress, then they ought to treat the whole SWS word as a unit, and not respond to just the initial segment.

The logic behind assessing whether infants segment whole words or parts of words from fluent speech stems from related findings by Jusczyk et al. (1999b). In experiments with bisyllables, Jusczyk et al. (1999b) argued that infants segmented strong/weak words from fluent speech as cohesive units rather than as strong + weak syllable sequences. They supported their claim by showing that infants oriented significantly longer to the passages containing strong/weak target items (e.g., *kingdom*) only when they were familiarised with the whole word and not when they were

---

[2] Videotapes of six participants were randomly selected and coded offline as a reliability check. The correlation between the offline and online codings was .994. On 90% of the trials, the difference in looking-time measurements was less than 0.50 s. The differences were greater than 1 s on 4% of the trials. The original coder's measurements were 7.39 s and 6.98 s for the familiar and unfamiliar passages, respectively. The second coders measurements were 7.28 s and 6.77 s. The original coder measured slightly longer looking times than the second coder, $F(1, 10) = 4.99$, $p < .05$, but there was no interaction with passage condition, $F(1, 10) < 1$.

[3] Note that during the familiarisation phase, words are presented in citation form with clear pauses between each repetition so that the beginning and ending of each word is clearly defined. Hence, we assume that infants encode the whole words as cohesive units during the familiarisation phase. What is in question is what they extract in the context of fluent speech.

familiarised with only the strong syllable (e.g., *king*). This suggests that the infants were treating the whole word *kingdom* as a cohesive unit, and were not simply segmenting out the initial strong syllable of this word. In other words, the infants were attending not only to the syllable that they are familiarised with, but also to the distribution of the familiarised element within the passages they hear. Although they are capable of detecting single syllable words from fluent speech (Jusczyk & Aslin, 1995), they will only segment out single syllables that appear in a variety of contexts. If the familiarised word occurred consistently followed by another syllable, then it appears that the infants did not treat the familiar initial syllable as an independent word, but as a new, unfamiliar word.

Further support for this hypothesis comes from Jusczyk et al. (1999b). This study also tested groups of infants with weak/strong words (*surprise*) by familiarising infants with the word (or part of the word) and then testing to see whether the infants listened longer to the passages that contained the familiar word. In contrast to the findings with strong/weak words, infants oriented longer to the passages containing the familiarised weak/strong words only when familiarised with the strong syllables (*prize*) and not when familiarised with the whole words (*surprise*). These results suggest that infants extracted only the final strong syllables from the weak/strong words in the passages. In this instance, the consistent syllable *preceding* the stressed syllable did not induce the infants to treat the WS group of syllables as an individual word.

However, these studies were conducted with two syllable words, which are common in English child-directed speech. How does distribution interact with familiarity in longer strings? We predict that if infants segment strong/weak/strong words (*baritone*) as two separate units (as the initial strong/weak trochaic foot (*bari*) separate from the final strong syllable (*tone*)), then they will respond to the familiar passages more strongly if familiarised with only the initial trochaic feet (*bari*). In addition, if they segment strong/weak/strong words as two separate units, then this predicts they will also respond if familiarised with only the final strong syllable (*tone*). The next two experiments investigated these two predictions. If this pattern is found, it would indicate that for strings longer than bisyllables, stress plays a larger role than distributional cues for segmentation. The next two experiments were designed to compare stress and distribution.

The first possibility tested is that infants attend only to the presence of strong syllables and segment the initial trochaic feet (*bari, canta*) of the strong/weak/strong (*baritone, cantaloupe*) as words in the passages. If infants attend only to the presence of strong syllables to mark new words, then this predicts that when they are presented with only the initial trochaic feet during familiarisation, they should also show a preference for the ‘‘familiar’’ passages.[4] However, if they are also attending to distributional and/or stress cues then they may fail to respond to the ‘‘familiar’’ passages with the full word because the initial trochees are consistently followed by a third strong syllable (*tone, loupe*), which could form a cohesive unit with the previous segments. This possibility is tested in Experiment 2. In Experiment 3, the possibility that infants are segmenting the final strong syllable as an independent unit is explored by testing infants on the passages after familiarising them on the final strong syllable.

EXPERIMENT 2

To determine if infants treat each strong syllable as a word onset and thus segment the initial strong/weak bisyllables from strong/weak/strong words from the passages as a separate unit, 7.5-month-olds were familiarised with the initial feet of the trisyllabic words and then presented with the passages with the whole trisyllabic words. If 7.5-month-olds segment the initial strong/weak (e.g., *bari ['beɹə]*) units as separate from the final strong syllables (e.g., *tone*) when they encounter strong/weak/strong words (e.g., *baritone*) we expect significantly longer orientation times to the ‘‘familiar’’ passages. In contrast, no differences in listening times would support the view that the infants are attending to more than the metrical stress patterns and treating the following strong syllable as a part of the preceding unit, yielding a strong/weak/strong trisyllablic unit. It is important to note that this experiment was not designed to test whether infants can segment a trisyllabic word (*baritone*) after being familiarised with an initial trochee (*bari*), rather it is intended to examine whether infants will segment the initial trochee as an independent unit despite the following strong syllable (*tone*) in the test passages. If this is the case, then it would indicate that strong syllables are the major cue that infants use for segmenting longer words in English.

Method

Participants. Twenty-eight infants (9 female, 19 male) were recruited from monolingual, English-speaking homes in the Buffalo, NY area. The infants were approximately 7.5-months old, with a mean age of 32 weeks, 3 days (range: 30 weeks, 1 day to 36 weeks, 2 days). Fourteen additional infants were tested but excluded for the following reasons: crying (9), restlessness (4), and experimenter error (1).

---

[4] ‘‘Familiar’’ passages are those that contain the words (or the whole-word versions of the part words) that infants heard during the familiarisation phase.

Stimuli. The same female talker recorded the new lists for the trochees *bari [ˈbeɹə], canta [ˈkʰæntə], para [ˈpʰeɹə],* and *vesti [ˈvɛstə].* The new lists were recorded in exactly the same manner and under the same recording conditions as for the original lists. The average duration of the lists was 19.12 s (ranging from 19.07 s for the *bari* and *para* lists to 19.16 s for the *canta* and *vesti* lists). Again, the speaker was instructed to produce the second syllable as reduced, so that the familiarisation tokens contained a schwa [ə] and not a full vowel such as [i] or [ɑ]. After the speaker created the lists, two independent listeners determined that the final vowels in all the tokens were reduced.

Acoustic analyses of the target items in the familiarisation lists revealed that the first syllables were significantly greater than the second syllables with respect to duration (308 vs. 292 ms), $F(1, 56) = 5.98$, $p < .02$, pitch peaks (452 vs. 360 Hz), $F(1, 56) = 11.51$, $p < .01$, and vowel amplitudes (67 vs. 63 dB), $F(1, 56) = 32.74$, $p < .001$. Measurements of the individual trochees are given in Table 3.

Apparatus. This was identical to Experiment 1.

Procedure. Half of the infants were familiarised with the *bari* and *canta* lists; the other half heard the *para* and *vesti* lists. Otherwise the procedure was identical to Experiment 1.

Results and discussion

The average orientation times to the passages containing the familiarised words and to the control passages were computed for each infant. Fifteen of the 28 infants had longer average orientation times to the passages containing the familiarised words. The average orientation times were 8.42 s (SD = 2.80 s) for the passages containing the familiarised words and 7.82 s (SD = 2.97 s) for the control passages.[5] Although the average orientation time to the passages containing the familiarised words was greater than the average orientation time to the control passages, the difference was not significant, $t(27) = 1.42$, $p = .17$; 95% CI: $-0.27 < 0.60 < 1.47$. The results suggest that familiarising the infants with only the initial trochees of

---

[5] Videotapes of six participants were randomly selected and coded offline as a reliability check. The correlation between the offline and online codings was .991. On 71% of the trials, the difference in looking-time measurements was less than 0.50 s. The differences were greater than 1 s on 11% of the trials. The original coder's measurements were 7.67 s and 6.67 s for the familiar and unfamiliar passages, respectively. The second coders measurements were 7.53 s and 6.46 s. The original coder measured slightly longer looking times than the second coder, $F(1, 10) = 9.51$, $p < .05$, but there was no interaction with passage condition, $F(1, 10) < 1$.

strong/weak/strong words does not induce them to orient significantly longer to the passages containing the whole trisyllabic words than to the control passages.

These findings are comparable with Jusczyk et al.'s (1999b) findings. When infants were familiarised with part of a word (e.g., *king*), they did not respond to the whole word (e.g., *kingdom*) in the passages. Presumably, this was because infants recognised the co-occurrence of *king* and *dom* in the passages, which led them to treat the two syllables as cohesive and not match them with the familiarisation word *king.* Both the present results and those by Jusczyk et al. (1996b) suggest that infants are sensitive to the distribution of syllables in fluent speech and not just stress patterns. This result also coheres with further experiments in Jusczyk et al. (1999b) which used weak/strong words and consistent distributional information in passages. The investigators found that infants segmented the strong final syllables (e.g., *tar*) of weak/strong words (e.g., *guitar*) in passages, but only when the target words in the passages were followed by varying words in the passage. In contrast, when the target word was always followed by the same word (e.g., *guitar is*) infants did not segment the final strong syllable. Instead they segmented the strong/weak units across word boundaries (e.g., *taris*). Taken together with the results from this experiment, this indicates that infants treat co-occurring stress-initial sequences as familiar cohesive units (e.g., *cantaloupe, tar_is*) only when familiarised with the whole sequence and not when familiarised with only part of the sequence.

Another possible explanation for the pattern of results in the present investigation is that the acoustic differences between the items presented during familiarisation and the target words in the passages were greater in Experiment 2 than in Experiment 1. Indeed, differences in duration and peak pitch between familiarisation items and the corresponding items in the passages were greater in Experiment 2 than in Experiment 1. The average duration of the initial strong syllables of words presented during familiarisation was only 6 ms longer than the initial strong syllables of the same items presented in the test passages for Experiment 1. By comparison, in Experiment 2 the mean duration of the initial strong syllables of the trochees presented during familiarisation was 69 ms longer than the initial strong syllables of the words presented in the test passages. Also, the familiarisation item/test item duration differences of the weak syllable were 33 ms for Experiment 1 and 189 ms for Experiment 2. Likewise, differences in average pitch peak for the first two syllables were 188 Hz and 101 Hz, respectively, for Experiment 1 and 209 Hz and 140 Hz for Experiment 2. It is possible that infants may not have been able to recognise the familiarised items in the passages in Experiment 2 because they were too acoustically different from what was produced in isolation.

However, in an experiment conducted by Jusczyk et al. (1999b), 7.5-month-olds recognised familiarised strong/weak words in passages, even though there were similarly large differences between the strong/weak words in the passages and the strong/weak words presented in isolation, with respect to average syllable durations (48 ms, 184 ms) and average pitch peaks (222 Hz, 91 Hz). Also, in the present set of experiments and the 15 experiments conducted by Jusczyk et al. (1999b) the stimuli were produced using infant-directed speech with substantial variation across tokens. Hence, infants in both studies would not likely be able to recognise words based on a strict acoustic pattern matching.

Thus, the findings do not support the notion that English-learning 7.5-month-olds posit word onsets at each occurrence of a strong syllable in fluent speech. If they had, then the occurrence of a second strong syllable in a word such as *canteloupe* should have led them to treat the initial strong/weak sequence, *cante*, as matching one of the items heard in familiarisation. Instead, the present findings lend support to the view that 7.5-month-olds segment strong/weak/strong words from fluent speech as cohesive units. However, another possible reason why infants may have failed to match the strong/weak sequences in the present experiment is that the phonotactics of these patterns are not typical for English words. Two-syllable patterns that end in a final schwa vowel are relatively uncommon in English. Perhaps, then, word segmentation processes were disrupted because infants perceived the familiarisation items to be non-English patterns. One way to avoid this problem was to familiarise the infants with only the final strong syllable of the target words that occurred in the passages. If the critical factor is whether familiarisation items do conform to English word structure, infants should find a match to these syllables in the test passages, because these syllables were all closed syllables. This possibility was tested in the next experiment.

EXPERIMENT 3

We tested whether infants might segment only the final syllables from trisyllabic words. Another group of 7.5-month-old English-learning infants was familiarised with the final strong syllables of the trisyllabic words (e.g., *tone*) and then tested on the original passages, containing the whole words (e.g., *baritone*). Again, it is important to note that this experiment was not designed to test whether infants can segment a trisyllabic word (*baritone*) after being familiarised with the final strong syllable (*tone*), but instead it examined whether infants will segment the final strong syllable as an independent unit. If this is the case, then it would indicate that strong syllables are the major cue infants are using for segmenting longer words in English.

Method

Participants. Twenty-eight infants (17 female, 11 male) were recruited from monolingual, English-speaking homes in the Baltimore, MD area. The infants were approximately 7.5-months old, with a mean age of 33 weeks, 1 day (range: 28 weeks, 6 days to 38 weeks, 6 days). An additional 11 infants were tested but excluded for the following reasons: failure to look at the flashing lights (2), restlessness (7), parental interference (1), and orientation times averaged less than 3 s (1).

Stimuli. The same female talker as in Experiments 1 and 2 recorded the new lists for the syllables *tone [ton], loupe [lop], chute [ʃut],* and *bule [bjul].* The new lists were recorded in exactly the same manner and under the same recording conditions as for the original lists. The average duration of the lists was 19.11 s (ranging from 19.07 s for the *loupe* list to 19.15 s for the *chute* list). The acoustic measurements of the syllables are given in Table 3.

Apparatus. This was identical to Experiment 1.

Procedure. Half of the infants were familiarised with the *tone* and *bule* lists; the other half heard the *chute* and *loupe* lists. Otherwise, the procedure was identical to Experiment 1.

Results and discussion

The average orientation times to the passages containing the familiarised words and to the control passages were computed for each infant. Eighteen of the 28 infants had longer average orientation times to the passages containing the familiarised words. The average orientation times were 7.26 s (SD = 2.06 s) for the passages containing the familiarised words and 6.56 s (SD = 2.69 s) for the control passages.[6] Similar to Experiment 2, the average orientation time for the passages containing the familiarised words was greater than the average orientation time for the control passages, but the difference was again not significant, $t(27) = 1.32$, $p = .20$; 95% CI: $-0.39 < 0.71 < 1.80$. The results suggest that familiarising 7.5-month-

---

[6] Videotapes of six participants were randomly selected and coded offline as a reliability check. The correlation between the offline and online codings was .965. On 87% of the trials, the difference in looking-time measurements was less than 0.50 s. The differences were greater than 1 s on 7% of the trials. The original coder's measurements were 7.10 s and 5.01 s for the familiar and unfamiliar passages, respectively. The second coder's measurements were 6.77 s and 5.17 s. The original coder measured slightly longer looking times than the second coder, but the difference did not approach statistical significance, $F(1, 10) < 1$. Also, the interaction between coder and passage conditions was not significant, $F(1, 10) = 2.92$, $p = .12$.

olds with only the final strong syllables of strong/weak/strong words does not induce them to orient significantly longer to the passages containing the whole trisyllabic words than the control passages. The pattern of findings here is the same as in the previous experiment. This suggests that the failure of infants in Experiment 2 to match the strong/weak familiarisation items to the appropriate patterns in the passages is not simply due to a problem with the phonotactics of the familiarisation items. In the present experiment, all of the familiarisation items were of the type commonly found in English, yet the infants did not generalise to the corresponding patterns in the passages.

This result also raises the possibility that the infants are sensitive to the differences between primary and secondary stress. These cues, while subtle (mainly amplitude differences for the test passages), may nonetheless be sufficient for the infant to use in their segmentation strategies. This issue will be considered in more detail below.

Taken together, the results of Experiments 1–3 suggest that 7.5-month-old English-learning infants can segment word patterns that are longer than a trochaic foot from fluent speech (Table 4), thereby extending the findings by Jusczyk et al. (1999b) for bisyllables. In fact, the pattern of results here very closely mirrors what Jusczyk et al. (1999b) found with strong/weak bisyllables. In both studies, infants oriented significantly longer to the passages with the familiarised whole words than to the control passages only when they were familiarised with the whole words. Moreover, in both cases, there was a nonsignificant trend toward listening longer to the passages that corresponded to the parts of the words played during familiarisation.

To further explore this apparent trend across the three experiments, an omnibus ANOVA was computed on the data from Experiments 1–3. There was a main effect of Word Familiarity, $F(1, 81) = 7.76, p < .01$, no significant main effect of Experiment, $F(2, 81) = 2.34, p > .1$, and no Experiment × Word Familiarity interaction ($F < 1$). Planned comparisons of the Word Familiarity effect in Experiment 1 (whole-word familiarisation) vs. Experiments 2–3 (part-word familiarisations) also

TABLE 4
Summary of results of Experiments 1–3

Experiment	*Familiarisation stimuli*	*Test stimuli*	*Evidence of segmentation?*
1	Isolated stress-initial S/W/S words	Passages with stress-initial S/W/S words	Yes
2	Isolated initial trochees	Same as above	No
3	Isolated final syllables	Same as above	No

revealed that the patterns of looking times were not significantly different in Experiment 1 compared with Experiments 2–3 ($F < 1$). Hence, although the Word Familiarity effect was significant in Experiment 1 but not significant in Experiments 2–3, there was a trend to attend longer to the ‘‘familiar’’ passages in all three experiments. Nevertheless, infants showed statistically significant longer orientation times to the ‘‘familiar’’ passages only when familiarised with the whole words. This pattern of results is certainly consistent with a segmentation of the entire target words from the passages rather than with a segmentation of only the strong syllables or initial trochees. However, there is another alternative to consider, namely, the possibility that the infants were more successful in Experiment 1 because there were more syllables that matched (i.e. three syllables per item) than in Experiment 2 (two syllables per item) or in Experiment 3 (one syllable per item). Although we cannot completely rule out this explanation on the basis of the three present experiments, it does run counter to what Jusczyk et al. (1999b) observed for 7.5-month-olds’ responses to weak/strong words. Specifically, Jusczyk et al. observed that their infants were actually more likely to match just the strong syllable rather than the whole weak/strong pattern to weak/strong words in passages. Because the findings from the next three experiments bear on this matter, we defer further consideration of this alternative until the General Discussion.

In any case, the present findings indicate that English-learning 7.5-month-olds are not limited to extracting only trochees from fluent speech. Instead, it appears that they open a processing window at a strong syllable and then pay attention to what follows that syllable. If more than one syllable, even another strong syllable, consistently follows then they treat the whole sequence as cohesive units. In other words, they are able to use statistical/distributional information, in addition to information about syllable type, to perceive two- and three-syllable sequences as cohesive. The present findings are also consistent with the idea that infants are more likely to posit word onsets at strong syllables when they carry primary stress than when they are a weaker stress.

Lexical stress may also play a role in infants’ segmentation of trisyllabic words. Consider that in Experiments 1–3, the target words all began with a strong syllable that also carried primary stress of the word, and that the final strong syllables of the trisyllabic words received only secondary stress. Might such stress differences affect the way that infants segment words from fluent speech? Mattys et al. (1999) have suggested that English-learning infants are sensitive to differences between primary and secondary stressed syllables. Suppose that infants use information about primary and secondary stress differences in segmenting words. For example, infants may open processing windows only at primary stressed syllables rather than at all strong syllables. In particular, only syllables with primary stress might be treated as word-initial syllables. If so, then one would expect infants to segment words with primary stress on the initial syllable (e.g., *baritone*) differently than they do words with primary stress on the final syllable (e.g., *magazine*). Alternatively, if infants are not sensitive to the differences in stress levels of strong syllables, one would expect them to segment both types of words in the same way. Experiments 4–6 explored these alternatives by examining English-learning 7.5-month-olds' segmentation of stress-final strong/weak/strong words, such as *magazine* and *lemonade*.

EXPERIMENT 4

To explore the role of lexical stress in infants' segmentation of fluent speech, we familiarised 7.5-month-olds with pairs of strong/weak/strong words with a different stress pattern than the words in Experiments 1–3. Rather than being stress initial, the words chosen for the next set of experiments had primary stress in final-syllable position. The words were *cavalier, jamboree, magazine,* and *lemonade.* As with Experiments 1–3, these words were intended to test whether infants would posit a new word at every strong syllable, or whether they were also attending to stress patterns.

The infants were tested with four 6-sentence passages—two with the familiarised target words and two control sentences with the other two target words. As with the earlier experiments, if infants detect familiarised targets of this type in fluent speech, they should listen significantly longer to the passages containing the familiarised words than to the control passages.

Method

Participants. Thirty-two infants (17 female, 15 male) were recruited from monolingual, English-speaking homes in the Baltimore, MD area. The infants were approximately 7.5-months old, with a mean age of 32 weeks, 1 day (range: 30 weeks, 2 days to 34 weeks, 3 days). An additional seven infants were tested but excluded for the following reasons: crying (1), restlessness (4), and equipment failure (2).

Stimuli. Because the talker who had recorded the passages used in the previous experiments was no longer available, a different native American–English female talker from Western New York recorded the stimuli for the experiment. A passage of six sentences was constructed for each of four target words (*cavalier, magazine, jamboree, lemonade*). The new passages were similar to the ones in Experiments 1–3 with respect to number of syllables and position of the target words (see Table 5). Once again, the talker was encouraged to read the passages in a lively voice, as if reading to a small child. The recordings were made in a sound attenuated room with a Shure microphone. The passages were digitised on a PC using CSL software at a sampling rate of 20 kHz. The average duration of the passages was 19.97 s (ranging from 19.41 s for the *jamboree* passage to 20.32 s for the *lemonade* passage).

Acoustic analyses of duration, pitch peak (F0), and vowel amplitude were performed on the target words in the sentences. The mean values for each syllable for each of these target words is given in Table 6. ANOVAs were conducted to evaluate possible differences among the syllables in the target words. The mean duration of the second, weak syllables (92 ms) was significantly shorter than that of both the initial (208 ms), $F(1, 20) = 124.62$, $p < .001$, and the final (250 ms), $F(1, 20) = 231.35$, $p < .001$, strong syllables. The final strong syllables were significantly longer than the initial syllables, $F(1, 20) = 16.38$, $p < .001$. The mean pitch peak was lower for the weak syllables (227 Hz) than for the initial (286 Hz), $F(1, 20) = 10.04$, $p < .003$, and the final (284 Hz), $F(1, 20) = 9.27$, $p < .005$, strong syllables. However, the mean pitch peak difference between the initial and final strong syllables was not significant, $F(1, 20) < 1.00$. The mean vowel amplitude was significantly lower in the weak syllables (54 dB) than in the

TABLE 5

Passages with stress-final strong/weak/strong target words

Jamboree passage

The old jamboree was still very exciting. Everybody left the funny jamboree in the morning. Tomorrow we'll buy tickets to the expensive jamboree. The baritone at the good jamboree was incredible. Their jamboree was truly spectacular. In the summertime he had a bright jamboree.

Magazine passage

On the sidewalk, we discovered an expensive magazine. The bright magazine was filled with colorful pictures. Unfortunately, my old magazine has deteriorated. A funny magazine was in the convenience store. You must ask politely to see their magazine. We recovered a good magazine from the parachute.

Lemonade passage

Their lemonade was advertised at the corner. Using strange ingredients, they mixed bright lemonade. We are replacing the old lemonade that is not refreshing. Sally knows a recipe for making good lemonade. At the supermarket, expensive lemonade is discounted. The funny lemonade tastes like cantaloupe.

Cavalier passage

In the front vestibule stands a good cavalier. Somebody took the old cavalier from the festivities. The expensive cavalier is rather ordinary. We prefer the funny cavalier in the entryway. The bright cavalier talks of extraordinary adventures. The citizens are going to cheer for their cavalier.

TABLE 6
Mean acoustic values for target words produced in passages (Experiments 4–6)

	Mean duration (ms)			*Mean pitch peak (Hz)*			*Mean vowel amplitude (dB)*		
	Syll 1	*Syll 2*	*Syll 3*	*Syll 1*	*Syll 2*	*Syll 3*	*Syll 1*	*Syll 2*	*Syll 3*
cavalier [ˌkʰævəˈliɹ]	176	127	264	286	244	299	64	55	58
jamboree [ˌdʒæmbəˈɹi]	279	57	198	242	199	270	56	51	55
lemonade [ˌlɛməˈned]	156	117	261	336	238	244	63	57	58
magazine [ˈmægəˈzin]	220	65	276	280	229	322	60	55	61

initial (61 dB), $F(1, 20) = 52.93$, $p < .001$, and final (58 dB), $F(1, 20) = 16.38$, $p < .001$, strong syllables. Furthermore, the vowel amplitudes were significantly greater in the initial than in the final strong syllables, $F(1, 20) = 10.42$, $p < .001$.

The acoustic analyses reveal a striking similarity between the stress-final and stress-initial strong/weak/strong words. In both cases the strong syllables tend to be greater than the weak syllables in measurements of duration, pitch, and loudness. Similarly, the initial strong syllables, on average, have greater amplitude, shorter duration, and about equal pitch peaks as the final strong syllables in both sets of words. Thus, the acoustic measurements do not clearly indicate differences between stress-initial and stress-final words in the context of fluent speech. However, this outcome is not surprising because, as noted earlier, finding clear acoustic correlates of stress levels is difficult across words in different sentential contexts and across words whose phonetic make up is quite different from each other (Lehiste, 1970).

The female talker also recorded isolated versions of the target words. She was asked to repeat each target word (*cavalier [ˌkʰævəˈliɹ], jamboree [ˌdʒæmbəˈɹi], lemonade [ˌlɛməˈned], magazine [ˌmægəˈzin]*) about 50–60 times with some variation, in a lively voice, and as if naming the objects for an infant. We selected 15 tokens of each target word. The selection was based on the quality of the pronunciation and assuring that the tokens used in the experiment all had word-final stress, and that the second vowel was reduced. The lists were then digitised on the computer in the same way as the passages. The average duration of the lists was 18.91 s (ranging from 18.38 s for the *cavalier* list to 19.36 s for the *magazine* list).

The same types of acoustic analyses were performed on the isolated target words. The mean duration of the second, weak syllables (74 ms) was significantly shorter than the mean duration of the initial (189 ms), $F(1, 56) = 91.73$, $p < .001$, and the final (467 ms), $F(1, 56) = 1069.85$, $p < .001$, strong syllables. The final strong syllables were significantly longer than the initial syllables, $F(1, 56) = 535.04$, $p < .001$. The mean pitch peak of the weak syllables (371 Hz) was significantly lower than the final (447 Hz), $F(1, 56) = 17.39$, $p < .003$, but not the initial (353 Hz), $F(1, 56) < 1.00$, strong syllables. The final syllables also had significantly higher pitch peaks than the initial syllables, $F(1, 56) = 26.54$, $p < .001$. The mean vowel amplitude of the final syllables (68 dB) was significantly higher than both the initial strong (65 dB), $F(1, 56) = 15.76$, $p < .001$, and the weak (65 dB), $F(1, 20) = 17.40$, $p < .001$, syllables. The vowel amplitudes of the initial strong and weak syllables did not differ significantly from each other, $F(1, 56) < 1.00$. Hence, the primary stressed (final strong) syllables differed from the secondary stressed (initial strong) syllables in pitch, duration, and amplitude. Acoustic measurements of the isolated versions of the target words are shown in Table 7.

Digitised versions of the passages and the lists were transferred to a Macintosh Centris 650 computer for playback during the experiment.

Apparatus. This was identical to the previous experiments.

Procedure. Half of the infants were familiarised with the *cavalier* and *magazine* lists while the other half heard the *jamboree* and *lemonade* lists. All other aspects of the procedure were identical to the previous experiments.

Results and discussion

The average orientation times to the passages containing the familiarised words and to the control passages were computed for each infant. Fifteen of the 32 infants had longer average orientation times to the passages containing the familiarised words. The average orientation times were 7.59 s (SD = 2.87 s) for the passages containing the familiarised words and 7.63 s (SD 2.88 s) for the control passages.[7] A *t*-test revealed that this

---

[7] Videotapes of six participants were randomly selected and coded offline as a reliability check. The correlation between the offline and online codings was .998. On 98% of the trials, the difference in looking-time measurements was less than 0.50 s. The differences were greater than 1 s on 1% of the trials. The original coder's measurements were 7.03 s and 6.21 s for the familiar and unfamiliar passages, respectively. The second coders measurements were 6.95 s and 6.19 s. The original coder measured slightly longer looking times than the second coder, but the difference was not statistically significant, $F(1, 10) = 2.42$, $p = .15$. Also, there was no interaction between coder and passage conditions, $F(1, 10) < 1$.

TABLE 7
Mean acoustic values for target words produced in isolation (Experiments 4–6)

	Mean Duration (ms)			*Mean Pitch Peak (Hz)*			*Mean Vowel Amplitude (dB)*		
	Syll 1	*Syll 2*	*Syll 3*	*Syll 1*	*Syll 2*	*Syll 3*	*Syll 1*	*Syll 2*	*Syll 3*
cavalier [ˌkʰævəˈliɹ]	181	106	412	373	361	377	69	64	68
jamboree [ˌdʒæmbəˈɹi]	272	23	438	336	380	419	63	63	66
lemonade [ˌlɛməˈned]	111	96	528	401	377	541	67	68	70
magazine [ˌmægəˈzin]	192	71	491	302	365	451	61	64	66
cava [kʰævə]	279	214		344	307		75	67	
jambo [dʒæmbə]	344	190		433	357		69	65	
lemo [lɛmə]	256	268		439	374		72	67	
maga [mægə]	358	214		383	391		72	66	
lier [liɹ]			525			468			69
ree [ɹi]			431			526			67
nade [ned]			577			461			70
zine [zin]			515			435			67

difference was not significant, $t(31) = -0.09$, $p = .93$; 95% CI: $-0.94 < -0.04 < 0.86$. The results suggest that infants did not respond to the target words in the passages that they were familiarised with.

In contrast to the findings in Experiment 1 with stress-initial trisyllabic words, 7.5-month-olds did not exhibit an ability to segment stress-final trisyllabic words from fluent speech. If the infants segmented anything of the target words from the passages it was not enough to be recognised with what they heard during familiarisation. There are a number of possibilities that could account for these results. First, 7.5-month-olds may extract nothing from stress-final strong/weak/strong words from fluent speech. Second, they may segment these words as two separate units—one corresponding to the initial trochee, the other to the final stressed syllable. Third, infants might segment only the primary stressed syllable. Any of these possibilities could involve a mismatch between the patterns they perceived during familiarisation to the isolated words and how they perceived the stress-final trisyllabic words in fluent speech.

To explore these various possibilities, we conducted two more experiments. In Experiment 5, we familiarised another group of infants with only the final, lexically stressed syllables and then tested them with passages containing the whole words. In Experiment 6, we tested the status of the initial trochee in strong/weak/strong trisyllables with final stress. Once again, the aim of these experiments was to examine whether the infants would extract part of the trisyllable as a word after having been familiarised with that part, not to examine whether they could extract the entire word after having been familiarised with part of the word.

EXPERIMENT 5

This experiment investigates the possibility that infants may identify word onsets with primary stressed syllables. English-learning 7.5-month-olds were familiarised with only the final syllables of stress-final strong/weak/strong and then tested with passages containing the whole words. Evidence of significantly longer listening times to the passages with the whole words that included the syllables from familiarisation would suggest that, in fluent speech contexts, infants tend to segment stress-final strong/weak/strong words at the onset of the final strong syllable.

Method

Participants. Thirty-two infants (18 female, 14 male) were recruited from monolingual, English-speaking homes in the Baltimore, MD area. The infants were approximately 7.5-months old, with a mean age of 32 weeks, 2 days (range: 28 weeks, 3 days to 34 weeks, 4 days). An additional 9 infants were tested but excluded for the following reasons: crying (3), restlessness (1), orientation times averaged less than 3 seconds (2), parental interference (1), failure to look at the side lights (1), and equipment failure (1).

Stimuli. The female talker from Experiment 4 recorded the new lists for the stressed syllables *lier [liɹ], ree [ɹi], nade [ned],* and *zine [zin].* The new lists were recorded in exactly the same manner under the same recording conditions as for the original lists. The average duration of the lists was 18.61 s (ranging from 18.29 s for the *ree* list to 18.87 s for the *nade* list). Average values from acoustic analyses of the individual syllables are given in Table 7.

Apparatus. This was identical to the previous experiments.

Procedure. Half of the infants were familiarised with the *ree* and *nade* lists; the other half heard the *lier* and *zine* lists. Otherwise the procedure was identical to the previous experiments.

Results and discussion

The average orientation times to the passages containing the familiarised words and to the control passages were computed for each infant. Twenty-two of the thirty-two infants had longer average orientation times to the passages containing the familiarised words. The average orientation times were 8.44 s (SD = 3.04 s) for the passages containing the familiarised syllables and 7.46 s (SD = 2.46 s) for the control passages.[8] A paired *t*-test indicated that this difference was significant, $t(31) = 2.06$, $p < .05$; 95% CI: $0.01 < 0.98 < 1.95$. Thus, 7.5-month-olds did detect the final stressed syllables of the strong/weak/strong words in the passages.

The pattern of results in Experiments 4 and 5 with stress-final strong/weak/strong words is similar to the findings by Jusczyk et al. (1999b) with stress-final weak/strong words. In both cases, when infants were familiarised with the final stressed syllables, they listened significantly longer to the passages containing the whole words than the control passages. Moreover, in both cases, when infants were familiarised with the whole words, they exhibited no signs of differentiating the passages with the familiarised words from the control passages. One conclusion that can be drawn from these patterns of findings is that 7.5-month-olds extract word-final stressed syllables from fluent speech as isolated one-syllable units. Moreover, the results of the current experiments suggest that the positioning of stress among strong syllables plays a role in infants' speech segmentation.

But what happens with the initial trochees of stress-final strong/weak/strong words? One possible scenario is that infants perceive a word onset at the initial strong syllable but then perceive a new word onset after the initial trochee because the final syllable is more stressed than the first. Hence, infants may extract both the initial trochees and the final syllables

---

[8] Videotapes of six participants were randomly selected and coded offline as a reliability check. The correlation between the offline and online codings was .994. On 81% of the trials, the difference in looking-time measurements was less than 0.50 s. The differences were greater than 1 s on 2% of the trials. The original coder's measurements were 6.16 s and 5.62 s for the familiar and unfamiliar passages, respectively. The second coder's measurements were 6.34 s and 5.67 s. The original coder measured slightly longer looking times than the second coder, $F(1, 10) = 5.66$, $p < .05$, but there was no interaction with passage condition, $F(1, 10) = 1.76$, $p = .21$.

of these words as two separate units. Another possibility is that infants do not perceive the initial strong syllable as a word onset because it does not carry primary stress and is, perhaps, less salient perceptually. The next experiment investigated these possibilities.

EXPERIMENT 6

If 7.5-month-old English-learning infants segment stress-final strong/weak/strong as two separate (strong/weak + strong) units, then they should exhibit recognition of the strong/weak units just as they demonstrated recognition of the final strong syllables. However, if infants only segment words or syllables that carry primary stress from fluent speech, then they should not respond to the initial trochees. To explore these possibilities, we familiarised 7.5-month-olds with the initial trochees of the stress-final strong/weak/strong words and tested them on passages containing the whole words.

Method

Participants. Thirty-two infants (16 female, 16 male) were recruited from monolingual, English-speaking homes in the Baltimore, MD area. The infants were approximately 7.5 months old, with a mean age of 32 weeks, 1 day (range: 30 weeks, 0 days to 35 weeks, 5 days). Fifteen additional infants were tested but excluded for the following reasons: crying (10), restlessness (4), and orientation times averaged less than 3 seconds (1).

Stimuli. The female talker from Experiments 4 and 5 recorded the new lists for the trochees *(cava [kʰævə], jamba [dʒæmbə], lemo [lɛmə], and maga [mægə]).* The new lists were recorded in exactly the same manner under the same recording conditions as for the original lists. The average duration of the lists was 17.91 s (ranging from 17.48 s for the cava *list to 18.29 s for the* maga *list). Once again, the speaker was instructed to reduce the second vowel; an independent listener determined that the second vowel was heard as a schwa [ə], rather than a full vowel.*

Acoustic analyses of the target items revealed that the first syllables were significantly greater than the second syllables with respect to duration (309 vs. 221 ms), $F(1, 56) = 369.43$, $p < .001$, pitch peaks (400 vs. 357 Hz), $F(1, 56) = 4.30$, $p < .05$, and vowel amplitudes (72 vs. 66 dB), $F(1, 56) = 131.82$, $p < .001$. The acoustic values of the individual trochees are shown in Table 7.

Apparatus. This was identical to the previous experiments.

Procedure. Half of the infants were familiarised with the *cava* and *maga* lists while the other half heard the *jamba* and *lemo* lists. All other aspects of the procedure were identical to the previous experiments.

Results and discussion

The average orientation times to the passages that contained the familiarised trochees and to the control passages were computed for each infant. Seventeen of the thirty-two infants had longer average orientation times to the passages containing the familiarised words. The average orientation times were 7.49 s (SD = 2.67 s) for the passages containing the familiarised words and 7.51 s (SD = 2.78 s) for the control passages.[9] A paired t test indicated that this difference was not significant, $t(31) = -0.02$, $p = .97$; 95% CI: $-0.97 < -0.02 < 0.93$. The results indicate that 7.5-month-olds did not detect the match between the familiarised strong/weak trochees and the strong/weak/strong words in the passages. Thus, the present findings contrast with those of the previous experiment in which the infants did detect a match between the final syllables of the strong/weak/strong words and the whole words in the passages. Apparently, whether a strong syllable carries primary stress or not does make a difference in how infants segment the speech stream. However, this conclusion must be tempered in light of the results of an omnibus ANOVA conducted with the data from Experiments 4–6. As expected, neither the main effect of Experiment ($F < 1$) nor main effect of Word Familiarity, $F(1, 93) = 1.33$, $p > .2$, was significant. However, the Experiment × Word Familiarity interaction was also not significant, $F(1, 93) = 1.60$, $p > .2$, despite the fact that only infants in Experiment 5 showed significant preferences for the passages containing the familiarised items (Table 8).

GENERAL DISCUSSION

The findings reported here suggest that 7.5-month-old English-learning infants have some ability to segment trisyllabic words from fluent speech and that lexical stress (i.e., which syllable carries the primary stress of the word) plays a role in their ability to do so. Experiments 1–3 tested infants' segmentation of stress-initial strong/weak/strong words in fluent speech. Infants oriented significantly longer to the passages containing the target

---

[9] Videotapes of six participants were randomly selected and coded offline as a reliability check. The correlation between the offline and online codings was 1.0. On all of the trials, the difference in looking-time measurements was less than 0.50 s. The original coder's measurements were 7.13 s and 6.54 s for the familiar and unfamiliar passages, respectively. The second coders measurements were 7.12 s and 6.53 s. The original coder measured slightly longer looking times than the second coder, $F(1, 10) = 4.81$, $p = .053$, but there was no interaction with passage condition, $F(1, 10) = 1.5$, $p = .25$.

TABLE 8
Summary of results of Experiments 4–6

Experiment	*Familiarisation stimuli*	*Test stimuli*	*Evidence of segmentation?*
4	Isolated stress-final S/W/S words	Passages with stress-final S/W/S words	No
5	Isolated final syllables	Same as above	Yes
6	Isolated initial trochees	Same as above	No

items when familiarised with the whole word and not when familiarised with only the initial trochees or the final syllables. In contrast, Experiments 4–6 showed that infants do not segment stress-final strong/weak/strong words from fluent speech. Instead, they only showed evidence of extracting the final, primary-stressed syllables. The latter experiments demonstrate that infants' performance in matching the familiarisation items to comparable strings in the passages is determined by more than the number of syllables that match. Whereas infants in Experiments 1–3 only seem to detect the patterns in the passages when all three syllables matched, infants in Experiments 4–6 were only successful in matching the patterns in the passages when they were familiarised with the primary stressed syllable. The latter finding seems to rule out a potential account that infants were successful in segmenting trisyllabic words in Experiment 1, and not in Experiments 2 and 3, because Experiment 1 had a greater number of syllables in common between the familiarisation and test sequences. If that were the case, then one would have expected infants in Experiment 4 (where they were familiarised with three syllables) to outperform those in Experiment 5 (where they were familiarised with a single, final stressed syllable). Instead, the reverse was true.

These results cohere with previous work on infants' segmentation of rhythmic units. Previous research has focused on strong/weak and weak/strong units, finding that the former is preferred over the latter and that the former is located in fluent speech while the latter type is missed (Echols et al., 1997; Jusczyk et al., 1999b). But 7.5-month-olds' processing capacity is apparently not limited to strong/weak units. Instead, our results suggest that English-learning infants can segment stress-initial strong/weak/strong words from fluent speech as cohesive units. In other words, infants do not always shut their processing window at the end of strong/weak units, nor do they posit that all strong syllables indicate new words. Instead, they can leave their processing window open to include at least one strong syllable in addition to a weak syllable. Similarly, Saffran et al. (1996) found that English-learning 8-month-olds treat strings of three CV syllables (all strong syllables) as cohesive units when they occur repeatedly in the same order.

Our results support those of Saffran et al., using more complex syllables and English passages.

However, infants do not always segment re-occurring trisyllabic sequences from fluent speech as cohesive units. In her thesis, Johnson (2003) has recently found that infants' segmentation of trisyllabic sequences was influenced by whether they were produced as cohesive units or if there was subphonemic information (i.e., allophonic and co-articulatory information) that indicated word boundaries. For example, Johnson (2003) found that 8-month-olds segmented trisyllabic words (e.g., *catalogue*) from fluent speech as whole words when they were produced as words but not when they were produced as 3-word phrases (e.g., *cat a log*). But when familiarised with passages containing 3-word idioms that are produced as fixed phrases (e.g., *piece of cake*), infants did segment these from fluent speech as whole words. Parallel to Johnson's findings, infants in the present study did segment trisyllabic segments (but not the parts) when primary stress fell on the first syllable, but they did not segment trisyllabic sequences from fluent speech as whole words when primary stress fell on the final syllable. Taken together, these results suggest that infants segment trisyllabic constituents from fluent speech when subphonemic, rhythmic, and distributional information converge to support segmentation of sequences as cohesive and not as separate units.

In addition to support for the fact that infants use multiple cues to segment words, one further implication of our findings is that subtle differences between primary and secondary stressed syllables may affect 7.5-month-olds' segmentation of trisyllabic words from fluent speech. Our findings that stress plays a role in the segmentation of strong/weak/strong trisyllables coheres with results from Mattys et al. (1999) with 9-month-olds. Mattys et al. presented infants with lists of bisyllabic words in which both syllables were strong (i.e., contained an unreduced vowel). For each word list, either the first or second syllables were stressed. The infants treated stress-initial words as cohesive and stress-final words as two separate units. Taken together, the findings from Mattys et al. and the present study are consistent with the notion that infants are sensitive to levels of stress and that stress level plays a role in segmentation.

Stress appears to play a role in infants' segmentation ability even though stress itself is somewhat difficult to measure acoustically. Stress level sometimes manifests itself in simple acoustic measurements when the stimuli are controlled for other phonetic properties (Mattys, 2000). For example, analysing stress difference between which had identical initial segments, such as *controversy* and *controversial,* Mattys found small but consistent differences in duration, pitch, and loudness. The acoustic analyses of the isolated target words in our study also revealed the predicted differences in duration, amplitude, and pitch between primary stressed and secondary stressed syllables in both types of words, despite using words with very different segments. However, measurements of these same properties failed to provide any systematic differences between primary and second stressed syllables in the fluent speech versions of the targets. As noted earlier, this is not entirely unexpected given the impact of sentential context on stress (Lehiste, 1970).

While acoustic measures did not provide reliable indicators of stress for the words in the passages, other factors such as rhythmic patterns may have provided additional cues for stress. Further analyses of the stimuli revealed a rhythmic difference that seems likely to have played a significant role. In particular, 21 of the 24 occurrences of the initially stressed strong/weak/strong words had a rhythmic beat on the initial syllable, whereas none of the stress final words had a rhythmic beat on the first syllable. The second syllables of both types of words never received a rhythmic beat, whereas the third syllables almost always did (23/24 and 24/24 for the stress-initial and stress-final words, respectively). Assuming that attention is likely to be drawn to rhythmically prominent locations in utterances (Jones, 1976), the fact that none of the words with final stress ever had rhythmic beats on their initial syllable certainly may have worked against infants' segmentation of the whole words from the passages. Thus, the stress-final and stress-initial strong/weak/strong words in the fluent speech contexts did embody rhythmic differences that seem to have affected how infants segmented the different types of strong/weak/strong patterns.

Another implication of our results is that they suggest that English-learning infants may be using a segmentation strategy that differs somewhat from the Metrical Segmentation Strategy (MSS) posited for adult English-listeners (Cutler, 1990; Cutler & Butterfield, 1992; Cutler et al., 1994; Cutler & Norris, 1988). In the MSS, the critical distinction is between strong syllables (ones with full vowels) and weak syllables (ones with reduced vowels) (Fear, Cutler, & Butterfield, 1995).[10] No distinction

---

[10] The role of stress in segmentation is currently a topic of debate in the literature on adult segmentation. The MSS postulates that English listeners treat stress in a binary fashion—either strong or weak (Fear, Cutler, & Butterfield, 1995). On the other hand, Mattys (1999) argues that levels of stress between strong syllables could be an important factor in segmentation. The findings here with 7.5-month-olds are that the degree of stress in syllables does indeed play a role. It is interesting to note that, at an age when infants have been shown to not be able to use such cues as phonotactic and allophonic cues, level of stress plays such a strong role in segmentation. These findings support the possibility that 7.5-month-olds posit onsets at the most salient syllables and that the likelihood that a syllable is considered an onset is related to its stress level in a continuous manner. This conclusion does not bear heavily on the debate of the role of stress in adult listeners because the behaviour observed is consistent with both end states. After 7.5-months, the amount of stress between strong syllables could continue to play a role in segmentation, as suggested by Mattys (1999), or listeners may, according to Cutler and colleagues, attend to vowel quality (i.e., weak or strong) as the primary rhythmic cue for segmentation.

is made with respect to the level of stress that strong syllables bear. However, the present findings suggest that infants are more likely to treat strong syllables with primary stress as word onsets than strong syllables with secondary stress. Why might English-learning 7.5-month-olds process strong/weak/strong words with initial stress differently than strong/weak/strong words with final stress? There are several possible explanations. One possibility is the one that we have been considering throughout the paper: English-learning 7.5-month-olds may use the locus of the primary stress as an indicator of word onsets in fluent speech. However, there are other factors that might also lead to greater success in segmenting strong/weak/strong words with initial stress. Strong/weak/strong words with initial stress are far more common in English than are strong/weak/strong words with final stress. In fact, the latter type are rare enough so as to be treated as exceptions to accounts of how English stress placement is determined (Burzio, 1994). Hence, the infrequent occurrence of such patterns in the linguistic input may bias infants against segmenting items as strong/weak/strong patterns with final stress. Another possible factor that may have influenced infants' segmentation of the different types of strong/weak/strong words has to do with pitch accenting. Pitch accents in utterances are thought to have the effect of drawing listeners' attention to an important syllable (Bolinger, 1958). Analyses of our passages indicated that there was a slight difference in pitch accenting for the stress-initial and stress-final words in the passages. None of the 24 stress-initial targets (4 passages with 6 occurrences of each target) was pitch accented, whereas 4 of the 24 stress-final targets were pitch accented. Although it is hard to see how such a small difference may have affected the infants' performance, together with differences in rhythmic properties described earlier, there may have been sufficient cues to signal primary and secondary stress for the infants.

While a difference in stress placement between words is one possible explanation for our findings, there is a potential alternative explanation. The average duration of the target words in Experiment 1–3 passages was 615 ms while the average duration of the target words in the Experiment 4–6 passages was 549 ms—a small but reliable, $F(1, 46) = 7.00$, $p < .05$, difference. This difference in duration suggests that there was a difference in speaking rate between the two types of passages, which may have played a role in differences in performance between the two types of passages. Indeed, infants were unable to segment the whole target words from the passages with the faster speaking rate (Experiments 4–6) but were able to segment the whole target words from the passages with the slower speaking rate (Experiments 1–3).

However, the pattern of performance found across all of the experiments cannot be fully explained by differences in speaking rate. For example, in both Experiments 3 and 5 infants were familiarised with only the final syllables. Infants attended longer to the familiar passages only in Experiment 5, in which the final syllable contained primary stress. If speaking rate alone, and not differences in stress placement, determined the performance of the infants, then we would expect equal or superior performance in the slower speaking rate condition than in the faster speaking rate condition. In contrast to that prediction, infants demonstrated segmentation of the final syllable in Experiment 5 but not in Experiment 3. Because there are no investigations on the role of speaking rate on infants' segmentation of words from fluent speech, we cannot rule out the possibility that it played a role in the findings of this investigation. However, the pattern of results is more consistent with an explanation based on differences in rhythmic and distributional properties of the target words.

While results of these experiments support a rhythmic-based segmentation strategy that gradually becomes supplemented with other cues for word segmentation, the experiments presented here leave a number of questions unanswered. Given that primary and secondary stress is relative, why do infants treat syllables that carry the primary stress as word onsets but not strong syllables that carry secondary stress? How could infants know that an initial strong syllable carries secondary rather than primary stress until they hear the rest of the word? One possibility is that infants' ability to perceive a re-occurring strong syllable is probabilistic rather than absolute, and the acoustic factors that play a role in the perception of primary stress (such as rhythmic beat placement, discussed above) also increase the chances that infants will perceive these syllables in the context of continuous speech. Further investigations into factors that influence infants' perception of syllables in fluent speech would be helpful for understanding the probabilistic nature of infants' speech segmentation. In addition, as noted above, little is known about the interaction between speaking rate and word segmentation, and further studies investigating these areas would be helpful for a thorough understanding of infants' abilities to segment speech.

How a rhythmic and distributionally based segmentation strategy may evolve can be seen through a consideration of what English-learning 7.5-month-olds must still achieve to develop adult-level word segmentation skills. Because a large majority of words in spoken English begin with a strong syllable (Cutler & Carter, 1987) identifying word onsets with the occurrence of these syllables will lead to a correct segmentation in the majority of cases. Indeed, at 7.5 months, English learners appear to follow such a strategy, even though it leads them to mis-segment words with weak/strong stress patterns (Jusczyk et al., 1999b). However, because English does have content words that begin with weak syllables, as well as grammatical words (function words) that are weak syllables, English-learning infants eventually have to be able to segment these types of words. A strategy of identifying word onsets with strong syllables that carry primary stress will not be sufficient for segmenting words beginning with weak syllables. However, breaking the speech stream up into smaller units at each strong syllable that carries primary stress may allow infants to learn to use other potential word segmentation cues, such as allophonic and phonotactic cues. Access to multiple sources of information about potential word boundaries has been shown to improve the accuracy of computer simulation models of word segmentation (e.g., Brent & Cartwright, 1996; Christiansen, Allen, & Seidenberg, 1998). In fact, there is evidence that between 7.5 and 10.5 months, English learners do develop the ability to use allophonic (Jusczyk et al., 1999a) and phonotactic (Mattys et al., 1999; Mattys & Jusczyk, 2001) cues to word boundaries in fluent speech. Hence, it is not surprising that, by 10.5-months, English learners show some ability to segment weak/strong words from fluent speech (Jusczyk et al., 1999b).

As infants' abilities continue to develop, other relevant sources of information become available for use in word segmentation. For example, infants' lexicons will expand and grow rapidly during the course of their second year. Many investigators studying word recognition abilities have pointed to the critical role that lexical competition plays in adult listeners' segmentation and recognition of words in fluent speech (e.g., Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994; Norris, McQueen, & Cutler, 1995; Vitevitch & Luce, 1999). One might well expect that, as new information comes on-line, infants would adjust the relative weighting of the various word boundary cues. Thus, some cues that play leading roles in word segmentation at the youngest ages (e.g., strong syllables carrying primary stress) may well turn out to play lesser roles in word segmentation by fluent speaker/listeners.

Finally, these studies, and others that involve recognising familiarised sound patterns (e.g., Houston & Jusczyk, 2000; Jusczyk & Aslin, 1995; Jusczyk et al., 1999; Saffran et al., 1996) show that infants have some ability to store word representations in memory. An important step to forming a lexicon is storing the sound patterns of words and matching them to the appropriate meanings. While it is very unlikely that the 7.5-month-olds in these experiments are close to learning the meanings of words like *vestibule*, they may well retain information about such sound patterns. Indeed, 9-month-olds have been shown to retain information about the sound patterns of frequently occurring words for as long as 2 weeks (Jusczyk & Hohne, 1997). The ability to extract and store such words from fluent speech may mark an important step towards learning the meaning of words and their grammatical functions. As infants are able to integrate more cues for word segmentation, they should improve their recognition of different words in a variety of contexts. In turn, these improved word recognition skills should put infants in a better position to pick up new information about the words' meanings and grammatical functions.

In conclusion, the present studies indicate that English-learning 7.5-month-olds have some ability to segment stress-initial trisyllabic (strong/weak/strong) words from fluent speech. However, for words with final stress, infants at this age appear to segment only the final syllables from trisyllabic words. Our findings suggest that for English learners, primary stressed syllables serve as better indicators of word onsets in fluent speech than secondary stressed syllables. In time, as infants learn other segmentation cues and integrate these with the rhythmic cues, they are able to segment words that do not begin with primary stress.

Manuscript received July 2000
Revised manuscript received June 2003

REFERENCES

Bolinger, D.L. (1958). A theory of pitch accent in English. *Word*, *14*, 109–149.

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.

Burzio, L. (1994). *Principles of English stress*. Cambridge: Cambridge University Press.

Christiansen, M.H., Allen, J., & Seidenberg, M.S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*, 221–268.

Cole, R.A., & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing*. New York: Academic Press.

Cole, R.A., & Jakimik, J. (1980). A model of speech perception. In R.A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In T.M.A. Gerry (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives. ACL-MIT Press series in natural language processing.* (pp. 105–121) Cambridge, MA: MIT Press.

Cutler, A. (1994). Segmentation problems, rhythmic 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., Baayen, H., & Drexler, H. (1994). *Words within words in a real-speech corpus.* Paper presented at the Fifth Australian International Conference on Speech Science and Technology, Perth.

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

Dehaene-Lambertz, G., & Houston, D. (1998). Faster orientation latencies toward native language in two-month-old infants. *Language and Speech*, *41*, 21–43.

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

Fear, B.D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, *97*(3), 1893–1904.

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

Goodsitt, J.V., Morgan, J.L., & Kuhl, P.K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, *20*, 229–252.

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

Houston, D.M., & Jusczyk, P.W. (2000). The role talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1570–1582.

Johnson, E.K. (2003). Word segmentation during infancy: The role of subphonemic cues to word boundaries. Unpublished doctoral dissertation. The Johns Hopkins University. Baltimore, MD.

Jones, M.R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*, 323–355.

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

Jusczyk, P.W., Cutler, A., & Redanz, N.J. (1993a). Infants' 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. (1993b). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*, 402–420.

Jusczyk, P.W., & Hohne, E.A. (1997). Infants' memory for spoken words. *Science*, *277*, 1984–1986.

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

Jusczyk, P.W., Houston, D.M., & Newsome, M. (1999b). 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.

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 and Development*, *18*, 111–116.

Klatt, D.H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*, 279–312.

Klatt, D.H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

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

Mattys, S. (2000). The perception of primary and secondary stress in English. *Perception and Psychophysics*, *62*, 253–265.

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). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465–494.

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

McQueen, J.M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 621–638.

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

Morgan, J.L. (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: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 756–766.

Nazzi, T., Jusczyk, P.W., & Johnson, E.K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*, 1–19.

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

Norris, D., McQueen, J., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1209–1228.

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.

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

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

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

van de Weijer, J. (1998). *Language input for word discovery*. Unpublished doctoral dissertation, University of Nijmegen, The Netherlands.

Vitevitch, M. & Luce, P.A. (1999). Probablistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374–408.

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