Spectral tilt as a cue to word segmentation in infancy and adulthood

ERIK D. THIESSEN and JENNY R. SAFFRAN
University of Wisconsin, Madison, Wisconsin

Across a variety of tasks, adults respond differently to syllables with multiple stress cues than to syllables with only one cue to stress. This series of experiments was designed to explore how infants and adults use partial stress as a cue to word boundaries. In the first experiment, 9-month-old infants treated syllables with only one cue to stress (spectral tilt) as a strong cue to word boundaries. The second experiment shows that whereas adults treat syllables with multiple cues to stress as word onsets, they do not consider syllables marked only by spectral tilt to be strong indicators of word boundaries. The third experiment shows that 1-year-old infants are more adultlike than 9-month-olds in their use of stress cues. Taken together, these results suggest a rapid development of stress cue knowledge in infancy, perhaps due to infants' experience with word segmentation.

Even before they speak their first word, infants learn to use the acoustic regularities of speech to identify word boundaries and other structural aspects of language. Like all languages, English is characterized by numerous correlations between sounds and word boundaries that proficient language users employ to identify word onsets and offsets. However, since these correlations differ across languages, they must be learned by novice language users. Infants are sensitive to a number of these correlations by the end of their 1st year of life, suggesting that the learning mechanism that underlies infants' ability to discover them is both flexible and powerful (e.g., Jusczyk, 1999; Saffran & Thiessen, 2003). By 9 months, infants are sensitive to phonotactic cues; they prefer to listen to syllables spoken with wordlike rhythmic structure when those syllables have consonant clusters that are likely to occur within words, as opposed to consonant clusters (such as /vt/) that are unlikely to occur within a word in English (Mattys, Jusczyk, Luce, & Morgan, 1999). Infants also appear to be sensitive to allophonic cues to word boundaries by 11 months. For example, they can use the differences in articulation between the /tr/ in "night rates" and "nitrates" as a cue to whether the /tr/ is word internal or crosses word boundaries (Jusczyk, Hohne, & Bauman, 1999).

Although there are many auditory cues to word boundaries, the one to which infants respond at the earliest age appears to be lexical stress (Jusczyk, Houston, & Newsome, 1999). Lexical stress may also be the cue to word boundaries weighted most heavily by infants (Johnson & Jusczyk, 2001), at least in English, where most words follow a very regular pattern of lexical stress. In a corpus study, Cutler and Carter (1987) found that over 90% of English content words are trochaic—that is, are stressed on their first syllable (as opposed to iambs, which are stressed on the second syllable). Depending on the task, infants will use rhythm as a cue to word boundaries by as early as 7.5 months of age (Jusczyk, Houston, & Newsome, 1999). By 9 months, infants treat stress syllables as word onsets even if there is other information—such as transitional probabilities between syllables (Saffran, Aslin, & Newport, 1996)—that indicates different word boundaries (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003).

Stress in English is realized through the alteration of three parameters: pitch, duration, and amplitude (Lieberman, 1960). Of these three, pitch is believed to be the most critical, with duration close behind, whereas amplitude is considered less important (e.g., Fry, 1955). Increasing a syllable's pitch or duration alone is often sufficient to lead adult listeners to treat a syllable as stressed, whereas increasing amplitude alone is a less effective stress cue (Fry, 1955; Streeter, 1978; although see Sluijter & van Heuven, 1996). Not surprisingly, across a number of paradigms, syllables more closely approximating natural stress are more effective at influencing adults' responses. That is, when there are more cues to stress present in a syllable, adults are more likely to treat that syllable as stressed and to consider stress to be a strong cue to structure (Streeter, 1978).

Adults' extensive experience with English has resulted in a flexible and adaptive response to prosodic information (e.g., Cutler & Darwin, 1981). However, it is unclear whether infants will respond to manipulation of the pa-

rameters of stress in the same way that adults do. One way to frame this question is to ask how broad is the group of syllables that infants will respond to as though they are stressed. By analogy with results from the development of categorization, there seem to be three likely possibilities. Infants might respond to the manipulation of stress cues in much the same way that adults do, suggesting either that they begin language learning with a complete knowledge of how lexical stress works or that they have had sufficient experience to have already acquired this information. Another possibility is that infants' relative inexperience with stress will lead them to be conservative in their categorization of syllables as stressed. That is, whereas adults will treat a syllable as *moderately stressed* when it has most, but not all, of the attributes of a stressed syllable, infants might not yet be capable of a graded response and will treat a syllable as stressed only if it is a near-perfect exemplar of stress. Finally, infants might be liberal in their identification of stressed syllables. Unlike adults, whose response to a stressed syllable depends on how strongly that syllable is stressed, infants might be willing to respond to a syllable with one cue to stress as being just as good a stressed syllable as a syllable with all cues to stress present. The categorization literature provides numerous examples of infants' categorizing both quite broadly and too narrowly (Eimas & Quinn, 1994; Eimas, Quinn, & Cowan, 1994). The breadth of categorization is highly dependent on the distribution of exemplars to which the infant is exposed during familiarization and on the requirements of the task (Mareschal & Quinn, 2001). Infants' response to stress may also be dependent on the distribution of stress cues to which they have been exposed. Unfortunately, it is not clear what the distribution of fully stressed versus partially stressed syllables is in the input that children receive, so it is difficult to know what specificity of categorization this distribution affords.

In order to better understand how infants learn about the relevance of auditory cues in speech and how that learning affects infants' later language acquisition, it is important to understand what sort of learning infants are doing. For example, the learning mechanisms involved may be very different if infants are narrowing their representation of stress, as opposed to broadening it. To gain an understanding of infants' response to stress cues and how it changes, we need to observe whether infants respond differentially to syllables with one and multiple cues to stress, in a task in which infants are known to be attentive to stress. Word segmentation provides such a task. Normally, 9-month-old infants respond to syllables with multiple cues to stress as word onsets even if sequential statistical information in the speech stream indicates different word boundaries (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). That is, even if two syllables usually or always occur together, which is a good signal that the two syllables are part of the same word, infants will not group the syllables together as a word if the second syllable is stressed. Instead, they treat the first (unstressed) syllable

as the end of a previous word, and the second (stressed) syllable as the onset of a new word.

Previous experiments have established that infants will missegment fluent iambic speech if the iambic stress is realized through the presence of multiple cues to stress (Thiessen & Saffran, 2003). In that series of experiments, stressed syllables were louder, longer, and higher pitched than unstressed syllables. In the present experiment, we asked whether syllables containing only one cue to stress are sufficient to cause infants to ignore statistical cues to word boundaries and missegment fluent speech. Importantly, the language we used in this experiment was identical to the one used in Thiessen and Saffran's (2003) experiments, except in the nature of the stress cues it contained. Whereas those languages contained multiple cues to stress (pitch, duration, and amplitude), the present experiment exposed infants to fluent speech containing only one stress cue, spectral tilt (ST, a cue that was not used in Thiessen and Saffran's, 2003, experiments). If infants missegment stimuli that include only one cue to stress in the same way that they missegmented a language that contained multiple cues to stress, the evidence would be consistent with the idea that infants are willing to use a very broad range of acoustic events as cues to word boundaries. However, if infants segment in a manner consistent with statistical information after hearing a language that is iambic only by virtue of a single cue to stress, this would suggest that infants, like adults, consider syllables with only one aspect of stress to be weaker cues to linguistic structure than are syllables marked by more aspects of stress.

EXPERIMENT 1

Because 9-month-old infants who are exposed to English expect words to be stressed on their first syllable, they will missegment iambic words from fluent speech (Jusczyk, Houston, & Newsome, 1999). For example, when they hear the word "guitar," 9-month-old infants incorrectly treat "tar" as the start of a word, not the end, because that syllable is stressed. Rhythm is apparently a very powerful cue to word boundaries at this age; infants will missegment in this manner even when there are statistical cues in the speech stream that point out word boundaries (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). In this experiment, we asked whether infants continue to missegment when only a single parameter of stress, ST, renders the fluent speech "iambic": the flattened distribution of energy across frequency (a particular manifestation of the alteration of amplitude in stressed syllables). All syllables have ST, a term that refers to the decrease in energy across the spectrum. Voiced speech has more energy at low frequencies than at high frequencies (e.g., Klatt & Klatt, 1990). The average value of this decrease is 12 dB per octave (Fant, 1960). However, in stressed syllables, this decrease is less steep, leading to a more equal distribution of energy across the frequency spectrum. Although we could have altered any single parameter of stress to op-

erationalize partial stress, we chose to use ST. We did so because our subjective impression, and that of several adult pilot subjects, was that syllables with altered ST were not particularly reminiscent of naturally stressed syllables. In contrast, syllables with altered pitch or duration seemed more similar to natural stress.

We exposed infants to fluent speech in which the second syllable of words had a flattened ST, and asked whether this single cue to stress caused them to missegment the speech. This experiment pitted two segmentation cues (statistical information and ST) against each other. Infants heard a stream of speech such as "diTI#buGO# daPU# doBI" (capitalized syllables signify syllables marked with flattened ST). If infants attend to statistical cues, they should segment words, such as "diTI" and "buGo" from the speech. If, however, they treat syllables marked with ST as word onsets, they should segment part-words (syllable sequences that cross word boundaries) such as "TIbu" or "GOda" from the speech.

In a previous series of experiments (Thiessen & Saffran, 2003), we found that 9-month-old infants, when presented with iambic speech containing multiple cues to stress (amplitude, duration, pitch) segmented the part-word "TIbu" from fluent iambic speech. The present experiment employed a language with the identical word order but containing only ST as a cue to stress, rather than multiple cues. In this experiment, we used the same method to ask whether infants respond to ST alone as though it is a stronger cue to word boundaries than is statistical information.

Infants might segment either words or part-words from fluent iambic speech, depending on the cue to which they choose to attend. These results will thus be uninterpretable unless infants' baseline direction of preference after learning is known. For example, imagine that the infants, after being familiarized with an iambic language, show a preference for part-words. Does this suggest that infants have *incorrectly* segmented the language and are showing a familiarity preference for the items that they have segmented? Or does this suggest that they have segmented the language according to statistical cues to word boundaries and are showing a novelty preference for the items that they (correctly) did not segment from fluent speech? Either interpretation is plausible.

In previous experiments using the head turn preference procedure, it has often been difficult to predict infants' direction of preference a priori (e.g., Aslin, 2000). In some experiments, infants act as though they are bored with the words that they have segmented from speech and listen longer to novel words during test trials (e.g., Echols, Crowhurst, & Childers, 1997; Saffran et al., 1996). In other experiments, infants prefer to listen longer to the familiar stimuli (the words they segmented from speech) during test trials (e.g., Jusczyk & Aslin, 1995; Jusczyk, Houston, & Newsome, 1999). Without knowing whether infants are expected to show a familiarity or a novelty preference given these materials, it will be impossible to interpret a preference for words or part-words.

In order to resolve this difficulty, we included a separate group of infants exposed to the same words, spoken with the first syllable receiving the cue to stress; these infants heard a stream of speech, such as "DIti#BUgo#DApu# DObi." In a language with trochaic words, both the stress cue and the transitional probabilities point to the *same* word boundaries. Therefore, we can be certain that infants will successfully segment words from this speech stream. If infants listen longer to part-words after hearing the trochaic language, we can say that infants at this age will exhibit a novelty preference in this kind of task. If instead they listen longer to the words, we can say that infants at this age will exhibit a familiarity preference in this task. With this baseline information available for comparison, it becomes possible to know which items infants segmented from the speech stream after listening to the iambic language, in which the stress cue is pitted against statistical information.

Method

Participants. The participants were 34 infants between the ages of 8.6 and 9.4 months. The average age of the participants was 9.13 months. To obtain the 34 infants for this experiment, it was necessary to test 54. The other 20 were excluded for the following reasons: crying (6), failure to complete at least two of one or more trial types (5), refusal to look at side lights (3), looking times of less than 3 sec, on average, to either side light (3), parent-reported ear infections on the day of the experiment (2), and experimental error (1). Of the 34 participants included in the data analysis, 17 were randomly assigned to the ST trochaic condition (average age: 9.05 months), and 17 were randomly assigned to the ST iambic condition (average age: 9.2 months).

Stimuli: Acoustics. All the stimuli were generated by the MacInTalk speech synthesizer running on a PowerMac 5300. Two artificial languages were synthesized for use during familiarization, one iambic and one trochaic, each consisting of the same four disyllabic nonsense words: "dapu," "dobi," "bugo," and "diti." Each language consisted of the same words spoken in the same randomized order. There were no pauses between words, and all the syllables were fully coarticulated. The synthesizer produced syllables with a monotonic $F0$ of 200 Hz, sampled at 16000 Hz and 16 bits. We used synthesized speech because it allowed better control over the acoustics of our languages. A human speaker would have found it quite difficult to alter ST—and only ST—in a manner consistent with stressing every other syllable while maintaining fluent speech.

Both familiarization languages[1] were 1 min 20 sec in length. In the trochaic language, the first syllable of every word was filtered to flatten its ST. In the iambic language, the second syllable was filtered. Thus, the infant heard a stream of speech like one of the following:

Iambic: "daPU#buGO#diTI#doBI . . ."

Trochaic: "DApu#BUgo#DIti#DObi . . ."

Past results indicate that amplitude alone is an ineffective cue to stress. However, this has been found in experiments in which amplitude was altered simply by superimposing the duration-adjusted amplitude envelope from a stressed syllable onto an unstressed syllable (Streeter, 1978). More sophisticated analyses of speech suggests that amplitude is not altered across the spectrum in a stressed syllable. Instead, stress is likely to be realized by an increase in amplitude (of between 5 and 10 dB) in regions of the spectrum above 0.5 kHz. Intensity below 0.5 kHz may be unaffected by the presence or absence of stress (Sluijter & van Heuven, 1996). Therefore, although there are amplitude changes in a stressed syllable, the changes

have the effect of altering the distribution of energy across frequency. Amplitude normally decreases by approximately 12 dB per octave in speech (Fant, 1960); this decrease is much less steep in stressed syllables.

In order to roughly simulate the decreased ST of a naturally stressed syllable, we used Matlab to apply a filter to the vocalic portions—as identified by steady-state formants—of the syllables in our fluent speech that would normally receive stress (in the trochaic language, the first syllables of words; in the iambic language, the second syllables). This filter was designed to accentuate high frequencies and affected frequencies between 300 and 3000 Hz, approximately the range in which the effects of stress on ST have been explored (Sluijter & van Heuven, 1996). Energy at 300 Hz and below was unaffected (amplified by 0 dB). The filter increased the amplification of the frequencies linearly between 300 Hz (0 db) and 3 kHz (20 dB). This resulted in a significant increase in the overall amplitude of the filtered syllables. At the final stage in the process, the filtered syllables were decreased in amplitude—as measured by average root-mean square (RMS) amplitude—to their original level, using CoolEdit; we then used an amplitude meter to verify that the two languages were of equal amplitude, approximately 60 dB SPL at the infant's head. Thus, the altered syllables that the infants finally heard were at the same level of overall amplitude as they had been before filtering but had a significantly flatter ST than they had before the filtering process. However, no effort was made to match these utterances in a precise fashion to the ST of actual utterances. These syllables resembled naturally stressed syllables only in that their ST was flatter than usual; the resultant ST was not necessarily a good match for the ST that a human speaker would have imparted to that syllable.

The infants were tested on their ability to distinguish words from part-words (disyllabic sequences that crossed word boundaries). Two words from the languages and two part-words served as the test items. Since English-hearing infants prefer to listen longer to trochees than to iambs (Jusczyk, Cutler, & Redanz, 1993), all the items were presented with neutral stress during test trials. This mismatch between the articulations of the test items from the familiarization session to the test trials likely made the task more difficult for the infants (Hunter & Ames, 1988). However, this design feature was necessary because, if the infants had heard the test items articulated in the same way during both the test trials and the familiarization period (i.e., with flattened ST), they might have preferred to listen to the trochees, whether or not they had segmented these items from speech, making interpretation difficult.

Stimuli: Statistical structure. Both the trochaic and the iambic languages contained identical statistical cues to word boundaries, because they had the same word order. The transitional probability between syllables within a word was 1.0, whereas the transitional probabilities across word boundaries ranged between .2 and .4. However, the relationship between flattened ST and statistical cues was critically different in the two languages. In the trochaic language, both ST (to the extent that infants believe ST is a cue to word onsets) and transition probabilities between syllables indicated the same word boundaries. But in the iambic language, ST and statistical cues conflicted. If the infants attended primarily to statistical cues, they would segment the same items from the iambic speech as those from the trochaic speech (since statistical cues are identical in both languages). However, if the infants identified syllables with flattened ST as word initial, they should missegment the words in this language. Since these words contained a cue to stress on their second syllable, the infants using a metrical segmentation strategy would treat the second syllable of a word as the word onset. Therefore, they would assume that a part-word such as "PUbu" (actually the second syllable of "daPU" and the first syllable of "buGO") was a word, as opposed to a syllable sequence that crossed word boundaries.

An interpretational difficulty that sometimes arises when testing infants' discrimination between part-words and whole words is that

words occur more frequently than part-words. This is because, after a word such as "dapu," any of the other words in the language (here, "bugo," "diti," or "dobi") can occur. Each of them will occur, on average, one third of the time after "dapu." Therefore, the part-word formed across the boundaries of words occurs only one third as often as whole words.

Differences in test item frequency present a problem for an experiment such as the present one, since they add another factor that could influence infant preference. In order to remove this potential confound, the two familiarization languages (iambic and trochaic) were constructed so that two of the words ("dapu" and "dobi") occurred twice as often (90 times) as the other two ("diti" and "bugo"), which occurred only 45 times each. Therefore, the part-words formed across the boundaries of the two frequent words ("dapu" and "dobi") occurred 45 times each, just as often as the infrequent whole words (see Aslin, Saffran, & Newport, 1998, for further discussion of the statistical features of this type of language). The two infrequent words ("diti" and "bugo") and the two part-words formed across the boundaries of the frequent words ("pudo" and "bida") served as the test items in these experiments.

Procedure. The infants were tested individually in a double-walled sound-attenuated room while seated on a parent's lap. An experimenter outside the booth observed the infants' looking behavior on a video monitor connected to an infrared camera inside the room and coded the direction of the infants' gaze on line. The parent inside the room listened to masking music to eliminate bias, and the observer was similarly unable to hear the stimulus being played to the infant.

At the beginning of the familiarization phase, a light in the center of the wall facing the infant began to flash, directing the infant's gaze forward. Simultaneously, one of the two languages (either iambic or trochaic; each infant heard only one language) began to play from the speakers beneath the two side lights—one light and speaker on each side wall—in the room. The familiarization phase lasted 2 min 20 sec.

Immediately after familiarization, 12 test trials were presented. All the infants heard the same test trials, regardless of familiarization condition. Six of these trials were part-word trials, and 6 were word trials. Each test item occurred on 3 trials during the testing session. A test trial began with the blinking light at the center of the wall facing the infant drawing the infant's gaze forward. When the observer signaled the computer that the infant had fixated on the center light, one of the side lights began to flash, and the center light simultaneously stopped. As soon as the infant made a head turn of at least 30° in the direction of the flashing side light, the experimenter signaled the computer, and one of the test items was presented from the speaker beneath the flashing light. Test items were presented in random order, with 6 trials (3 words and 3 part-words) presented from each side speaker. The test item continued to play for as long as the infant continued to look at the flashing side light. When the infant looked away for more than 2 sec, the test item stopped playing, and the center light began to blink again. This procedure was repeated until the infant had completed all 12 test trials.

Results and Discussion

We first compared listening times to words and part-words for infants exposed to the ST trochaic language. As is shown in Figure 1, the infants listened to words for 6.9 sec ($SE = 0.6$) during the test trials and to part-words for 6.1 sec ($SE = 0.5$). Twelve of the 17 infants listened longer to words than to part-words during the test trials after exposure to the trochaic language. A paired t test (all t tests reported are two-tailed) indicated that the difference in looking times between words and part-words was significant [$t(16) = 2.53, p < .05$].
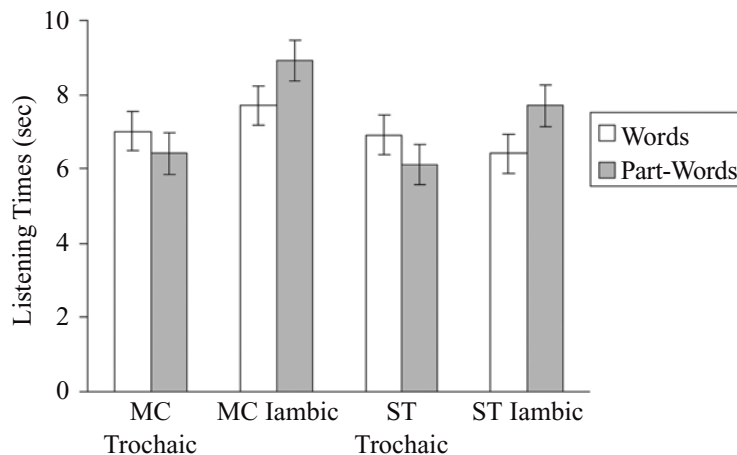
Figure 1. Nine-month-old infants' looking times to words and part-words in the multiple cues to stress (MC) trochaic, MC iambic, spectral tilt (ST) trochaic, and ST iambic conditions (data for the MC conditions are from Thiessen & Saffran, 2003, Experiment 1).

We then compared listening times to words and part-words for the infants exposed to the ST iambic language. As is shown in Figure 1, the infants listened to words for 6.4 sec ($SE = 0.5$) and to part-words for 7.6 sec ($SE = 0.4$). Thirteen of the 17 infants listened longer to part-words than to words during the test trials after exposure to the iambic language. A paired t test indicated that the difference in looking times between words and part-words was significant [$t(16) = 2.77, p < .05$].

After hearing the trochaic language, the infants listened longer to words than to part-words. In contrast, the infants listened longer to part-words than to words after listening to the iambic language. To assess whether this difference in direction of preference was significant, a 2 × 2 analysis of variance (ANOVA; test item × stress position) was performed. There was no main effect for item [word vs. part-word, $F(1,32) = 0.55, p > .05$]. There was also no significant main effect for stress position [trochaic vs. iambic, $F(1,32) = 0.54, p > .05$]. Importantly, however, there was a significant interaction between item and stress position [$F(1,32) = 13.93, p < .05$], which indicates that the infants exhibited a significantly different direction of preference in the two conditions (the infants in the trochaic condition listened longer to words, and the infants in the iambic condition listened longer to part-words).

Nine-month-old infants showed a familiarity preference for words over part-words after listening to the trochaic language. This result is important, because it establishes that infants are able to match the words they segmented from fluent speech (which are marked by ST) to the test items (which are neutrally stressed). That is, the acoustic mismatch between ST and neutral stress is not so great that infants are unable to distinguish between neutrally stressed words and part-words on the basis of their experience with a language containing ST. If infants can match the items they segmented from the trochaic language to the test items, they should also be able to do so after exposure to the iambic language, because the iambic language results in the same degree of acoustic mismatch between familiarization and test items.

Although the infants were able to segment the iambic language and match the items they segmented from that language to the test items (as shown by their significant preference), the infants showed a different pattern of responses after listening to the iambic language. After exposure to the ST iambic language, the infants listened longer to the part-words. Assuming that infants continue to show the same direction of preference as they did after hearing the trochaic language—listening longer to the items that they have segmented from fluent speech—these results indicate that the infants in the iambic language condition treated part-words as words. Such a pattern of missegmentation would occur only if the infants were being consistently misled by the ST of the second syllable of the iambic language. Thus, the finding that the infants listened longer to part-words (which were stressed on their first syllable during familiarization) after hearing the iambic language suggests that they are segmenting part-words as words.

This is exactly the same pattern that 9-month-old infants showed in a previous experiment (Thiessen & Saffran, 2003), in which they were exposed to fluent trochaic and iambic synthesized speech containing multiple cues to stress: altered pitch, amplitude, and duration (with exactly the same word order and statistical cues; for a description of the acoustics of these stimuli, see Experiment 2). In Thiessen and Saffran's (2003) experiment, infants missegmented the iambic language and showed the opposite pattern of preference after hearing the two languages. As is shown in Figure 1, after hearing the trochaic language with multiple cues to stress (MC trochaic), infants showed a familiarity preference for words: They listened to words for 7.03 sec ($SE = 0.34$) and to part-words for 6.43 sec ($SE = 0.36$). By contrast, after hearing the MC iambic

language, infants listened to words for 7.73 sec ($SE = 0.51$) and to part-words for 8.92 sec ($SE = 0.59$); the same comparisons as those in the present experiment (words vs. part-words in both conditions, the cross-condition interaction) were significant.

To more thoroughly compare the results of Thiessen and Saffran's (2003) Experiment 1, we entered the data from the participants in that experiment (16 in the MC trochaic condition, 15 in the MC iambic condition), along with the data from the present experiment, into a $2 \times 2 \times 2$ (stress position $\times$ stress type $\times$ test item) ANOVA. There was a main effect for stress position [trochaic vs. iambic, $F(1,61) = 5.49, p < .05$], due to the fact that overall looking time was longer in the iambic conditions of the two experiments. This may have resulted from the more unfamiliar rhythmic pattern holding the infants' interest longer and increasing looking times. There was a trend toward a main effect for stress type [multiple cues vs. spectral tilt, $F(1,61) = 3.04, p = .09$], because the infants' overall looking times were slightly longer in Thiessen and Saffran's (2003) Experiment 1 than in the present experiment. There was no main effect for test item [words vs. part-words, $F(1,61) = 1.88, p > .05$]. There was a significant interaction between test item and stress position [$F(1,61) = 27.94, p < .05$], because in both the natural stress and the ST conditions, the infants showed the opposite pattern of preference after the iambic language than after the trochaic language. There were no other significant effects.

Critically, there were no significant interactions involving stress type (multiple cues vs. ST) in this $2 \times 2 \times 2$ ANOVA. This indicates that the pattern of preference for words and part-words after exposure to the trochaic and iambic languages was not significantly different whether the segmentation languages used multiple cues (pitch, duration, and amplitude) to realize stress or only one cue (ST). Although the absolute looking times are slightly different in Thiessen and Saffran's (2003) Experiment 1 than in the present ST conditions, the pattern is identical: Infants show a different preference after hearing an iambic language than after hearing a trochaic language, regardless of whether those languages contain full natural stress or only one cue to stress (ST). These results suggest that whether a language contains one or multiple stress cues, 9-month-old infants segment different items from fluent iambic speech than from fluent trochaic speech.

It is important to note that the interpretation that infants segmented different items from the trochaic and the iambic speech depends on the assumption that infants show the same direction of preference after listening to both languages. However, previous research has indicated that infants at this age are likely to show the same direction of preference after segmenting both a trochaic and an iambic language (e.g., Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003, 2004). Furthermore, since the trochaic language contains no conflicting cues to word boundaries (both cues, stress and statistical information, point to the same word boundaries), it should be easier to segment than the iambic language. If infants show a familiarity prefer-

ence in an easy task, they will continue to show a familiarity preference in a more difficult task (Hunter & Ames, 1988).

These results suggest that, at least in this paradigm, a single cue to stress (ST) has the same effect as multiple cues to stress: nine-month-old infants missegment given either of these manifestations of stress placed on the second syllables of words (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). This stands in stark contrast to existing results with adults, where one cue to stress influences adults' responses far less effectively than do multiple cues (e.g., Streeter, 1978). However, previous adult studies employed different experimental paradigms, such as locating clause boundaries. It may be that word segmentation is particularly sensitive to the presence of stress cues. In order to ascertain whether the infants' present performance is truly different from adults', we need to assess the performance of adults in an experimental paradigm that is closer to the one used in Experiment 1. This was the motivation for Experiment 2.

EXPERIMENT 2

Infants in a word segmentation task are as strongly influenced by a single cue to stress (ST) as they are by syllables that have multiple cues to stress. Is this true of adults as well? In order to answer this question, we first need to know how adults will perform when they are asked to segment iambic and trochaic languages containing multiple cues to stress. It is unclear whether adults will missegment (or fail to segment) consistently iambic fluent speech. Experiments with infants indicate that they can successfully segment iambic words from fluent speech by 11 months (Jusczyk, Houston, & Newsome, 1999). However, other results indicate that English-hearing adults are biased to expect trochees and have more difficulty segmenting iambs (Cutler & Norris, 1988). If adults segment a naturally stressed iambic language less proficiently than they segment a trochaic language (as infants do), it is possible to ask whether one cue to stress (ST) interferes with adult word segmentation in the way it interferes for 9-month-old infants.

Therefore, it is necessary to first determine how adults use multiple cues to stress as cues to word boundaries. Adults should segment fluent speech more easily when stress cues are placed on the first syllables of words than on the second syllables. If this is the case, we can then ask how adults use only one cue to stress (ST) as a cue to word boundaries. For infants, one cue to stress appears to be as effective a cue to word boundaries as multiple cues. If this is the case for adults, adults should perform as poorly when only one cue to stress is placed on the second syllable of words as when multiple cues to stress make a word iambic.

Method
Participants. The participants were 60 native English-speaking undergraduates from the University of Wisconsin–Madison, who received extra credit in an introductory psychology course. Of the 30

participants, 15 were randomly assigned to the MC trochaic condition, 15 were randomly assigned to the MC iambic condition, 15 were randomly assigned to the ST trochaic condition, and 15 were randomly assigned to the ST iambic condition.

Stimuli. All the languages were generated by the MacInTalk speech synthesizer running on a PowerMac 5300. These languages consisted of four disyllabic nonsense words: "dapu," "dobi," "bugo," and "diti." Each language consisted of the same words spoken in the same randomized order as that used for the languages in Experiment 1. There were no pauses between words, and all the syllables were fully coarticulated. The synthesizer produced syllables with a monotonic $F0$ of 200 Hz.

For the participants in the two ST conditions, the segmentation languages were identical to those used in Experiment 1. In the MC conditions, both segmentation languages were 2 min 20 sec in duration. Stress was synthesized by altering three parameters of the stimuli: vowel length, amplitude, and pitch (Lieberman, 1960). In the iambic language, each word was stressed on the second syllable, and in the trochaic language, each word was stressed on the first syllable.

According to Crystal and House (1987), the ratio between stressed and unstressed syllable duration in fluent speech is approximately 2:1. Most of the syllable lengthening due to stress occurs on the vowels. Although the consonants in stressed syllables do lengthen, the ratio between stressed and unstressed consonants is not as large; it is approximately 1.3:1 (Crystal & House, 1987). Since consonants are much shorter than vowels to begin with, vowel lengthening makes a much greater contribution to the increase in the duration of stressed syllables.

Vowels were lengthened to match Crystal and House's (1987) estimates of the ratio of stressed vowels to unstressed vowels (range, 1.8:1 to 2:1; mean, 1.87:1). Consonants were not lengthened, due to the fact that consonant lengthening could result in the extension of formant transitions, which could, in turn, make the consonants themselves more difficult to recognize. Thus, the stressed syllables were, on average, 310 msec long, and the unstressed syllables were all approximately 185 msec long. This ratio (1.67:1) is a close match with Crystal and House's (1990) report of the ratio of stressed CV syllables to unstressed CV syllables in the fluent speech of a fast talker (1.85:1).

Amplitude and fundamental frequency also increase in stressed syllables. Stressed syllables can have a peak amplitude between 4 and 8 dB higher than unstressed syllables (Bernstein-Ratner & Pye, 1984; Schwartz, Petinou, Goffman, Lazowski, & Cartusciello, 1996); the stressed syllables in this experiment were 4 dB louder (average RMS value) than their unstressed counterparts. Amplitude was increased across the spectrum so that ST was unaltered. Note that amplitude is altered in both the MC and the ST languages: In the MC languages, amplitude is increased across the frequency spectrum by a consistent 4 dB, whereas in the ST language, a filter alters amplitude differentially across the spectrum. Thus, the MC languages do not have an altered ST consistent with naturally produced stressed syllables. Instead, they have an increased amplitude that is consistent with the overall amplitude increase in naturally produced stressed syllables.

$F0$ values in the MC languages were based on the pitch contours of an adult female native speaker of English. Average pitch peak values varied from 255 to 270 Hz, depending on the vowel. The pitch contour, resynthesized in stressed syllables, using Kay Elemetrics' Analysis and Synthesis Lab, was somewhat different depending on whether the syllable began with a voiced or a voiceless consonant. In the case of a voiced consonant, the pitch contour was a roughly inverted parabola, whereas in the case of voiceless consonants, the pitch contour described a falling plateau. This is due to the fact that during the onset of a voiceless consonant, there are no glottal pulses and, thus, no $F0$ contour (Stevens, 1998). Therefore, when voicing began in the syllable, after a voiceless consonant, the value of the $F0$ was roughly where it would have been had the consonant been

voiced, rather than starting at the ambient value (200 Hz). The beginning part of the pitch contour parabola has, in these cases, simply been cut off. Unstressed syllables remained at the monotonic pitch of 200 Hz at which they were synthesized.

Both the MC trochaic and the MC iambic languages have previously been used in experiments with infants (Thiessen & Saffran, 2003). Of most importance, the iambic language has been found to cause infants to missegment and treat part-words, which are stressed on their first syllable during familiarization, as words. Thus, even though this language is synthesized, it results in the same pattern of word segmentation as natural speech (e.g., Johnson & Jusczyk, 2001; Jusczyk, Houston, & Newsome, 1999). Finally, the word order and statistical properties of these languages are identical to those used in Experiment 1.

Intelligibility and perception of stress. Because these stimuli are synthesized, it is unclear whether adults will treat them as stressed, even though they have been altered to display many of the auditory characteristics associated with stressed syllables. One way to alleviate this concern is to ascertain whether this synthesized speech elicits the same behavioral response as natural stress. To that end, Thiessen and Saffran (2003) used these stimuli with 9-month-old infants and found that the infants used the synthesized stress as a cue to word segmentation, just as they use natural stress. Another way to address this concern is to ask whether these synthesized syllables give rise to the perception of stress.

To answer this question, we excised two tokens of each nonsense word ("diti," "bugo," "dapu," and "dobi") from the synthesized MC trochaic, MC iambic, ST trochaic, ST iambic, and neutrally stressed fluent speech and presented them in isolation to 10 naive adult listeners. The participants heard a total of 40 words, 8 from each category, and were asked to perform two tasks: first, to transcribe the word, and second, to decide whether the word was stressed on the first syllable, the second syllable, or neither syllable.

Table 1 shows adults' response rates (out of a possible 8) to each of the five types of words: MC trochaic, MC iambic, ST trochaic, ST iambic, and neutrally stressed words. The participants responded correctly (i.e., selecting the first syllable as stressed for the MC trochaic words, the second syllable as stressed for the ST iambic words, etc.) at a rate significantly above chance in response to each of the five word types; a score of chance would be 2.67 out of 8.

After exposure to our MC speech, adults are clearly able to detect which syllable was altered in a manner consistent with lexical stress. They are also able to do so after exposure to the ST speech, but seemingly not as easily. To assess this difference, we performed a 2×2 (stress position $\times$ stress type) ANOVA. There was a significant main effect for stress position [first or second syllable, $F(1,36) = 5.65, p < .05$], indicating that the participants were more successful at identifying first-syllable stress than second-syllable stress. Most important for the present discussion, there was also a main effect of stress type [multiple cues vs. ST, $F(1,36) = 17.95, p < .05$]. The participants were more successful identifying the position of multiple cues to stress than that of ST alone. This suggests that although

Table 1
Average Number of Times (Out of Eight) Adults Identified the First, the Second, or Neither Syllable as Stressed After Hearing Multiple-Cue (MC) Trochaic, MC Iambic, Spectral Tilt (ST) Trochaic, ST Iambic, and Neutrally Stressed Words

Stress	First Syllable		Second Syllable		Neither	
	Aver.	SE	Aver.	SE	Aver.	SE
MC trochaic	7.5	0.3	0.4	0.3	0.1	0.1
MC iambic	1.6	0.6	5.5	0.8	0.9	0.3
ST trochaic	4.8	0.5	1.6	0.6	1.6	0.4
ST iambic	1.2	0.4	4.3	0.7	2.5	0.5
Neutral	1.7	0.4	1.3	0.4	5.0	0.7

both types of speech gave rise to the perception of stress, multiple cues did so more effectively than ST.

Adults were also asked to transcribe the words, a step that was taken to assess their intelligibility. Each word had two consonants and two vowels, and adults received a score of 0–4, based on how many of these they identified correctly (participants' transcriptions contained very few additions, so for the sake of simplicity we did not score them). Overall, the participants were 98.6% correct. Percentage correct varied very little across conditions. After hearing MC trochaic words, the participants' transcriptions were 97.5% correct; after MC iambic, their score was 98.7% correct; after ST trochaic, 98.7%; after ST iambic, 98.7%; and after hearing neutrally stressed words, the participants' transcriptions were 99.3% correct. There were no significant differences across conditions, suggesting that the phonemes were equally intelligible whether they occurred in words marked by natural stress, ST, or no stress.

Procedure. Since the head turn preference procedure is inappropriate for use with adults, we used a forced-choice methodology in this experiment. The adults were seated in individual booths while the familiarization language played. After the segmentation language ended, the adults received 12 forced-choice test trials. In each trial, two test items were played, one a word and one a part-word. Trials were arranged pseudorandomly, with the constraint that words occurred first on half of the trials and part-words occurred first on the other half. There was a 1-sec pause between items on each trial, and there was a 5-sec pause between trials. The participants were asked to circle a number (1 or 2), corresponding to which test item, the first or the second, sounded more like the fluent speech they had heard in the first part of the experiment. Circling the number corresponding to the word was scored as a correct answer, whereas circling the number corresponding to the part-word was scored as an incorrect answer. There were four test items (two words and two part-words), and these were identical to the four test items used in Experiment 1.

Although this methodology is somewhat different from the test procedure used with the infants in Experiment 1, previous studies have suggested that these procedures yield comparable results across the two age groups (e.g., Saffran & Griepentrog, 2001). Although the adults are asked only to identify which test items sound most like the language they heard (and are not explicitly asked to segment the fluent speech), these results provide insight into the participants'

segmentation. All test items, both words and part-words, occurred equally often in the fluent speech. Thus, if the adults indicated that the words were more reminiscent of the fluent speech, they must have segmented the languages. If the adults were not able to distinguish between the words and the part-words, there would be no reason for them to have a response bias toward the words.

Results and Discussion

First, we analyzed the participants' performance after exposure to the languages containing multiple cues to stress (the MC languages). As is shown in Figure 2, the adults averaged 7.9 (out of 12) correct ($SE = 0.7$) after hearing the MC trochaic language. This level of performance is significantly different than chance [$t(14) = 2.61, p < .05$]. The fact that the adults were consistently able to choose words over part-words indicates that they successfully segmented the trochaic language.

The adults performed more poorly after exposure to the MC iambic language (see Figure 2); they averaged 4.9 (out of 12) correct ($SE = 0.6$), a level that is not quite significantly worse than chance [$t(14) = 1.76, p = .10$]. A paired t test comparing performance on the two languages indicates that performance was significantly worse in the iambic language than in the trochaic language [$t(14) = 3.35, p < .05$].

After hearing the trochaic language, the adults successfully distinguished words from part-words. This is a task at which they failed after hearing the iambic language (in fact, the trend toward choosing part-words over words indicates possible missegmentation). These results suggest that adults, like infants, treat syllables with multiple cues to stress as word onsets. By contrast, our participants showed a very different pattern of results after exposure to the ST languages.

As is shown in Figure 2, adults averaged 9.3 (out of 12) correct ($SE = 0.7$) after hearing the ST trochaic language. This level of performance is significantly different from
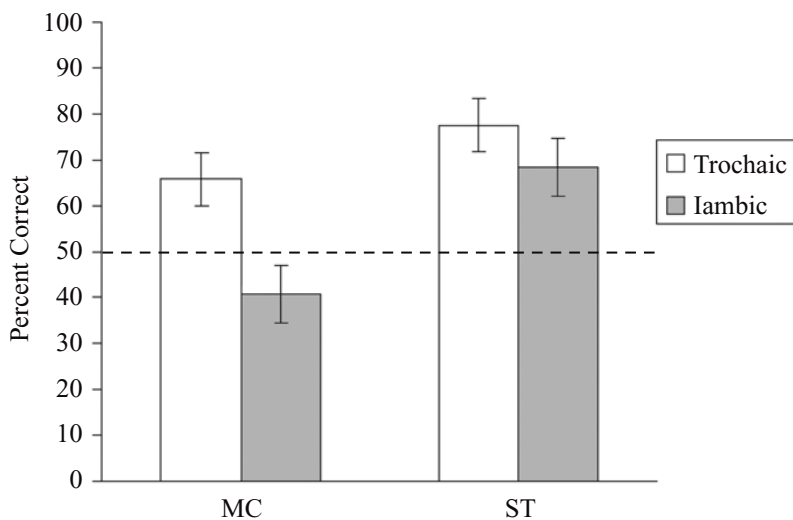
Figure 2. Percentages correct word versus part-word discriminations by adults in the multiple-cue (MC) trochaic, MC iambic, spectral tilt (ST) trochaic, and ST iambic conditions (the dashed line represents chance).

chance [$t(14) = 4.42, p < .05$]. The fact that the adults were consistently able to distinguish between words and part-words suggests that they successfully segmented the trochaic language.

The adults performed just as well after exposure to the ST iambic language (see Figure 3); they averaged 8.2 (out of 12) correct ($SE = 0.6$), a rate that is significantly better than chance [$t(14) = 2.57, p < .05$]. A paired t test indicated that there were no significant differences between the participants' responses after hearing the trochaic versus the iambic language [$t(14) = 1.01, p > .05$].

To more thoroughly assess the differences between the participants' performances across the four conditions, we performed a 2×2 (stress position $\times$ stress type) ANOVA. There was a significant effect of stress position [first vs. second syllable, $F(1,56) = 17.4, p < .05$]. Across conditions, performance was better after exposure to the trochaic languages than after that to the iambic languages, mainly due to the participants' success at the MC trochaic language, in comparison with the MC iambic language. There was also a significant effect of stress type [multiple cues vs. spectral tilt, $F(1,56) = 22.2, p < .05$]. This is partially due to the fact that the participants scored better than chance after hearing both ST languages, whereas they only scored above chance after hearing one of the MC languages. It may also partially be due to the fact that the adults were better at recognizing the neutrally stressed test items after hearing the ST languages than after hearing the MC languages, since the MC languages had more acoustic parameters altered (amplitude, pitch, and duration), in comparison with the test items. Most important, there was a significant interaction between stress position and stress type [$F(1,56) = 3.9, p < .05$]. The difference between the participants' performances after hearing the trochaic and the iambic languages was significantly greater in the MC condition than in the ST condition.

Unlike infants, adults perform equally well whether the flattened ST occurs in the first syllable of words or the second syllable. As the results of the MC conditions show, this is not because adults are insensitive to stress as a cue to word boundaries. Instead, adults seemingly do not consider flattened ST to be as strong a cue to word boundaries as natural stress. Therefore, they are able to successfully segment both the ST trochaic and the ST iambic languages (likely by attending to statistical information). Nine-month- old infants, by contrast, are more influenced both by a single cue to stress (ST) and by multiple convergent cues to stress (amplitude, pitch, and duration) than by statistical information.

The results of Experiments 1 and 2, taken together, imply a developmental progression in infants' responses to stress cues. At an early age, infants consider ST to be a strong cue to word boundaries. However, by adulthood, English speakers have a different response to ST, and do not consider ST to be a strong cue to word boundaries. This response to ST is more appropriate to the acoustics of the language, where ST flattening rarely occurs in the absence of other cues to stress, and never as a sole cue to word boundaries (Blomgren, Chen, Ng, & Gilbert, 1998).

Although our results suggest a developmental progression, they do not specify the nature of that progression. There are a number of possible explanations and time lines, since there is a huge developmental gap between 9-month-olds and adults. When, in that time, do infants learn that ST is not a particularly strong cue to word boundaries, and what are the mechanisms that drive that development? Two different kinds of experiences may play a role. Infants may learn that ST is a poor cue to word boundaries simply by gaining more exposure to speech. As infants become more experienced with lexical stress, they may gain a more adultlike understanding of the contribution of the various acoustic factors that contribute to the perception of a syllable as *stressed*.

The second experience that may be important is producing stressed syllables. When infants begin to talk and to produce a distinction between stressed and unstressed syllables, they may realize that their own production of stress alters all of the parameters of stress (amplitude, duration, pitch, and ST), and not just one. This, in turn, may lead them to shift from reacting to any cue as a good cue to word boundaries to a more adultlike response, requiring the presence of most or all the cues in a syllable before treating a syllable as stressed.

One of the remarkable features of infant babbling is that it contains rhythmic structure, and infants produce both stressed and unstressed syllables as they babble, marking the distinction with (at least) pitch, amplitude, and duration (Davis, MacNeilage, Matyear, & Powell, 2000). However, even by 18 months, toddlers have difficulty manipulating the parameters of stress in a linguistically appropriate manner. That is, their marking of stressed and unstressed syllables in words is often absent or acoustically indistinct (Kehoe, Stoel-Gammon, & Buder, 1995; Schwartz et al., 1996; Scukanec & Watson, 1995). Therefore, if some minimal production or perception of stress is all that is necessary for infants to note that ST, alone, plays little functional role in identifying word boundaries, we would expect infants to begin to disregard ST during the babbling stage, which usually occurs between 6 and 12 months. If, by contrast, proficiency at using stress to functionally mark stressed and unstressed linguistic units plays a decisive role in shaping infants' and toddlers' representation of stress, we would not expect to see a developmental shift before 18 months of age. In order to further investigate the time course of infants' use of syllables marked only by ST versus their use of syllables marked by multiple cues to stress, we exposed a group of 11- to 12-month-old infants to both natural and ST trochaic and iambic languages.

EXPERIMENT 3

How soon do infants become more adultlike in their use of the acoustic parameters of stress as cues to word boundaries? This development may be a function of early productive or perceptual experience, or it may occur in response to later-arising abilities to proficiently mark stressed and unstressed syllables in a lexically appropriate manner.
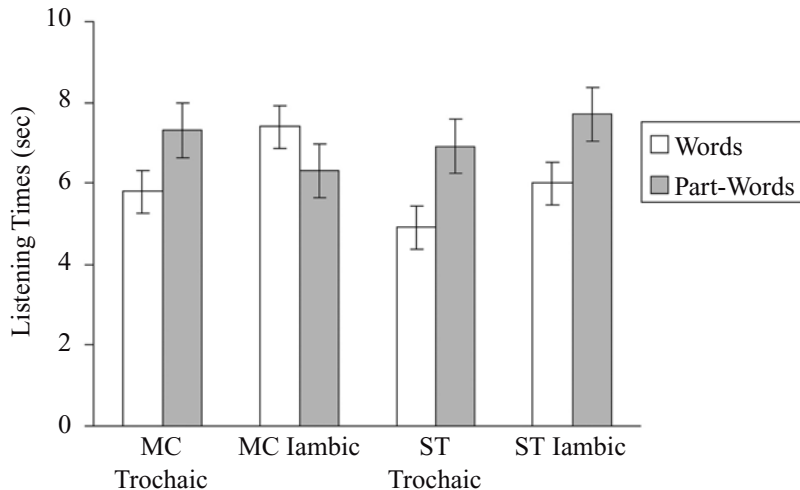
Figure 3. Eleven-month-old infants' looking time to words and part-words in the multiple-cue (MC) trochaic, MC iambic, spectral tilt (ST) trochaic, and ST iambic conditions.

Since 12-month-old infants are not yet proficient at the production of lexical stress (although they can produce rhythmic contour), in Experiment 3 we explored their response to multiple parameters of stress versus ST alone as a cue to word boundaries. Experiment 1, in tandem with the data from Thiessen and Saffran (2003), suggested that infants respond to syllables marked by ST in a manner very similar to the way they respond to syllables marked by multiple cues to stress: Nine-month-old infants treat both kinds of syllables as though they are strong cues to word boundaries. By contrast, Experiment 2 indicated that adults treat syllables with multiple cues to stress as word onsets, but not syllables with flattened ST. In Experiment 3, we presented 11- to 12-month-old infants with MC trochaic, MC iambic, ST trochaic, and ST iambic languages. We expected them, like 9-month-olds and adults, to use syllables with multiple cues to stress as word onsets. The question of interest was whether the placement of ST would also prompt them to change their segmentation of fluent speech, as with the younger infants, or whether their segmentation would be unaffected by the placement of the spectral tilt cue, as with adults.

Method

Participants. The participants were 52 infants between the ages of 10.9 and 12.4 months. The average age of the participants was 11.52 months. To obtain the 52 infants for this experiment, it was necessary to test 69. The other 17 were excluded for the following reasons: crying (9), failure to complete at least two of one or more trial types (5), looking times of less than 3 sec, on average, to either side light (2), and falling asleep during the experimental session (1). Of the 52 participants included in the data analysis, 13 were randomly assigned to the MC trochaic condition (average age: 11.61 months), 13 were assigned to the MC iambic condition (average age: 11.55 months), 13 were assigned to the ST trochaic condition (average age: 11.45 months), and 13 were assigned to the ST iambic condition (average age: 11.48 months).

Stimuli. Because 12-month-old infants are generally unwilling to sit on a parent's lap for the same length of time as younger infants, we used shorter versions of the MC and ST languages from Experiments 1 and 2. In the original versions of those languages, frequent words occurred 90 times, and infrequent words occurred 45 times. In the languages used in the present experiment, frequent words occurred 60 times, and infrequent words occurred 30 times. The MC trochaic and iambic languages were 1 min 15 sec in duration; the ST languages were 1 min in duration (the lack of durational stress cues rendered this language shorter than the naturally stressed stimuli). The test items (two words and two part-words) were identical to those used in Experiment 1. All the infants received the same test items, produced without stress cues.

Procedure. The procedure was identical to that in Experiment 1.

Results and Discussion

First, we compared the listening times to words and part-words for infants exposed to the MC trochaic language. As is shown in Figure 3, the infants listened to words for 5.8 sec ($SE = 0.4$) during the test trials and to part-words for 7.3 sec ($SE = 0.7$). Ten of the 13 infants listened longer to part-words than to words during the test trials after exposure to the trochaic language. A paired t test indicated that the difference in looking times between words and part-words was significant [$t(12) = 2.33, p < .05$].

Second, we compared listening times to words and part-words for infants exposed to the MC iambic language. As is shown in Figure 3, the infants listened to words for 7.4 sec ($SE = 0.8$) and to part-words for 6.3 sec ($SE = 0.7$). Nine of the 13 infants listened longer to words than to part-words during the test trials after exposure to the iambic language. A paired t test indicated that the difference in looking times between words and part-words was not significant [$t(12) = 1.45, p = .17$].

Infants showed very different patterns of preference after exposure to the two MC languages. In order to assess that these patterns were truly different, we performed a 2×2

(stress position × test item) ANOVA. There was no main effect of stress position [trochaic vs. iambic, $F(1,24) = 0.132, p > .05$]. There was also no main effect of test item [word vs. part-word, $F(1,24) = 0.275, p > .05$]. Importantly, though, there was a significant stress position × test item interaction [$F(1,24) = 6.99, p < .05$]. This indicates that the infants' patterns of preference across the two conditions were significantly different.

After listening to the MC trochaic language, the 11- to 12-month-old infants listened longer to part-words than to words. This distinction indicates that the infants successfully segmented the trochaic language and showed a novelty preference. This direction of preference differs from the 9-month-old infants tested in Experiment 1, who showed a familiarity preference; the difference in direction of preference is not surprising, since this group of infants was much older, and older infants are far more likely, given comparable tasks, to show a novelty preference (Hunter & Ames, 1988).

In contrast, after listening to the MC iambic language, the 11- to 12-month-old infants did not show a significant distinction between words and part-words. This suggests that the infants found this language more difficult to segment. These results differ from previous experiments with 9-month-olds (e.g., Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003), where infants have shown a reliable preference after hearing iambic languages, and one that suggests that they have missegmented, rather than no preference. We believe this is due to the fact that by 11–12 months, infants are no longer attending primarily to one cue (lexical stress) but attempt to synthesize multiple cues (Jusczyk, 1999; Morgan & Saffran, 1995). Because the iambic language contains multiple conflicting cues (stress and statistics) to word boundaries, infants are unable to integrate the various cues and successfully segment the language.

The infants showed a very different pattern after exposure to the ST languages. As is shown in Figure 3, after exposure to the ST trochaic language, the infants listened to words for 4.9 sec ($SE = 0.6$) and to part-words for 6.9 sec ($SE = 0.6$). Twelve of the 13 infants listened longer to part-words than to words. A paired t test indicated that the difference in looking times between words and part-words was significant [$t(12) = 5.73, p < .05$].

As is shown in Figure 3, after exposure to the ST iambic language, the infants listened to words for 6.0 sec ($SE = 0.4$), and to part-words for 7.7 sec ($SE = 0.6$). Eleven of the 13 infants listened longer to part-words than to words during the test trials after exposure to the ST iambic language. A paired t test indicated that the difference in looking times between words and part-words was significant [$t(12) = 3.47, p < .05$].

After exposure to the two kinds of MC languages (trochaic and iambic), the infants showed different patterns of segmentation, indicating that they found the trochaic language easier to segment than the iambic language. By contrast, the infants showed the same pattern (a novelty preference) after exposure to either the ST trochaic or the

ST iambic language. To assess whether this difference was significant, we performed a 2 × 2 × 2 (stress type × stress position × test item) ANOVA. There was no main effect of stress type [MC vs. ST, $F(1,48) = 0.37, p > .05$], since looking times were approximately equal across stress type conditions. Similarly, there was no main effect of stress position [trochaic vs. iambic, $F(1,48) = 1.21, p > .05$]. There was, however, a main effect of item [words vs. part-words, $F(1,48) = 13.66, p < .05$]. This was due to the fact that the infants listened longer to part-words than to words in three out of the four conditions, so across conditions the looking time to part-words was significantly greater.

There was not a significant interaction between stress type and stress position [$F(1,48) = 0.31, p > .05$]. Critically, however, there was a significant interaction between stress type and test item [$F(1,48) = 7.85, p < .05$]. In the MC conditions, the infants showed different patterns of preference after exposure to the iambic and the trochaic stress position. In the ST conditions, the infants showed the same pattern of preference; this interaction indicates that the difference in those patterns was significant and indicates that the infants responded differently to the two kinds of stress as cues to segmentation. There was also an interaction between stress position and test item [$F(1,48) = 6.66, p < .05$], because the infants in both the MC and the ST trochaic condition listened longer to part-words, whereas the infants in the two iambic conditions showed conflicting patterns: In the MC iambic condition, the infants listened longer to words, whereas in the ST iambic condition, the infants listened longer to part-words. The three-way interaction between stress type, stress position, and test item was not quite significant [$F(1,48) = 3.72, p = .06$].

These results suggest that by 12 months, ST is a much less powerful cue to word boundaries than are multiple convergent cues to stress. Whereas multiple cues are salient enough to negate statistical information to word boundaries, ST is not. Since the statistical information is likely no more accessible in the ST languages than in the languages containing multiple cues to stress (both languages contain the same statistical information), ST must be a weaker cue to word boundaries than are multiple stress factors occurring together, for infants at this age.

GENERAL DISCUSSION

The results of these three experiments, taken together, suggest a rapid developmental course in infants' response to one acoustic cue, stress, to word boundaries. As was seen in Experiment 2, adults consider ST to be a less effective cue to word boundaries than alterations to pitch, amplitude, and duration, occurring simultaneously. This result is consistent with past research in other paradigms, which indicates that adults' responses are not affected as much by one cue to stress as by multiple cues to stress (e.g., Streeter, 1978). Experiment 1 indicates that 9-month-old

infants—at least in this paradigm—respond to syllables marked only by ST in the same manner that they respond to syllables marked by multiple cues to stress. At 9 months, infants respond to both kinds of syllables as though they are both strong cues to word boundaries. However, Experiment 3 suggests that by 11–12 months, infants are behaving in a more adultlike fashion and treat ST as a less strong index of word boundaries than is a convergence of multiple cues to stress.

One explanation for this pattern of results is that infants at 9 months are particularly insensitive to statistical information. That is, although they will attend to sequential statistical cues when no other cues are present, it is the least important cue to word boundaries for 9-month-old infants. Therefore, even though they, like adults, consider ST to be a less important cue to word boundaries than multiple stress cues are, ST is still more salient than any available statistical information. This hypothesis is consistent with the fact that lexical stress does seem to be a more potent cue to word boundaries than statistical information at 9 months (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). However, recent results from our laboratory suggest that 9-month-old infants can be made to favor statistical cues over stress cues after only a brief exposure to iambs in isolation, which appears to shake their reliance upon trochaic stress (Thiessen & Saffran, 2004). The fact that a relatively brief (2-min) training can cause infants to weight statistical information more heavily than stress cues is inconsistent with the suggestion that stress cues are so much more salient at 9 months that even a weak cue to the presence of stress (ST) would be more potent than statistical information.

Another possible explanation for the developmental progression that emerged in our results is that infants, between 9 and 12 months, become more experienced with the distribution of acoustic factors that contribute to the presence or absence of lexical stress. At 9 months, infants have a very general representation of stress. Thus, it is possible that any factor that is associated with stress (amplitude, duration, pitch, ST) is as good or nearly as good an index of lexical stress as all of these cues functioning in concert. Alternatively, infants may consider ST to be a particularly important cue to the presence of stress. Regardless, at this age, infants are not yet as sophisticated at identifying the presence or absence of stress as adults, who locate stress flexibly and in a gradated manner, depending on how many cues are present and how fully those cues are articulated (Streeter, 1978). This may be because they lack experience with stress, or because the distribution of stress cues in infant-directed speech is different from the distribution of stress cues in adult-directed speech.

By 12 months, however, infants have begun to realize that one cue to stress is not as reliable an indicator of word boundaries as the presence of multiple cues to stress. This may be due to increased experience at hearing lexical stress or at producing cues to stress themselves. It is also possible that word segmentation plays a role in helping infants to refine their ability to use the acoustic correlates of stress as cues to word boundaries. At 12 months, infants

have much more experience at segmenting words from fluent speech than 9-month-old infants do (e.g., Jusczyk & Aslin, 1995). At 9 months, infants may be willing to accept a wide range of acoustic variance as cues to word boundaries, simply because they have not yet had sufficient experience to note that these acoustic events are uninformative. By 12 months, infants may have learned that certain cues to word boundaries are more reliable than others in their native language. Furthermore, they may have realized that multiple acoustic cues are more likely to indicate a word boundary than is one acoustic cue appearing in isolation (Jusczyk, 1999).

Whichever of these proposed accounts best describes the developmental progression between 9 and 12 months, it is clear that 9-month-old infants react very differently to the acoustic correlates of stress than do adults. Furthermore, they appear to progress toward an adultlike state very quickly, showing adultlike performance at 12 months of age. Although the exact mechanism underlying this change is currently unclear, the change itself is consistent with a wide range of results from experiments with infants in this age range, suggesting an ability to rapidly identify acoustic regularities and apply those regularities to the language-learning tasks at hand (e.g., Maye, Werker, & Gerken, 2002; Saffran & Thiessen, 2003).

REFERENCES

ASLIN, R. N. (2000, July). *Interpretation of infant listening times using the headturn preference technique*. Paper presented at the International Conference on Infant Studies, Brighton, U.K.

ASLIN, R. N., SAFFRAN, J. R., & NEWPORT, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, **9**, 321-324.

BERNSTEIN-RATNER, N., & PYE, C. (1984). Higher pitch in BT is not universal: Acoustic evidence from Quiche Mayan. *Journal of Child Language*, **11**, 515-522.

BLOMGREN, M., CHEN, Y., NG, M. L., & GILBERT, H. R. (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *Journal of the Acoustical Society of America*, **103**, 2649-2658.

CRYSTAL, T. H., & HOUSE, A. S. (1987). Segmental duration in connected-speech syllables: Syllabic stress. *Journal of the Acoustical Society of America*, **83**, 1574-1585.

CRYSTAL, T. H., & HOUSE, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, **88**, 101-112.

CUTLER, A., & CARTER, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, **2**, 133-142.

CUTLER, A., & DARWIN, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics*, **29**, 217-224.

CUTLER, A., & NORRIS, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 113-121.

DAVIS, B. L., MACNEILAGE, P. F., MATYEAR, C. L., & POWELL, J. K. (2000). Prosodic correlates of stress in babbling: An acoustical study. *Child Development*, **71**, 1258-1270.

ECHOLS, C. H., CROWHURST, M. J., & CHILDERS, J. B. (1997). Perception of rhythmic units in speech by infants and adults. *Journal of Memory & Language*, **36**, 202-225.

EIMAS, P. D., & QUINN, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, **65**, 903-917.

EIMAS, P. D., QUINN, P. C., & COWAN, P. (1994). Development in exclu-

sivity in perceptually based categories of young infants. *Journal of Experimental Child Psychology*, **58**, 418-431.

FANT, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

FRY, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, **27**, 765-768.

HUNTER, M. A., & AMES, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, **5**, 69-95.

JOHNSON, E. K., & JUSCZYK, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory & Language*, **44**, 548-567.

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

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. (1993). Preference for the predominant stress pattern of English words. *Child Development*, **64**, 675-687.

JUSCZYK, P. W., HOHNE, E. A., & BAUMAN, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, **61**, 1465-1476.

JUSCZYK, P. W., HOUSTON, D. M., & NEWSOME, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, **39**, 159-207.

KEHOE, M. M., STOEL-GAMMON, C., & BUDER, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech & Hearing Research*, **38**, 338-350.

KLATT, D. H., & KLATT, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, **87**, 820-857.

LIEBERMAN, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, **32**, 451-454.

MARESCHAL, D., & QUINN, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, **5**, 443-450.

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.

MAYE, J., WERKER, J. F., & GERKEN, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, **82**, B101-B111.

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

SAFFRAN, J. R., ASLIN, R. N., & NEWPORT, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926-1928.

SAFFRAN, J. R., & GRIEPENTROG, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, **37**, 74-85.

SAFFRAN, J. R., & THIESSEN, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, **39**, 484-494.

SCHWARTZ, R. G., PETINOU, K., GOFFMAN, L., LAZOWSKI, G., & CARTUSCIELLO, C. (1996). Young children's production of syllable stress: An acoustic analysis. *Journal of the Acoustical Society of America*, **99**, 3192-3200.

SCUKANEC, G. P., & WATSON, M. M. (1995). Early prosodic development: A case study. *Infant Toddler Intervention*, **5**, 131-140.

SLUIJTER, A. M. C., & VAN HEUVEN, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, **100**, 2471-2485.

STEVENS, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

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

THIESSEN, E. D., & SAFFRAN, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries in 7- to 9-month-old infants. *Developmental Psychology*, **39**, 706-716.

THIESSEN, E. D., & SAFFRAN, J. R. (2004). *Learning to learn: Infants' acquisition of stress-based strategies for word segmentation*. Manuscript submitted for publication.

NOTE

1. Samples of these stimuli can be accessed at http://www.waisman.wisc.edu/infantlearning/infant_stimuli.htm.