The Official Journal of the International Society on Infant Studies
Infancy

Infancy, 1–32, 2011

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/j.1532-7078.2011.00110.x

WILEY-BLACKWELL

Rhythmic Grouping Biases Constrain Infant Statistical Learning

Jessica F. Hay
Department of Psychology
University of Tennessee – Knoxville

Jenny R. Saffran
Department of Psychology and Waisman Center
University of Wisconsin – Madison

Linguistic stress and sequential statistical cues to word boundaries interact during speech segmentation in infancy. However, little is known about how the different acoustic components of stress constrain statistical learning. The current studies were designed to investigate whether intensity and duration each function independently as cues to initial prominence (trochaic-based hypothesis) or whether, as predicted by the Iambic-Trochaic Law (ITL), intensity and duration have characteristic and separable effects on rhythmic grouping (ITL-based hypothesis) in a statistical learning task. Infants were familiarized with an artificial language (Experiments 1 and 3) or a tone stream (Experiment 2) in which there was an alternation in either intensity or duration. In addition to potential acoustic cues, the familiarization sequences also contained statistical cues to word boundaries. In speech (Experiment 1) and nonspeech (Experiment 2) conditions, 9-month-old infants demonstrated discrimination patterns consistent with an ITL-based hypothesis: intensity signaled initial prominence and duration signaled final prominence. The results of Experiment 3, in which 6.5-month-old infants were familiarized with the speech streams from Experiment 1, suggest that there is a developmental change in infants' willingness to treat increased duration as a cue to word offsets in fluent speech. Infants' perceptual systems interact with linguistic experience to constrain how infants learn from their auditory environment.

Correspondence should be sent to Jessica F. Hay, Department of Psychology, University of Tennessee – Knoxville, 1404 Circle Drive, Knoxville, TN 37996. E-mail: jhay@tennessee.edu

To understand spoken language, listeners must be able to locate lexical items in fluent speech. This is a nontrivial challenge, especially for infants, as the majority of infant-directed speech is produced continuously, without consistent pauses (e.g., Cole & Jakimik, 1980). One only needs to imagine listening to a foreign language, in which words seem to be spoken at lightning pace and melt into one another, to appreciate the complexity of this task. Although there are no infallible acoustic cues to word boundaries (Cutler & Carter, 1987; Klatt, 1980), natural languages do contain numerous overlapping cues. Cue redundancy may function to both bootstrap subsequent cue learning (Sahni, Seidenberg, & Saffran, 2010) and help infants accomplish this seemingly daunting task.

By the end of their first year of life, infants demonstrate sensitivity to both language-general (e.g., Saffran, Aslin, & Newport, 1996) and language-specific (e.g., Christophe, Dupoux, Bertoncini, & Mehler, 1994; Friederici & Wessels, 1993; Johnson & Seidl, 2009; Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Hohne, & Bauman, 1999a; Jusczyk, Houston, & Newsome, 1999b; Mattys & Jusczyk, 2001; Morgan, 1996; Nazzi, Dilley, Jusczyk, Shattuck-Hufnagel, & Jusczyk, 2005; Seidl & Cristià, 2008) cues to word boundaries. One language-general cue to word boundaries that may be particularly useful, and that has been the focus of much research, is sequential relationships between syllables: syllables that reliably co-occur often belong to the same word, while syllables that rarely co-occur are more likely to span word boundaries (e.g., Swingley, 2005). This type of sequential co-occurrence relationship is referred to as transitional probability[1] (TP), and young infants appear to possess powerful domain-general learning mechanisms that allow them to track transitional probabilities, facilitating the discovery of words in fluent speech (e.g., Aslin, Saffran, & Newport, 1998b). In fact, infants as young as five and a half months of age are sensitive to the co-occurrence relationship between syllables and can use those statistical relationships to locate "words" in an artificial language (Johnson & Tyler, 2010), English-learning 8-month-olds can track transitional probabilities in an unfamiliar natural language (Pelucchi, Hay, & Saffran, 2009a,b), and the ability to track transitional probability has been demonstrated in different infant populations, including those learning English (e.g., Pelucchi et al., 2009a,b;

[1]Transitional probability, also termed *conditional probability*, is the probability of one event given the occurrence of another event. This statistic refers to more than the frequency with which one element follows another, as it adjusts for the base rate of the first event or element. The transitional probability of Y given X is represented by the following equation:

$$\mathrm{P}(Y|X) = \frac{\text{frequency of pair } XY}{\text{frequency of } X}$$

Saffran et al., 1996), French (Mersad & Nazzi, in press), and Dutch (Johnson & Tyler, 2010). Other recent research has shown that even newborns are sensitive to statistical structure in continuous speech (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009) and nonspeech auditory streams (Kudo, Nonaka, Mizuno, Mizuno, & Okanoya, 2011), as well as in visual sequences (Bulf, Johnson, & Valenza, 2011).

Infants also show sensitivity to language-specific information embedded in fluent speech and can use these cues to segment speech (e.g., Christophe et al., 1994; Friederici & Wessels, 1993; Johnson & Seidl, 2009; Jusczyk et al., 1993, 1999a,b; Mattys & Jusczyk, 2001; Morgan, 1996; Nazzi et al., 2005; Seidl & Cristià, 2008). One language-specific cue to word boundaries that may be particularly important in English, and that has also been the focus of much research, is the pattern of stressed syllables within words: the majority of English disyllabic content words have a trochaic (strong–weak) stress pattern (Cutler & Carter, 1987). Consistent with this pattern, 7.5-month-old English-learning infants are better at recognizing trochees than iambs (weak–strong units) in fluent speech (Jusczyk et al., 1999b). Further, by 9 months of age, English-learning infants treat trochees as more coherent units than iambs (Echols, Crowhurst, & Childers, 1997) and show a listening preference for trochaic over iambic words (Jusczyk et al., 1993; Turk, Jusczyk, & Gerken, 1995). English-speaking adults are also biased to treat stressed syllables as a cue to word onset (Cutler & Norris, 1988).

In natural language, language-general cues (e.g., transitional probability) and language-specific cues (e.g., native-language stress patterns) typically provide redundant information to word boundaries. For example, consider the phrase "pretty baby." In this example, *baby* is a better word than *tyba* on at least two levels. First, the within-word transitional probability from *ba* to *by* is relatively high in speech to infants (~.80), whereas the transitional probability from *ty* to *ba* is relatively low because it spans a word boundary (~.0002). Second, *baby* has a trochaic stress pattern, whereas *tyba* has an iambic stress pattern. Thus, *baby* is a "better" English word than *tyba* both by virtue of its internal statistics and its prototypical stress pattern.

Although redundant cues potentially provide infants with a wealth of information relevant for speech segmentation, it is difficult to determine exactly which cues infants are relying on at any given stage of development. Studies that examine cues in isolation can reveal whether or not infants are sensitive to any given cue, but they cannot speak to how that particular cue is weighted with respect to other cues to word boundaries, nor can they speak to the developmental trajectory of cue weighting.

In the initial statistical learning studies (e.g., Aslin et al., 1998b; Saffran et al., 1996), 8-month-old infants were presented with an artificial language made up of nonsense words concatenated together, in which the transitional probabilities between syllable sequences were expressly manipulated, while all other potential cues to word boundaries were removed. Syllable pairs within "words" generally had high transitional probabilities (1.0) and pairs spanning word boundaries had lower transitional probabilities (.33–.5). To investigate the developmental trajectory of infants' sensitivity to stress as a cue to word boundaries, more recent studies have pitted linguistic stress against statistical regularities (Johnson & Jusczyk, 2001; Johnson & Seidl, 2009; Thiessen & Saffran, 2003, 2007). For example, Thiessen and Saffran (2003) familiarized infants with one of two artificial languages, each with a different stress pattern. In the trochaic language, the target words had initial prominence; the first syllable of each word was stressed relative to the second syllable. In this case, the stress and statistical cues provided consistent cues to word boundaries, and both 7- and 9-month-olds successfully discriminated words from sequences spanning word boundaries (part-words). In the iambic language, the target words had final prominence; the second syllable of each word was stressed relative to the first syllable. In this case, the stress and sequential statistical cues were placed in conflict with one another for these English-learning infants. Seven-month-olds continued to track sequential statistical cues, seemingly ignoring any stress cues to word boundaries. Nine-month-old infants, however, mis-segmented the words, and instead used stress as a cue to word boundaries, finding trochees (part-words spanning word boundaries) instead of the iambic target words. In follow-up studies, Thiessen and Saffran (2007) found that a brief exposure to nonsense words with either an iambic or trochaic stress pattern induced 7-month-old infants to use stress instead of statistics as their primary cue to word boundaries in a subsequent segmentation task. The results from these studies suggest that linguistic experience facilitates learning about stress cues, allowing language-specific stress cues to constrain and override statistics as a cue to word boundaries.

Thus far, we have discussed linguistic stress as though it is a discrete entity that is either applied or not applied to a syllable. However, stress is itself signaled by a constellation of acoustic correlates. How those acoustic correlates are weighted differs from one language to the next, and it is the alternation of these acoustic correlates that gives languages their characteristic rhythmic patterns. For example, in English, stressed syllables tend to be louder, longer, and higher pitched than unstressed syllables (e.g., Beckman, 1986; Delattre, 1965; Fry, 1955; Lehiste, 1970), and English listeners tend to perceive stressed syllables word initially (e.g., Cutler & Norris, 1988). In French, however, phrase-final syllables have a higher pitch, a slightly lower intensity, and a large increase in duration as compared to nonphrase-final syllables (Delattre, 1966). And although French-speaking adults have a difficult time encoding contrastive stress in their phonological representations (sometimes referred to as stress "deafness"; Dupoux, Pallier, Sebastián, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008), French listeners tend to perceive phrase-final prominence (e.g., Dell, 1984; Tranel, 1987).

The extent to which the acoustic cues to stress affect perception depends, in part, on how many cues are present in speech. For example, English-speaking adults are more likely to assign prosodic boundaries between words/syllables when those boundaries are marked by a greater number of acoustic correlates (Streeter, 1978). English-speaking adults also treat syllables marked by multiple correlates of stress as stronger cues to word boundaries than syllables marked by a single correlate of stress (Thiessen & Saffran, 2004). Not surprisingly, the majority of studies on the perception of linguistic stress have presented listeners with multiple acoustic cues to lexical stress. For example, in Thiessen and Saffran's (2003, 2007) studies, stressed syllables were louder, longer, and higher pitched than nonstressed syllables. It is of interest to understand whether infants treat individual acoustic cues to stress as cues to word boundaries, particularly because adults tend to require multiple cues before stress provides a good cue to word boundaries. In a follow-up study, using the same paradigm, Thiessen and Saffran (2004) found that 9-month-old infants do not require multiple cues to stress, but instead treat syllables with a single cue to stress (spectral tilt, which is an index of the relative balance between the intensity of low- and high-frequency energy across the spectrum) as a strong cue to word boundaries. On the other hand, 12-month-olds and adults require multiple cues to stress to override transitional probability during speech segmentation. Based on these data, Thiessen and Saffran (2004) argued that infants' representation of stress starts out broad and is narrowed through linguistic experience, with older infants requiring a more precise realization of stressed syllables if they are to function as strong cues to word boundaries. How flexible are infants' representations of linguistic stress? Will infants also treat other individual acoustic cues to stress, namely intensity, duration, and pitch, as potential cues to word boundaries?

This question is particularly interesting in light of the work suggesting that the individual acoustic cues to stress (i.e., intensity, duration, and pitch) may have characteristic and separable effects on rhythmic grouping (e.g., Bion, Benavides, & Nespor, 2011; Hay & Diehl, 2007; Iversen, Patel, & Ohgushi, 2008; Rice, 1992; Woodrow, 1909, 1951; Yoshida et al., 2010). For example, Hay and Diehl (2007) found that when a sequence of sounds alternates in intensity such that every other sound is more intense, English and French adult listeners tend to group those sounds into a two-beat rhythmic pattern with the more intense sound coming *first* (i.e., trochees). Conversely, when sounds alternate in duration, so that every other sound is longer, listeners tend to group sounds so that the longer one always comes *last* (i.e., iambs). These results provide an empirical demonstration of the principles of rhythmic grouping first described at the beginning of the 20th century to account for rhythmic grouping in music (Bolton, 1894; Cooper & Meyer, 1960; Woodrow, 1909, 1951) and later adapted for language by Hayes (1995). In his *Iambic/Trochaic Law*,[2] Hayes (1995) stated that (a) elements contrasting in intensity naturally form groupings with initial prominence (intensity principle) and (b) elements contrasting in duration naturally form groupings with final prominence (duration principle).

Hay and Diehl's (2007) results were unaffected by the native language of the participants (English versus French) or by whether or not the materials were linguistic. This pattern of results suggests that these auditory rhythmic grouping strategies may be independent of speech and linguistic experience. On this view, the principles laid out in the Iambic/Trochaic Law describe universal perceptual biases. A recent replication of Hay and Diehl (2007) with German- and French-speaking adult listeners generally supports the universality of the Iambic/Trochaic Law, while showing some language-specific modulations (Höhle, Unger, Gonzalez Gomez, & Nazzi, 2011).

Other recent research suggests that while the intensity principle may be universal, the duration principle is not. Using a task similar to Hay and Diehl (2007), Iversen et al. (2008) found that both English- and Japanese-speaking adult listeners grouped intensity-varying nonspeech sequences trochaically. However, when presented with duration-varying sequences, only the English-speaking adults perceived iambic units; Japanese-speaking adults did not show any predictive grouping patterns for duration-varying sequences.

Comparing the findings of Hay and Diehl (2007) with those of Iversen et al. (2008) leads to a conundrum: why does one study support universality of the Iambic/Trochiac Law, whereas the other suggests that the law is susceptible to linguistic experience? The answer may be found in cross-linguistic similarities and differences in phrase structure. Although French and English differ in their placement of stressed syllables, they share a predominantly iambic phrasal structure; in both languages, functors, which are typically monosyllabic and very short because of vowel reduction, precede content words, which may be mono- or multi-syllabic, fully articulated, and thus tend to be longer (e.g., *the dog/le chien*). English and Japanese, however, differ in their phrasal structure; Japanese tends to place functors after content words (Baker, 2001), leading to a predominantly trochaic phrasal structure.

[2]Although the Iambic/Trochaic Law does not predict the effects of pitch on rhythmic grouping, it has been shown that pitch may also play a crucial role in signaling trochees at the phrasal level (Bion et al., 2011; Nespor et al., 2008) and lead to trochaic rhythmic grouping (Rice, 1992).

Thus, Iversen et al. (2008) proposed that duration-based grouping biases may not be tied to linguistic stress but may instead emerge as a result of experience with native-language phrasal structure (for an alternate account based on the head/complement structure across languages, see Nespor et al., 2008). On this view, both English and French listeners have experience with short (functors) followed by longer (content words) sound sequences, and thus group duration-varying sequences iambically. Japanese listeners have more experience with long–short (content word – functor) sequences and thus group duration-varying sequences trochaically.

Developmental studies provide contradictory evidence for the universality of the grouping principles set out in the Iambic/Trochaic Law. For example, Trainor and Adams (2000) used a gap detection task to study the role of duration and intensity in auditory segmentation. They presented 8-month-old English-learning infants and adults with a sequence of complex tones that were identical, with the exception that every third tone was either longer (Experiment 1) or more intense (Experiment 2). They reasoned that gaps should be easier to detect if they fall within a segmented unit (thereby interrupting the unit) than if they fall after a segmented unit (leaving the segmented unit intact). Consistent with the duration principle, infants and adults were less likely to notice a gap following a long tone (segmentation boundary) than a gap between two short tones or between a short and a long tone. Contrary to the intensity principle, there were no differences in gap detection performance when every third syllable was more intense for either age group.

Yoshida et al. (2010) proposed a developmental trajectory for duration-based rhythmic grouping. In their study, English- and Japanese-learning infants were familiarized with a sequence of duration-varying tones. Subsequently, both English- and Japanese-learning 7- to 8-month-olds discriminated short–long from long–short tone sequences in a preferential listening paradigm. English-learning 7- to 8-month-olds preferred listening to trochees (long–short sequences), whereas Japanese-learning 7- to 8-month-olds preferred listening to iambs (short–long sequences). Younger infants (i.e., 5- to 6-month-olds), from both language backgrounds, failed to show consistent listening preferences. As there was no control group (i.e., a group that did not receive familiarization with the duration-varying tone sequence), it is possible that listening preferences in the older age group were not linked to the familiarization sequence, but instead reflect language-specific biases (e.g., a trochaic bias for English-learning infants).

Finally, Bion et al. (2011) familiarized 7- to 8-month-old Italian-learning infants and adults to sequences of syllables that alternated in either duration or pitch. At test, they found that syllable duration influenced the grouping of, and memory for, short–long speech sequences in adults but not in infants, whereas syllable pitch influenced rhythmic grouping of high pitch-low pitch sequences in both age groups.

Given that intensity and duration both signal syllable stress in English but lead to opposite rhythmic grouping strategies for English-speaking infants and adults, it is unclear whether these individual acoustic correlates of stress will, like spectral tilt, function as a strong enough cue to word boundaries in infancy to override statistical cues. We expect that infants should be sensitive to both sequential statistical cues and to cues resulting from the acoustic characteristics of the syllables. However, a number of factors may play a role in how cues are weighted at any given point in development. For example, to implement stress-based cues, infants must know something about the distribution of stressed syllables relative to word onsets and offsets in their native language, which may require a significant amount of language-specific knowledge. Infants may learn this mapping by using transitional probability as an initial segmentation strategy (Thiessen & Saffran, 2003; for a counterargument see the *early rhythmic segmentation hypothesis*: Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006; Nazzi, Jusczyk, & Johnson, 2000).

The adaptiveness of the given cue may also play a role in its weighting. In comparison with statistical cues to word boundaries (such as transitional probabilities) that require computations across a corpus, stress-based cues are computationally simpler; infants only need to hear a single presentation of a word's syllables to decide their relative levels of stress. Grouping sounds rhythmically in response to intensity and duration variation may require a longer exposure for the perceived rhythm to develop, because rhythm is a repeated pattern that necessarily unfolds over time. In fact, adult listeners subjectively report that they can flip their rhythmic grouping of a series of intensity-varying or duration-varying sounds at any given moment, even though their overall percept is consistent with the Iambic/Trochaic Law (Hay & Diehl, 2007). Thus, once learned, stress cues may provide a more robust cue to word boundaries than either statistical cues or individual rhythmic grouping cues.

To investigate whether syllable intensity and duration each function independently as cues to initial prominence (trochaic-based hypothesis), or instead lead to opposing rhythmic groupings as laid out in the Iambic/Trochaic Law (ITL-based hypothesis), the current study employed a paradigm in which we know that infants demonstrate sensitivity to linguistic stress as a cue to word boundaries. We adapted the artificial language used by Thiessen and Saffran (2003) to create four different speech streams by manipulating the acoustic structure of the syllables. The languages were created by concatenating four disyllabic nonsense words, such that the transitional probability within words was high (TP = 1.0), and the transitional probability at word boundaries was low (TP = .2–.4). Following a brief familiarization, infants were tested on their listening preference for "words" (high transitional probability) and "part-words" (low transitional probability) from the language. Two of the languages had a trochaic structure; we independently manipulated syllable intensity or duration so that the *first* syllable of each disyllabic word was either more intense or longer than the second syllable (initial prominence). The other two languages had an iambic structure; the *second* syllable of each disyllabic word was more intense or longer than the first syllable (final prominence). If 9-month-old infants treat both intensity and duration as cues to word-initial stress, then the statistics and acoustic cues should direct infants to the same word-onset boundary locations in the two Trochaic languages and to different word boundaries in the two Iambic languages. In both cases, we would expect infants to segment trochees from the artificial languages and thus discriminate words from part-words. However, if 9-month-old infants have perceptual grouping biases that are consistent with the principles outlined in the Iambic/Trochaic Law, then intensity and duration should elicit opposite rhythmic grouping patterns, with greater intensity signaling trochaic grouping (consistent with the intensity principle) and greater duration signaling iambic grouping (consistent with the duration principle). On this view, infants group syllables into *trochees* when intensity is varied and group syllables into *iambs* when duration is varied. This latter result would be particularly interesting given that 9-month-old English-learning infants typically prefer trochaic groupings when tested with materials containing multiple correlated cues indicating the placement of lexical stress.

EXPERIMENT 1

In Experiment 1, 9-month-old infants were familiarized with one of four artificial languages which contained two potential cues to word boundaries: sequential statistics (transitional probabilities) and an acoustic cue (either syllable intensity or duration). Infants were then tested on their listening preference for "words" (high transitional probability) and "part-words" (low transitional probability) from the familiarization language. In the Trochaic languages, the first syllable of each word was either more intense (Intensity Trochaic language) or longer (Duration Trochaic language) than the second syllable. In the Iambic languages, the second syllable of each word was either more intense (Intensity Iambic language) or longer (Duration Iambic language) than the first syllable.

If 9-month-old English-learners treat intensity and duration as independent cues to word-initial stress, they should exhibit the same pattern of segmentation regardless of whether the cue is intensity or duration. On this trochaic-based view, when the acoustic cue marks the first syllable (Intensity Trochaic and Duration Trochaic languages), infants should correctly segment words from fluent speech, because statistical cues and word-initial stress cues direct infants to the *same* word boundaries. Similarly, when the acoustic cue marks the second syllable (Intensity Iambic and Duration Iambic languages), infants should incorrectly segment part-words from fluent speech, because the statistical cues and word-final stress cues direct infants to *different* word boundaries. However, if increased intensity signals word beginnings and increased duration signals word endings, as predicted by the Iambic/Trochaic Law, infants should perceive a trochaic rhythm when syllables alternate in intensity and an iambic rhythm when syllables alternate in duration – regardless of whether the acoustic cue marks the first or the second syllable. According to this ITL-based view, infants should successfully extract words when the statistical and rhythmic cues direct them to the *same* word boundaries, that is, when the *first* syllable is more intense (Intensity Trochaic language) or when the *second* syllable is longer (Duration Iambic language). However, we would expect a different pattern of results in the conditions where the statistical cues and rhythmic cues direct infants to *different* word boundaries, that is, when the *second* syllable is more intense (Intensity Iambic language) or the *first* syllable is longer (Duration Trochaic language). In these cases, we would expect the rhythmic grouping to interfere with infants' ability to correctly exploit the sequential statistical cues to word boundaries. On this view, infants may be sensitive to both rhythmic and statistical cues to word boundaries, with neither cue sufficiently strong to override the other one. In this case, we would expect rhythmic cues to limit infants' abilities to track sequential statistics in this segmentation task, resulting in no significant preferences for words versus part-words. Finally, if infants ignore the acoustic cues altogether, then they should extract statistical words in all conditions irrespective of whether those words are trochees or iambs, or whether the syllables are marked by intensity or duration.

Method

Participants

Four groups of 22 healthy, full-term, 9-month-old infants participated in this experiment (mean age: 9 months, range: 8.1–9.5 months). An additional 63 infants were tested but were not included in the analysis for the following reasons: fussiness or crying (54); missing more than one trial of the same test item (5); and equipment failure (4). All infants came from monolingual English-speaking homes, were free of ear infections at the time of testing, and had no history of hearing problems or of pervasive developmental delays, according to parental report.

Stimulus materials

We manipulated the artificial language used by Thiessen and Saffran (2003) to create four different speech streams. All of the syllables were synthetically produced and fully co-articulated. Each language contained four disyllabic statistical "words," with a transitional probability of 1.0 between the two syllables. Each language also contained "part-words," which spanned word boundaries; these items had transitional probabilities ranging from .2 to .4. The language was frequency balanced, such that the words and part-words used during the test occurred with equal frequency during familiarization (e.g., Aslin et al., 1998b). This was accomplished by manipulating word frequency: two of the words were high frequency (*diti* and *bugo*), occurring 114 times each, and two of the words were low frequency (*dapu* and *dobi*), occurring 57 times each. The two low-frequency words and the two part-words created from the high-frequency words (*godi* and *tibu*) served as test stimuli; both types of test items had the same overall frequency in the corpus. The use of frequency-balanced test stimuli ensures that differences in looking times are not a result of the absolute frequency with which syllables co-occur, but instead are because of the probability structure of the artificial language.

The four languages were created by independently manipulating syllable intensity and duration, while preserving the statistical structure described previously. Two of the languages had a trochaic structure; we independently manipulated syllable intensity or duration so that the *first* syllable of each disyllabic word was either louder (Intensity Trochaic language) or longer (Duration Trochaic language) than the second syllable. Thus, according to a stress-based account, words in both of these languages have initial prominence. The other two languages had an iambic structure; the *second* syllable of each disyllabic word was either louder (Intensity Iambic language) or longer (Duration Iambic language) than the first syllable. According to a trochaic-based account, part-words in both of these languages should be perceived as having initial prominence. An ITL-based account makes a very different prediction. According to the Trochaic/Iambic Law, intensity should signal initial prominence and duration should signal final prominence regardless of whether the languages have trochaic or iambic structure.

In the two languages where syllable intensity was manipulated (Intensity Trochaic and Intensity Iambic), syllable duration was set at a constant length of 260 ms, while syllable intensity alternated between 65 and 56 dB SPL. In the two languages where syllable duration was manipulated

(Duration Trochaic and Duration Iambic), syllable intensity was set at a constant volume of 60.5 dB SPL, while syllable duration alternated between 173 and 346 ms. The values for intensity and duration were selected based on the ratios found in stressed versus unstressed syllables in English bisyllabic nonsense words produced with either initial or final stress (Schwartz, Petinou, Goffman, Lazowski, & Cartusciello, 1996). In addition, English-speaking adults show the same proportion of trochaic grouping in response to a 9 dB difference in intensity as they do iambic grouping in response to a doubling of duration (Hay & Diehl, 2007), suggesting that these values are of parallel strength (at least for English-speaking adults). Crucially, the test stimuli did not contain any stress or rhythmic information and were identical in all conditions: all syllables had equal loudness (60.5 dB SPL) and duration (300 ms).[3] This ensures that testing was comparable across all four conditions.

In statistical learning studies using artificial materials, infants often exhibit a novelty preference for nonwords/part-words (e.g., Saffran et al., 1996). However, Thiessen and Saffran (2003, 2004) observed a familiarity preference for words. For ease of interpretation of our results across conditions, we increased the duration of the exposure to attempt to elicit a novelty preference. Thiessen and Saffran (2003, 2004) presented high- and low-frequency words 90 and 45 times, respectively; we increased the duration of familiarization to include 114 and 57 presentations of high- and low-frequency words, respectively. Each speech stream lasted for a total of 3 min.

Apparatus and procedure

Infants were tested individually in a sound-attenuated booth that was equipped with three flashing lights: a red center light and two yellow side lights. During the experiment, infants were seated on their parent's lap in the center of the booth, approximately 1 m from each of the lights. The study included three phases: familiarization, re-familiarization, and test. During familiarization, infants listened to one of the artificial languages via laterally placed speakers. While the language played continuously, infants watched an age-appropriate silent video clip. Following familiarization, the researcher entered the booth, covered up the video monitor, and placed headphones on the parent. Testing began with a brief re-familiarization. During this phase, infants listened to an additional 28.8 sec of the same language presented over laterally placed speakers. Although the language was

[3]Syllable duration was increased slightly in the test syllables as compared to the familiarization languages to approximate syllable duration increases found when words are spoken in citation form.

played continuously, lights flashed contingent on the infant's looking behavior. This re-familiarization period, in which the overall statistics of the language were maintained, was included to give infants an opportunity to become familiar with the lights. Directly following re-familiarization, the test phase began. Each test trial began with a center flashing light. Once the infant fixated on the center light, a button pressed by an external observer signaled the center light to extinguish and one of the side lights to start flashing. When the infant looked in the direction of the flashing side light, a single test stimulus repeated continuously until the infant looked away for 2 sec or if 20 sec had elapsed, whichever came first. The two low-frequency words, *dapu* and *dobi*, and two part-words, *godi* and *tibu*, served as test stimuli. Each test word was repeated on three trials, in blocked random order, for a total of 12 test trials. Infants' looking times to the test stimuli were recorded by the same PC computer that controlled the stimulus presentation, using custom-designed software.

Results and discussion

To determine whether our acoustic manipulation affected discrimination of words versus part-words, we first performed a 2 (Language Type; trochaic versus iambic) X 2 (Cue Type; intensity versus duration) analysis of variance, with looking time difference scores (looking time to words minus looking time to part-words) as our dependent measure. There was a significant interaction between Language Type and Cue Type [$F(1, 84) = 5.6, p = .02, \eta^2 = .05$], suggesting that the placement of intensity and duration cues affected subsequent discrimination (see Figure 1). None of the main effects were significant (all p values $> .10$).

Planned comparisons revealed that infants looked longer during the presentation of part-words than words in both the Intensity Trochaic [$t(21) = 5.6$, $p < .001$, $d = .99$] and Duration Iambic [$t(21) = 2.7$, $p = .014$, $d = .58$] familiarization conditions [all t-tests, 2-tailed; effect sizes reported for t-tests are Cohen's d[4] (Cohen, 1988)]. By increasing infants' familiarization with the artificial language used by Thiessen and Saffran (2003), we were able to elicit a novelty instead of a familiarity preference. Eighteen and 17 of 22 infants exhibited this pattern for the Intensity Trochaic and Duration Iambic conditions, respectively. Infants successfully discriminated words from part-words when the first syllable of each word was marked by greater intensity, listening longer to the relatively novel part-words over more familiar words. Contrary to a trochaic-based hypothesis,

[4]Cohen's d (Cohen, 1988) was calculated using an average standard deviation that corrects for the dependence between the means for paired-samples tests (Morris & DeShon, 2002).

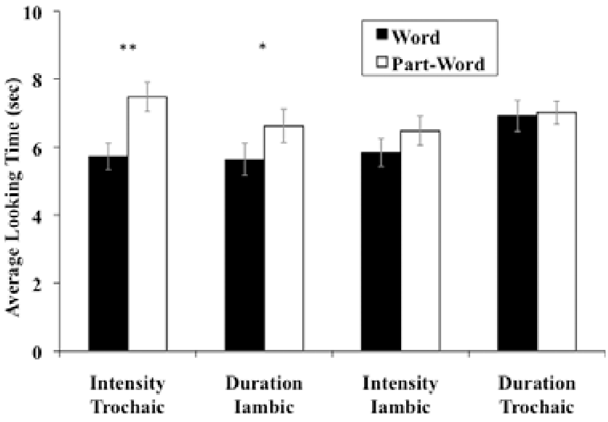

Figure 1 Results of Experiment 1: Average looking times ($\pm$1 SE) to words and part-words in each of the four familiarization conditions (9-month-old infants). $^{**}p < .001$, $^{*}p < .05$.

infants also listened longer to the part-words when the second syllable of each word was marked by longer duration. Interestingly, there were no significant differences in looking time during the presentation of words and part-words in either the Intensity Iambic [$t(21) = 1.4$, $p = .18$, $d = .29$] or Duration Trochaic [$t(21) < 1$, $p = .84$, $d = .04$] familiarization conditions (see Figure 1). Fourteen and 11 of 22 infants looked longer to part-words than words in the Intensity Iambic and Duration Trochaic conditions, respectively.

These results suggest that acoustic cues (i.e., syllable intensity and duration) interact with sequential statistics for 9-month-old learners. However, intensity and duration do not appear to function as independent cues to word-initial stress, despite infants' documented preference for trochees by this age (Jusczyk et al., 1993; Turk et al., 1995). Instead, consistent with an ITL-based account, the results from Experiment 1 may reflect perceptual rhythmic grouping biases as predicted by the Iambic/Trochaic Law – increased intensity provides a cue to word onsets and greater duration provides a cue to word offsets. At 9 months of age, infants may be sensitive to both statistical and rhythmic grouping cues to word boundaries, with neither cue receiving sufficient weight to override the other (see General Discussion).

These results provide indirect support for the principles outlined in the Iambic/Trochaic Law, proposed to account for the minimal set of rhythmic units in the world's languages (Hayes, 1995), but may, more generally, reflect universal grouping biases. However, cross-linguistic (Hay & Diehl, 2007; Iversen et al., 2008) and developmental work (Bion et al., 2011; Trainor & Adams, 2000; Yoshida et al., 2010) has resulted in conflicting support for the principles outlined in the Law. One approach used to test the universality of a perceptual phenomenon is to look at differences between the perceptual processing of speech and nonspeech stimuli (e.g., Hay & Diehl, 2007). Thus, it is of interest to investigate whether listeners respond similarly to speech and nonspeech sounds in the current paradigm. To this end, Experiment 2 was designed as a conceptual replication of Experiment 1, using a nonspeech segmentation task. In this study, we tested the ITL-based hypothesis that by 9 months of age, infants show intensity- and duration-based rhythmic grouping of *nonspeech* sounds, consistent with the Iambic⁄Trochaic Law.

EXPERIMENT 2

In Experiment 2, we tested the universality of the Iambic⁄Trochaic Law by addressing the linguistic specificity of infants' rhythmic grouping abilities. We created four continuous tone streams with the same sequential statistical and acoustic cues found in Experiment 1. Previous research has demonstrated that both infants and adults can use transitional probability cues to segment tone streams (Saffran, Johnson, Aslin, & Newport, 1999). As in Experiment 1, in the Trochaic tone streams, the first tone of each tone "word" was either more intense (Intensity Trochaic tone stream) or longer (Duration Trochaic tone stream) than the second tone. In the Iambic tone streams, the second tone of each tone "word" was either more intense (Intensity Iambic tone stream) or longer (Duration Iambic tone stream) than the first tone. Following familiarization, infants were tested on their discrimination of tone "words" (tone sequences with high transitional probability) versus tone "part-words" (low transitional probability tone sequences). If the rhythmic grouping biases unfold similarly across auditory domains, we would expect 9-month-old infants to show a similar pattern of results independent of the type of input (i.e., speech versus nonspeech).

Method

Participants

Four groups of 22 healthy, full-term, 9-month-old infants participated (mean age: 9.0 months, range: 8.5–9.6 months). An additional 49 infants were tested but were not included in the analysis for the following reasons: fussiness or crying (37); parental interference (4); equipment failure (1);

missing more than one trial of the same test item (1); and looking times averaging < 3 sec to one or both sides (6). All other participant characteristics were identical to Experiment 1.

Stimulus materials

By replacing the eight syllables used in Experiment 1 with 8 distinct pure (sine wave) tones, we created a stream of tones analogous to the speech streams used in the prior experiments. Pure tones were created and acoustically manipulated using Wavax, a custom-designed waveform generating and editing program. The statistical structure of these nonspeech stimuli was identical to those used in Experiment 1. The two high-frequency words from Experiment 1 (*diti* and *bugo*) were replaced with two high-frequency tone words: A4C4 (i.e., pure tones at 440 and 261 Hz) and F4C5 (i.e., pure tones at 349 and 523 Hz). The two low-frequency words (*dapu* and *dobi*) were replaced with two low-frequency tone words: B4G4 (i.e., pure tones at 493 and 391 Hz) and D4E4 (i.e., pure tones at 293 and 329 Hz). As in the previous experiments, the two high-frequency tone words (A4C4 and F4C5) were presented 114 times and the two low-frequency tone words (B4G4 and D4E4) were presented 57 times each. The high- and low-frequency tone words were concatenated together to create a continuous tone stream. Two tone part-words (C4F4 and C5A4), which resulted from the juncture of the two high-frequency tone words, and the two low-frequency tone words (B4G4 and D4E4) occurred equally often in the tone stream and served as test stimuli.

Four tone streams were created by independently manipulating tone duration and intensity, while preserving the statistical structure of the languages from Experiment 1. The tone streams were presented at a slightly slower rate than the speech streams from the previous experiments to more closely resemble the rate used in previous tone segmentation studies (Saffran et al., 1999). Thus, in the two streams where tone intensity was manipulated (Intensity Trochaic and Intensity Iambic), tone duration was set at a constant length of 300 ms, while syllable intensity alternated between 65 and 56 dB SPL. In the two streams where tone duration was manipulated (Duration Trochaic and Duration Iambic), syllable intensity was set at a constant volume of 60.5 dB SPL, while syllable duration alternated between 200 and 400 ms. While all the tones were drawn from the key of C major, the tone stream did not follow the conventions of Western tonal music. Crucially, the test stimuli did not contain any rhythmic information and were identical in all conditions: all tones had equal loudness (60.5 dB SPL) and duration (300 ms). This ensures that testing was comparable across all four conditions.

Apparatus and procedure

The apparatus and procedure were identical to Experiment 1.

Results and discussion

To determine whether our acoustic manipulations affected discrimination of tone "words" versus tone "part-words," we first performed a 2 (Stream Type; trochaic versus iambic) X 2 (Cue Type; intensity versus duration) analysis of variance, with looking time difference scores (looking time to words minus looking time to part-words) as our dependent measure. There was a significant interaction between Language Type and Cue Type [$F(1,84) = 5.80$, $p = .02$, $\eta^2 = .06$], suggesting that the placement of intensity and duration cues affected subsequent discrimination (see Figure 2). None of the main effects were significant (all p values $> .35$).

Planned comparisons revealed that infants looked longer during the presentation of tone part-words than tone words in the Intensity Trochaic tone stream [$t(21) = 2.2$, $p = .039$, $d = .47$]. There was also a marginally significant effect in the Duration Iambic tone stream [$t(21) = 2.0$, $p = .057$, $d = .43$] in the direction of the predicted novelty preference. Fifteen and 13 of 22 infants exhibited this pattern for the Intensity Trochaic and Duration Iambic conditions, respectively. There were no significant differences in looking time during the presentation of tone words and tone part-words in either the Intensity Iambic [$t(21) < 1$, $p = .62$, $d = .11$] or Duration Trochaic

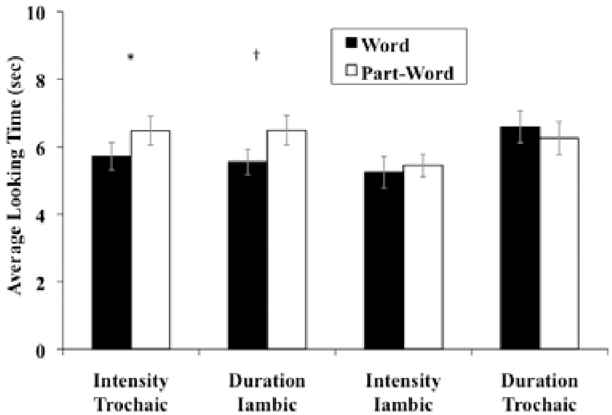

Figure 2 Results of Experiment 2: Average looking times (± 1 SE) to tone words and tone part-words in each of the four familiarization conditions (9-month-old infants). * $p < .05$, †$p = .057$.

[$t(21) = 1.1$, $p = .30$, $d = .23$] familiarization conditions (see Figure 2). Ten and 9 of 22 infants looked longer to tone part-words than tone words in the Intensity Iambic and Duration Trochaic conditions, respectively. This is the same pattern of results observed in Experiment 1.

Thus, even in nonspeech stimuli, acoustic cues (i.e., intensity and duration) interact with sequential statistics for 9-month-old learners. Like Experiment 1, these results are consistent with an ITL-based account. However, the marginal effects in the Duration Iambic condition suggest that variation in duration might not affect rhythmic grouping in nonspeech as strongly as in speech. This conclusion must be interpreted with some caution as the duration parameters of the tones were slightly modified from the linguistic version. Additionally, the linguistic stimuli used in Experiment 1 were comprised of a more complex alternation of consonants (C), vowels (V), and CV combinations, whereas the nonlinguistic stimuli had simple tones that alternated, potentially impacting rhythmic grouping.

Coupled with Experiment 1, these results raise an interesting question about the origins of rhythmic grouping biases. These findings provide indirect support for the hypothesis that the Iambic/Trochaic Law may reflect universal grouping principles (Hay & Diehl, 2007; Hayes, 1995). However, by the time they are 9 months old, infants' perceptual systems are already partially tuned to their native language (e.g., Friederici & Wessels, 1993; Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Jusczyk et al., 1993, 1999a; Mattys & Jusczyk, 2001; Morgan, 1996; Skoruppa et al., 2009; Thiessen & Saffran, 2003; Werker & Tees, 1984). Thus, grouping biases that appear to reflect universal principles may in fact develop through experience with native-language regularities. Another way to test the universality of a perceptual phenomenon is to examine younger infants, who typically give less weight to language-specific cues in perceptual tasks. Thus, in Experiment 3, we replicated Experiment 1 with younger infants, to see whether they also show segmentation patterns that are consistent with the principles set out in the Iambic/Trochaic Law.

EXPERIMENT 3

While nine-month-old infants already have access to native-language cues to word (and phrase) boundaries, younger infants tend to rely on more language-general cues. For example, although English-learning 6.5-month-old infants are aware of stress cues, they are not biased toward their native language's use of those cues (e.g., Morgan & Saffran, 1995; though see Höhle et al., 2009 for evidence from other languages). Similarly, younger infants tend to weight sequential statistics over stress cues when the two are placed in conflict (Thiessen & Saffran, 2003). Thus, if the principles set out in the Iambic/Trochaic Law are universal, they may provide infants with access to rhythmic grouping cues before they have learned relevant language-specific regularities. In Experiment 3, we explored the universality of the grouping principles set out in the Iambic/Trochaic Law by asking whether 6.5-month-old infants accept increased syllable intensity as a cue to word onsets and increased syllable duration as a cue to word offsets in a statistical segmentation task. Infants were familiarized with the same artificial languages used in Experiment 1. If these younger infants have grouping biases consistent with the principles of the Iambic/Trochaic Law, then they should readily discover words in the continuous speech stream when the first syllable of each word is more intense or when the second syllable of each word is longer, but not when the first syllable is longer or the second syllable is more intense.

Method

Participants

Four groups of 22 healthy, full-term, 6.5-month-old infants participated in this experiment (mean age: 6.6 months, range: 6.1–7.0 months). An additional 29 infants were tested but were not included in the analysis for the following reasons: fussiness or crying (20); parental interference (2); not noticing lights (3); and looking times averaging < 3 sec to one or both sides (4). All other participant characteristics were identical to Experiment 1.

Stimulus materials and procedure

The stimulus materials and procedure were identical to those used in Experiment 1.

Results and discussion

To determine whether our acoustic manipulations affected discrimination of words versus part-words, we first performed a 2 (Language Type; trochaic versus iambic) X 2 (Cue Types; intensity versus duration) analysis of variance, with looking time difference scores (looking time to words minus looking time to part-words) as our dependent measure. There were no significant interactions or main effects (all p values $> .10$; see Figure 3).

Although the omnibus ANOVA did not yield any significant interactions, planned individual *t*-tests revealed that these younger infants looked longer during the presentation of part-words than words in the Intensity Trochaic

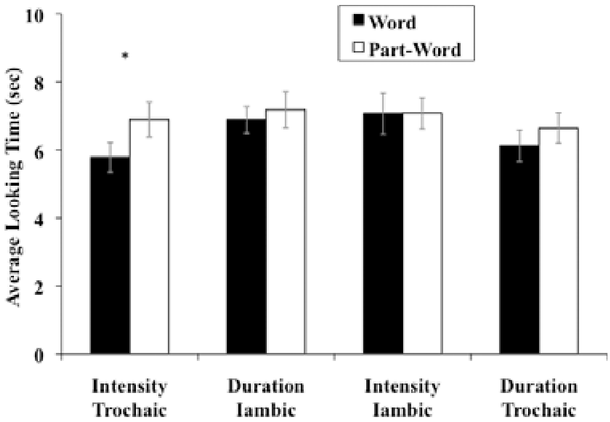

Figure 3 Results of Experiment 3: Average looking times (± 1 SE) to words and part-words in each of the four familiarization conditions (6.5-month-old infants). *$p < .05$.

familiarization condition [$t(21) = 2.8$, $p = .01$, $d = .61$].[5] However, they did not show any significant looking time differences between words and part-words in the Duration Iambic familiarization condition [$t(21) < 1$, $p = .46$, $d = .17$]. Seventeen of 22 infants looked longer during the presentation of part-words than during the presentation of words in the Intensity Trochaic familiarization condition, while only 11 of 22 infants exhibited this pattern following familiarization with the Duration Iambic language. Thus, infants seemingly successfully discriminated words from part-words when the first syllable of each word was marked by greater intensity, but not when the second syllable of each word was marked by a longer duration. These younger infants also failed to discriminate words from part-words in the Duration Trochaic [$t(21) = 1.3$, $p = .18$, $d = .30$] and Intensity Iambic familiarization conditions [$t(21) < 1$, $p = .99$, $d = .004$].

These findings suggest that intensity-varying syllables may indeed signal initial prominence for 6.5-month-old infants, as predicted by the ITL-based account, and as supported by the Iambic/Trochaic Law's intensity principle (e.g., Hay & Diehl, 2007; Iversen et al., 2008). When sequential statistics and syllable intensity directed infants to the same word boundary location, infants extracted the words from the continuous speech stream. However, contrary to the ITL-based account and to the Iambic/Trochaic Law's duration principle, 6.5-month-old infants did not treat increased duration as a rhythmic cue to word offsets. These findings are consistent with two recent developmental studies that suggest that young infants do not group

[5]This t-test remains significant with a Bonferroni correction for multiple comparisons.

duration-varying sounds into short-long (iambic) units (Bion et al., 2011; Yoshida et al., 2010). Interestingly, variation in duration did limit 6.5-month-old infants' use of transitional probability as a cue to word boundaries. Had these younger infants simply ignored syllable duration, they should have successfully segmented words according to their sequential statistics, as in Thiessen and Saffran (2003). This pattern of results suggests that these younger infants are either (i) unable to use duration variation to group sounds iambically, in contradiction of the second tenet of the Iambic/Trochaic Law, or (ii) are unable to use this particular rhythmic cue in a consistent manner (see General Discussion for additional explanations).

Taken together, the results from Experiments 1 and 3 suggest that there is a shift in infants' perceptual rhythmic grouping capabilities (or in the relative weighting of rhythmic grouping and statistical cues) from 6.5- to 9-months of age, possibly as a result of linguistic experience. These results suggest that the tenets of the Iambic/Trochaic Law, which is claimed to reflect universal grouping principles, may need to be revised. The data are consistent with the intensity principle, namely that elements contrasting in intensity naturally form groupings with initial prominence; this perceptual heuristic may represent a universal operating characteristic of the human auditory system (see General Discussion for an alternate hypothesis). Conversely, the duration principle, which purports that elements contrasting in duration naturally form groupings with final prominence, may emerge through language experience.

GENERAL DISCUSSION

The current set of studies investigated how the acoustic characteristics of sounds interact with sequential statistical cues to word/unit boundaries in a segmentation task. In Experiment 1, we tested the hypothesis that 9-month-old infants would treat two acoustic correlates of linguistic stress, namely syllable intensity and duration, as independent cues to word-initial prominence. Consistent with this trochaic-based hypothesis, infants extracted trochees when the words' first syllables were more intense than their second syllables as evidenced by their successful discrimination of words from part-words. Sequential statistical cues to word boundaries were constrained when the second syllables were marked by greater intensity. In contradiction of the trochaic-based hypothesis, infants failed to extract trochees when the first syllables were marked by greater duration. Instead, they extracted *iambs* when the words' second syllables were longer. Thus, the results from Experiment 1 instead provide indirect support for rhythmic grouping biases outlined in the Iambic/Trochaic Law, which predicts that intensity and duration have characteristic and opposite effects on rhythmic grouping; intensity variation leads to rhythmic grouping with initial prominence and duration variation leads to rhythmic grouping with final prominence.

In Experiments 2 and 3, we investigated the universality of rhythmic grouping biases. In Experiment 2, we examined the linguistic specificity of infants' rhythmic grouping abilities by testing the hypothesis that for 9-month-old infants, intensity and duration interact with sequential statistics in a nonspeech segmentation task in a manner consistent with the principles of the Iambic/Trochaic Law. Our hypothesis was confirmed, suggesting that speech and nonspeech are processed in a similar manner, although a developmental examination of rhythmic grouping of nonspeech using this experimental paradigm would be needed to confirm this conclusion (see Yoshida et al., 2010 for supporting evidence). Further support for the domain generality of rhythmic grouping principles set out in the Iambic/Trochaic Law comes from recent work by Peña, Bion, and Nespor (2011) demonstrating ITL-based rhythmic grouping of visual sequences. Our findings are particularly interesting in light of the ongoing debate as to whether perceptual processing is domain general or domain specific. Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) and Liberman and Matthingly (1985) have long argued that speech and nonspeech are processed independently, and their Motor Theory of Speech Development has received a great deal of attention (see Galantucci, Fowler, & Turvey, 2006 for a review). Additionally, a number of recent studies have suggested differential processing of speech and nonspeech stimuli in infancy, including differential rule-learning abilities (Marcus, Fernandes, & Johnson, 2007) and category formation (Fulkerson & Waxman, 2007) in the two domains, and preferences for linguistic over nonlinguistic sounds (Vouloumanos & Werker, 2007). There is, however, a growing body of literature that suggests that perceptual restructuring as a result of linguistic experience has domain-general effects (e.g., Bent, Bradlow, & Wright, 2006; Hay & Diehl, 2011; Iversen et al., 2008; Yoshida et al., 2010). For example, adult Mandarin listeners' impressive ability to identify language-relevant pitch contours impairs their ability to identify nonlinguistic pitch contours (Bent et al., 2006). Similarly, native-language experience with stop consonants (e.g., /p/ versus /b/) affects English- and Spanish-speaking adults' discrimination of nonspeech analogues of stop consonants (Hay & Diehl, 2011). And experience with the rhythmic structure of one's native language impacts nonspeech rhythmic perception (Iversen et al., 2008; Yoshida et al., 2010). Linguistic experience can affect the perception of nonspeech sounds that share certain features with speech sounds, suggesting the presence of intricate connections between auditory domains.

In Experiment 3, we took a developmental approach to the question of universality by replicating Experiment 1 with younger infants. We asked whether the use of rhythmic cues reflects a predisposition for rhythmic grouping or whether the use of rhythmic cues instead reflects the development of language-specific perceptual abilities. This type of perceptual restructuring as a result of linguistic experience is pervasive during language acquisition. For example, from birth, infants show a remarkable ability to discriminate many, if not most, of the sounds found in the world's languages (for a review, see Aslin, Jusczyk, & Pisoni, 1998a). During the first year of postnatal life, these perceptual abilities become more language specific and by about 10 months of age, infants show decreased responsiveness to some previously discriminable nonnative phoneme contrasts, while at the same time show increases in native-language phoneme discrimination (Werker & Tees, 1984). A similar trajectory is seen in the speech segmentation literature: for example, by 9 months of age, but generally not before, infants can use native-language prosodic patterns (Jusczyk et al., 1993; Morgan, 1996; Thiessen & Saffran, 2003), phonotactics (Friederici & Wessels, 1993; Mattys & Jusczyk, 2001), and allophonic variation (Jusczyk et al., 1999a) to locate lexical items in fluent speech. Thus, infants' perceptual capacities change significantly over the first year of life, primarily in response to linguistic experience. Experiment 3 addressed the following question: does infants' use of rhythmic cues reflect a predisposition for rhythmic grouping or instead the development of language-specific perceptual abilities? Results from Experiments 1 and 3 are consistent with the hypothesis that auditory experience affects the rhythmic grouping of duration-varying sequences; 9-month-old infants, but not 6.5-month-olds, were able to find iambs when the second syllable of each word was marked by increased duration. However, both younger and older infants extracted trochees when the first syllable of each word was marked by an increased intensity. Both age groups failed to demonstrate segmentation when the second syllable of each word was more intense or when the first syllable of each word was longer, and thus, no developmental change was observed in these conditions.

Taken together, Experiments 1–3 demonstrate that, consistent with the intensity principle of the Iambic/Trochaic Law, variations in intensity appear to elicit rhythmic groupings with initial prominence in infancy. Coupled with the related adult literature (Hay & Diehl, 2007; Iversen et al., 2008), our results suggest that infants may have a predisposition to group intensity-varying sounds trochaically. Conversely, and in contradiction to the Iambic/Trochaic Law's duration principle, 6.5-month-old infants do not appear to be consistently assigning longer syllables to word endings. The current findings do not allow us to rule out the possibility that greater differences in syllable durations may have lead to iambic grouping in our younger infants. A future study using larger duration differences may be a way of further exploring this issue. However, given the developmental changes observed, our results do suggest some sort of perceptual restructuring for duration-varying speech sounds, and/or a shift in the relative weighting of duration-based rhythmic grouping and statistical cues, and/or a shift in the level of representation that infants take into account when establishing perceptual groupings, between 6.5 and 9 months of age (see also Yoshida et al., 2010).

The precise mechanism by which duration-based rhythmic grouping develops is unclear. One proposal is that the rhythmic shape of words drives rhythmic grouping preferences (Jakobson, Fant, & Halle, 1952; Kusumoto & Moreton, 1997). In this account, placement of stressed syllables within words, or the relative syllable weight, may impact duration-based rhythmic grouping.

A second proposal is that the Iambic/Trochaic Law applies at the phrasal level instead of at the level of the word and thus the phrase is the most important unit affecting rhythmic grouping preferences (Iversen et al., 2008). Iversen et al. (2008) argue that function-content word order may drive perceptual grouping. In English and French, short function words (functors), including articles, prepositions, conjunctions, and pronouns, are overwhelmingly monosyllabic and acoustically reduced. These functors typically precede longer content words, which can be monosyllabic or multisyllabic and contain at least one fully articulated syllable. Thus, the majority of resulting phrases in English and French have a short-long pattern, and differences in rhythmic grouping would not be expected for listeners of these two languages. In Japanese, functors generally appear *after* related content words (Baker, 2001), resulting in a long-short pattern. Iversen et al. (2008) suggest that experience with native-language minimal phrasal units may drive developmental and cross-linguistic differences in rhythmic grouping.

Nespor and Vogel (1986), Nespor, Guasti, and Christophe (1996) and Nespor et al. (2008) provide an alternate phrasal-level account based on the head-complement structure found across languages. On this view, the physical realization of complement prominence within phrases varies depending on its location in a manner that tends to be consistent with the Iambic/Trochaic Law. Within a given language, if the primary prominence of a phrase is realized mainly through pitch and intensity, then that primary prominence is in a phonological phrase that is stress-initial and the syntactic heads (e.g., verbs, nouns) follow their complements (e.g., objects, nominal compliments). Correspondingly, when a language has phrases in which syntactic heads precede their complements then that language tends to have final prominence marked by duration. Although some languages, such as German, have both stress-initial (i.e., trochaic; object-verb) and stress-final (i.e., iambic; verb-object) phrases, many languages show a predominantly trochaic (e.g., Japanese and Turkish) or iambic (e.g., English and French) structure. Thus, experience with the acoustic correlates of the head-complement structure of the native language may drive the development of rhythmic grouping, such that English- and French-speaking listeners, but not Japanese, would develop iambic grouping biases for duration-varying sounds. Further, the grouping principles derived from the Iambic/Trochaic Law (and its extension to trochaic pitch-based rhythmic grouping) may help infants discover word order regularities (i.e., object-verb or verb-object) in their native language (Bion et al., 2011; Gervain & Werker, 2011).

A hybrid explanation may also account for the current data; as infants gain knowledge about the prosodic patterns in their native language, the level of representation used to establish perceptual groupings may shift. On this view, 6.5-month-old infants may be sensitive to the word-level stress patterns characteristic of their native language, and hence our English-learning infants demonstrate a trochaic bias (when processing intensity-varying stimuli). By 9 months of age, infants might be able to track rhythm both at the word and the phrase levels (with the latter being iambic in English, Nespor et al., 2008) and thus also demonstrate an iambic bias (when processing duration-varying stimuli). Thus, there may not be a perceptual restructuring or native-language attunement between 6.5 and 9 months of age, but rather an increase in knowledge about the prosodic patterns of their native language. Importantly, no matter which explanation is adopted to explain the full set of results obtained in this study, one needs to assume some form of the Iambic/Trochaic Law, as no grouping biases were induced when inappropriate cues were used (i.e., duration for trochees or intensity for iambs).

While our results do not differentiate between these hypotheses, they do suggest that duration-based rhythmic grouping is not a universal perceptual bias, as has previously been implied by the Iambic/Trochaic Law. The current results do, however, suggest that intensity-based rhythmic grouping may act as a perceptual bias that is either universal or comes online very early as a result of linguistic experience.

One final alternative explanation is that younger infants may have a more difficult time than older infants in processing multiple cues to word boundaries. On this account, our 6.5-month-old English-learning infants are successful in the Intensity Trochaic condition because the trochaic unit is more familiar to them given the prosodic patterns typical of their native language. Their failure in the Duration Iambic condition may have little to do with the nature of the cue (i.e., duration), but instead may reflect difficulty extracting the less familiar iambic units. Older infants may also have difficulty processing multiple cues when they are realized over less familiar auditory input, in this case tone sequences, and thus be less likely to extract iambic units. However, even in the absence of statistical cues, duration-based rhythmic grouping is typically not observed in younger infants (e.g., Bion et al., 2011; Yoshida et al., 2010).

An interesting puzzle emerges when comparing the discrimination patterns observed in the current study with the findings of Thiessen and Saffran (2003). In both studies, 9-month-old infants appear to be sensitive to both acoustic and sequential statistical cues during speech segmentation. However, our results suggest that individual rhythmic cues are not strong enough to trump sequential statistics (i.e., transitional probability), whereas Thiessen and Saffran (2003) found that stress cues override sequential statistics. Why do stress cues trump sequential statistics when individual rhythmic cues do not? The answer may be found in the acoustics. In Thiessen and Saffran's (2003) study, stressed syllables were marked by three acoustic cues: intensity, duration, and pitch. A small increase in a sound's duration tends to be perceived as an increase in the loudness of that sound (Nishinuma, di Cristo, & Espesser, 1983). Thus, small increases in duration in stressed syllables, coupled with increases in intensity, may provide a stronger cue to word onset than any of our manipulations. In addition, although the effects of pitch on rhythmic grouping are not predicted by the Iambic/Trochaic Law, variations in pitch also lead to trochaic rhythmic grouping (Bion et al., 2011; Nespor et al., 2008; Rice, 1992). Taken together, the intensity, duration, and pitch of the stressed syllables in Thiessen and Saffran (2003) provided three redundant cues to initial prominence, allowing the stress cues to trump sequential statistical cues in their study.

It should be noted, however, that providing multiple redundant cues to initial prominence is not a necessary condition for the acoustics to trump transitional probability in a segmentation task. Thiessen and Saffran (2004) found that spectral tilt (also a correlate of linguistic stress) overrides transitional probability cues for 9-month-old infants, when the two cues are placed in conflict. However, by 12 months of age, spectral tilt is a less effective cue to word boundaries. Sluijter and van Heuven (1996) have suggested that spectral tilt is the most powerful cue to linguistic stress, and thus young infants may treat spectral tilt as functionally equivalent to the three redundant cues used in Thiessen and Saffran (2003).

A second hypothesis is that segmenting words from fluent speech using stress is an adaptive process (Thiessen & Saffran, 2003). As discussed in the Introduction, infants only need to hear a single presentation of a word's syllables to decide their relative levels of stress. However, grouping sounds rhythmically in response to intensity and duration variation may require a longer exposure for the perceived rhythm to develop (Hay & Diehl, 2007). Thus, stress cues may provide a more robust cue to word boundaries than individual rhythmic grouping cues.

Another interesting puzzle emerges when comparing the discrimination patterns observed for 6.5-month-olds in Experiment 3 with the findings of Thiessen and Saffran (2003). In the current study, the younger infants failed to discriminate words from part-word in the both Duration Trochaic and Duration Iambic conditions. If infants simply ignored or gave very little weight to duration cues, we would have expected them to extract the statistical words in both conditions. Thiessen and Saffran (2003) found that infants ignore stress cues to word boundaries in favor of sequential statistical cues at this age. One possibility is that these young infants are getting conflicting information regarding the linguistic relevance of duration. While duration is used to mark initial stress at the word level in English, it marks final units at the phrase level (Iversen et al., 2008; Nespor & Vogel, 1986; Nespor et al., 1996, 2008). Young infants may be sensitive to both of these uses, and thus both might interfere with the sequential statistical cues to word boundaries. Over time, English-learning infants may discover that duration is a more robust cue to final prominence than initial prominence. Moreover, in order for duration to function as a cue to final prominence, younger infants may also crucially rely on a correlated decrease in syllable pitch and intensity – cues that were lacking in the present experimental conditions.

In sum, acoustic cues had a robust impact on learning during our segmentation tasks; infants were only able to successfully discriminate words from part-words (speech or nonspeech) when the rhythmic cues and sequential statistical cues were consistent, signaling the same groupings of syllables. When the two types of cues provided inconsistent information about word boundaries, segmentation was disrupted. Rhythmic grouping biases appear to impact perceptual learning by making sounds that cohere together rhythmically easier to segment than sounds that do not. These findings are broadly consistent with recent work by Shukla, White, and Aslin (2011), who demonstrate that statistically defined words are more easily mapped onto novel objects by 6-month-old infants when those words are aligned with, rather than straddle, prosodic boundaries. More generally, our results suggest that infants may be able to start learning about the world around them, not only by taking advantage of regularities in their perceptual environment, but also by taking advantage of perceptual biases that direct their computations. The infant's environment is notoriously complex; combining intrinsic and extrinsic limitations may be one piece of the solution that allows infants to learn so much so rapidly.

ACKNOWLEDGMENTS

This research was funded by grants from NICHD to J.F.H. (F32-HD557032) and to J.R.S. (R01HD37466), a grant from the James F. McDonnell Foundation to JRS and by a NICHD core grant to the Waisman Center (P30HD03352). We would like to thank the members of the Infant Learning Lab and especially Jessica Rich and Jessica Hersh for their assistance in the conduct of this research, the families who generously contributed their time, and Erik Thiessen and three anonymous reviewers for helpful comments on a previous version of this manuscript.

REFERENCES

Aslin, R. N., Jusczyk, P. W., & Pisoni, D. B. (1998a). Speech and auditory processing during infancy: Constraints on and precursors to language. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language* (Vol. 2, pp. 147–254). New York: Wiley.

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

Baker, M. C. (2001). *The atoms of language*. New York: Basic Books.

Beckman, M. E. (1986). *Stress and non-stress accent*. Netherlands Phonetic Archives 7. Dordrecht: Foris.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 97–103.

Bion, R. A. H., Benavides, S., & Nespor, M. (2011). Acoustic markers of prominence influence adults' and infants' memory of speech sequences. *Language & Speech*, *54*(1), 123–140.

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology*, *6*, 145–238.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*, 127–132.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition *Journal of the Acoustical Society of America*, *95*, 1570–1580.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd edn. Hillsdale, NJ: Lawrence Earlbaum Associates.

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

Cooper, G., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago, IL: The University of Chicago Press.

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., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology*, *14*(1), 113–121.

Delattre, P. (1965). *Comparing the phonetic features of English, German, Spanish and French*. Heidelberg: Julius Groos.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics in Language Teaching*, *4*, 183–198.

Dell, F. (1984). L'accentuation dans les phrases français. In F. Dell, D. Hirst & J.-R. Vergnaud (Eds.), *Forme sonore du langage: Structure des représentations en phonologie* (pp. 65–122). Paris: Hermann.

Dupoux, E., Pallier, C., Sebastián, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory Language*, *36*, 406–421.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress' deafness. *Journal of the Acoustical Society of America*, *110*, 1608–1618.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, *106*(2), 682–706.

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

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

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

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition*, *105*(1), 218–228.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377.

Gervain, J., & Werker, J. F. (2011). Learning two grammars through prosody: 7-month-old bilinguals' acquisition of the word order of their native language. Talk presented at the Society for Research in Child Development Biennial Meeting, Montreal, PQ (March 31).

Hay, J. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics*, *69*(1), 113–122.

Hay, J. F., & Diehl, R. L. (2011). *Linguistic experience has general auditory consequences: The role of experience with VOT in English and Spanish*. Manuscript submitted for publication.

Hayes, B. (1995). The rhythmic basis of the foot inventory. In B. Hayes (Ed.), *Metrical stress theory: Principle and case studies* (pp. 79–85). Chicago: The University of Chicago Press.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, *32*(3), 262–274.

Höhle, B., Unger, A., Gonzalez Gomez, N., & Nazzi, T. (2011). Rhythmical grouping in French and German adults: A cross-linguistic investigation of the Iambic-Trochaic Law. Paper presented at the Society for Research in Child Development Biennial Meeting, Montreal, PQ.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, *124*(4), 2263–2271.

Jakobson, R., Fant, C. G. M., & Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

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

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, *12*, 131–141.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*, 339–345.

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

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999a). Infant's sensitivity to allophonic cues for word segmentation. *Perception & 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.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971–995.

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., & Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Developmental Science*, *14*, 1100–1106.

Kusumoto, K., & Moreton, E. (1997). Native language determines parsing of nonlinguistic rhythmic stimuli. *Journal of the Acoustical Society of America*, *102*, 3204.

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

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

Liberman, A. M., & Matthingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Marcus, G., Fernandes, K., & Johnson, S. (2007). Infant rule-learning facilitated by speech. *Psychological Science*, *18*, 387–391.

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

Mersad, K., & Nazzi, T. (in press). When Mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning & Development*.

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 segmental and suprasegmental information in prelingual speech segmentation. *Child Development*, *66*, 911–936.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105–125.

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

Nazzi, T., Dilley, L. C., Jusczyk, A. M., Shattuck-Hufnagel, S., & Jusczyk, P. W. (2005). English-learning infants' segmentation of verbs from fluent speech. *Language and Speech*, *48*, 279–298.

Nazzi, T., Iakimova, I., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, *54*, 283–299.

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.

Nespor, M., Guasti, M. T., & Christophe, A. (1996). Selecting word order: The rhythmic activation principle. In U. Kleinhenz (Ed.), *Interfaces in phonology* (pp. 1–26). Berlin: Akademie Verlag.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio*, *2*, 1–29.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Berlin: Mouton de Gruyter, 1st edn. Dordrecht: Foris.

Nishinuma, Y., di Cristo, A., & Espesser, R. (1983). Loudness as a function of vowel duration in CV syllables. *Speech Communication*, *2*(2–3), 167–169.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Statistical learning in a natural language by 8 month-old infants. *Child Development*, *80*(3), 674–675.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*, 244–247.

Peña, M., Bion, R. A. H., & Nespor, M. (2011). How modality specific is the Iambic-Trochaic Law? Evidence from vision *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1199–1208.

Rice, C. C. (1992). Binary and ternarity in metrical theory: Parametric extensions. Diss., The University of Texas at Austin, Austin.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

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

Sahni, S. D., Seidenberg, M., & Saffran, J. R. (2010). Connecting cues: Overlapping regularities support cue discovery in infancy. *Child Development*, *81*, 727–736.

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 Acoustic Society of America*, *99*, 3192–3200.

Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, *11*(4), 596–606.

Shukla, M., White, K., & Aslin, R. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, *108*(15), 6038–6043.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., et al. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*(6), 914–919.

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.

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

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*, 86–132.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21.

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

Thiessen, E. D., & Saffran, J. R. (2004). Spectral tilt as a cue to word segmentation in infancy and adulthood. *Perception & Psychophysics*, *66*, 779–791.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning & Development*, *3*, 73–100.

Trainor, L. J., & Adams, B. (2000). Infants' and adults' use of duration and intensity cues in the segmentation of auditory patterns. *Perception & Psychophysics*, *62*, 333–340.

Tranel, B. (1987). *The sounds of French: An introduction*. Cambridge: Cambridge University Press.

Turk, A. E., Jusczyk, P. W., & Gerken, L. A. (1995). Do English-learning infants use syllable weight to determine stress? *Language and Speech*, *36*, 143–158.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*, 159–164.

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

Woodrow, H. (1909). A quantitative study of rhythm: The effect of variations in intensity, rate, and duration. *Archives of Psychology*, *14*, 1–66.

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224–1236). New York: Wiley.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition*, *115*(2), 356–361.