Journal of Experimental Psychology: Learning, Memory, and Cognition

How Modality Specific Is the Iambic–Trochaic Law? Evidence From Vision

Marcela Peña, Ricardo A. H. Bion, and Marina Nespor

Online First Publication, June 27, 2011. doi: 10.1037/a0023944

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/11/$12.00 DOI: 10.1037/a0023944

How Modality Specific Is the Iambic–Trochaic Law? Evidence From Vision

Marcela Peña
International School for Advanced Studies and Universidad Católica de Chile

Ricardo A. H. Bion
Stanford University

Marina Nespor
University of Milano-Bicocca

The iambic–trochaic law has been proposed to account for the grouping of auditory stimuli: Sequences of sounds that differ only in duration are grouped as iambs (i.e., the most prominent element marks the end of a sequence of sounds), and sequences that differ only in pitch or intensity are grouped as trochees (i.e., the most prominent element marks the beginning of a sequence). In 3 experiments, comprising a familiarization and a test phase, we investigated whether a similar grouping principle is also present in the visual modality. During familiarization, sequences of visual stimuli were repeatedly presented to participants, who were asked to memorize their order of presentation. In the test phase, participants were better at remembering fragments of the familiarization sequences that were consistent with the iambic–trochaic law. Thus, they were better at remembering fragments that had the element with longer duration in final position (iambs) and fragments that had the element with either higher temporal frequency or higher intensity in initial position (trochees), as compared with fragments that were inconsistent with the iambic–trochaic law or that never occurred during familiarization.

Keywords: iambic–trochaic law, speech, vision, memory, grouping

Supplemental materials: http://dx.doi.org/10.1037/a0023944.supp

The iambic–trochaic law (ITL) has been proposed to account for the grouping of sequences of sounds alternating in prominence into either iambs or trochees. Pairs of sounds are defined as iambs if their most prominent element is in final position, and they are defined as trochees if their most prominent element is in initial position instead. Studies spanning over a century (from Bolton, 1894, to Bion, Benavides-Varela, & Nespor, 2011) have demonstrated that sequences of sounds are grouped into iambic pairs when their elements alternate in duration (the longest and most prominent element is placed in final position), but they are grouped into trochaic pairs when their elements alternate only in pitch or intensity (the highest pitched or most intense element is placed in initial position). Initially proposed to explain the grouping of musical sequences (Bolton, 1894; Cooper & Meyer, 1960; Woodrow, 1951), the ITL has since been extended to account for regularities in speech production and biases in speech perception.

For language, the first formulations of the ITL focused on the effects of duration and intensity on the grouping of syllables at the word level. In speech production, the ITL was found to account for the location of word secondary stress across languages. Specifically, cross-linguistically, if secondary stresses are in final position in the phonological constituent known as the metrical foot, they are produced mainly with increased duration; if they are in initial position, they are manifested mainly with increased intensity (Hayes, 1995). In speech perception, a similar pattern holds: Adult participants report hearing iambic words when confronted with speech streams alternating in duration, and they report hearing trochaic words when confronted with speech streams alternating in intensity (Hay & Diehl, 2007).

The most recent extensions of the ITL focused on prominence at the phrasal level and on the role of pitch, in addition to duration and intensity, in grouping. In speech production, it has been proposed that the location of stress in phrasal constituents—on the first or on the last word of a phonological phrase—depends on whether the syntactic phrase is final or initial (Nespor & Vogel, 2008). It has also been proposed that the physical realization of phonological phrase stress varies depending on its location: Increased duration characterizes iambic patterns, whereas increased pitch and intensity characterize trochaic patterns (Nespor et al.,

Marcela Peña, Cognitive Neuroscience Sector, International School for Advanced Studies, Trieste, Italy, and Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago, Chile; Ricardo A. H. Bion, Department of Psychology, Stanford University; Marina Nespor, Department of Psychology, University of Milano-Bicocca, Milan, Italy.

The research presented was supported by the following grants: Fondecyt 1060767; Programa de Investigación Asociative–Comisión Nacional de Investigación Científica y Tecnológica, Project CIE-05; European Union NEST Pathfinder CALACEI (project reference: 012778); McDonnell Foundation grant. Marina Nespor had a fellowship (2006–2009) of the Interdisciplinary Center Beniamino Segre of the Lincei Academy that allowed her to be affiliated with the International School for Advanced Studies, Trieste.

Correspondence concerning this article should be addressed to Marina Nespor, Via Bonomea 265, 34136 Trieste, Italy. E-mail: marina.nespor@unimib.it

2008). Bion et al. (2011) investigated the role of the ITL in speech perception. They restricted their attention to pitch and duration, as their study was carried out on both adults and infants, and infants are sensitive to variations in pitch but are not as attentive to variations in intensity (Singh, White, & Morgan, 2008). In the adult experiments, they found that alternations of pitch and duration influence memory for speech sequences. In that work, perceptual grouping was tested in an indirect way, by focusing on whether linguistic units consistent with the ITL are better remembered than inconsistent units. Adult participants were better at remembering pairs of syllables that had been presented with a short syllable followed by a long one (iambs) and pairs in which a high-pitched syllable preceded a low-pitched one (trochees). A similar pattern of perceptual grouping in music has been found to be present already at 4.5 months of age (Jusczyk & Krumhansl, 1993; Krumhansl & Jusczyk, 1990). From this body of studies, it is possible to conclude that in auditory streams, the units having long duration mark the end of speech sequences, whereas units with higher pitch or intensity mark the beginning of speech sequences.

In a recent study, Nespor et al. (2008) found correlations between these regularities in speech production and the syntactic properties of different languages. Specifically, languages with final prominence within phonological phrases (iambic prominence marked mainly by duration) have syntactic heads that precede their complements, whereas languages with initial prominence within phonological phrases (trochaic prominence marked mainly by pitch and intensity) have syntactic heads that follow their complements. Head direction refers to the ordering of complements with respect to their head (e.g., objects with respect to either verbs or adpositions [prepositions and postpositions], and nominal complements with respect to nouns). In head-initial languages, like English and French, for example, verbs precede objects (e.g., *John loves India*), whereas in head-final languages, like Turkish and Japanese, verbs follow objects (e.g., the correspondent of *John India loves*). In addition, in head-initial languages, function words tend to precede content words, whereas in head-final languages, function words tend to follow them.

Interestingly, this correlation between the acoustic signal and syntax might help infants learn about the ordering of words in their language (e.g., Nespor, Guasti, & Christophe, 1996). This claim is supported by the fact that 6- to 12-week-old infants discriminate two languages that differ in head location exclusively on the basis of phonological phrase prominence (Christophe, Nespor, Guasti, & van Ooyen, 2003), and also by the fact that prelinguistic German infants rely on prosody to discriminate German phrases that differ only in head location (Bion, Höhle, & Schmitz, 2007). These findings give plausibility to the prosodic bootstrapping hypothesis and suggest that infants could potentially rely on prosodic patterns to learn some of the syntactic properties of their language (Gleitman & Wanner, 1982).

A central issue in the field of language acquisition concerns whether learning mechanisms are specific or general in nature. Several learning mechanisms proposed to be involved in the acquisition of language have been shown to be domain and modality general. For example, the ability to keep track of transitional probabilities between adjacent elements is used to segment both visual and auditory sequences (Fiser & Aslin, 2002; Saffran, Aslin, & Newport, 1996). Similarly, it has been shown that rule extraction mechanisms are recruited in both auditory (Marcus, Vijayan, Bandi Rao, & Vishton, 1999) and visual modalities (Saffran, Pollak, Seibel, & Shkolnik, 2007). In the present study, we investigate whether the ITL, so far proposed only for the auditory modality, is used to organize stimuli in the visual modality as well.

Language processing and acquisition are certainly enriched by the visual information provided by hands, body, face, and lips (McGurk & MacDonald, 1976; McNeill, 2005; Wright & Wareham, 2005). The analysis of the spatiotemporal properties of visual sequences is particularly relevant for sign languages (Corina & Knapp, 2008, for a recent review), as they involve a complex coordination and sequencing of nonvocal motor gestures (Poizner & Kegl, 1993). Yet little is known about the way in which different cues are used to group visual sequences.

Previous research, however, has suggested that humans can track a wide range of changes in duration, temporal frequency, and intensity in visual stimuli concatenated into sequences (Kelly & Burkeck, 1984; Spillmann, 2006). In the visual domain, duration refers to the time elapsed from the onset to the offset of a visual event; intensity can be estimated from the brightness of the visual stimuli (Lamb, 1991); and temporal frequency alludes to the rate at which a visual stimulus changes in a specific physical property (Hess & Snowden, 1992), such as its location in space. Adults are able to segment continuous visual sequences into distinct visual events exploiting information about their duration, temporal frequency, and intensity (Hard, Tversky, & Lang, 2006; Kurby & Zacks, 2008; Zacks, 2004; Zacks, Kumar, Abrams, & Mehta, 2009; Zacks & Swallow, 2007). Visual segmentation influences the way in which visual sequences are memorized and learned: Individuals who are better at segmenting are better at remembering (Swallow, Zacks, & Abrams, 2009). However, to the best of our knowledge, there are no previous reports about how visual sequences are grouped with respect to systematic changes in the nature of the cues to prominence—an aspect of grouping that has robustly been shown for auditory sequences.

In the present study, we focus on three visual cues, and on the ways in which they influence visual event segmentation, with the intent of drawing parallels between some properties of visual and auditory stimuli. Specifically, we test whether the most recent formulations of the ITL, which concern the grouping of phrasal chunks, also applies to the segmentation of visual sequences. Therefore, our goal is to evaluate the extent to which changes in duration, temporal frequency, and intensity of visually presented stimuli can influence perceptual grouping. Although the concept of duration in the auditory modality can be easily transferred to the visual modality, the visual analogues of auditory pitch and intensity are less obvious. In the auditory modality, the duration of a sound represents the amount of time the sound is audible. In the visual domain, duration can thus be operationalized as the amount of time a stimulus is visible. Similarly to auditory sequences, visual intensity reflects the stimulus energy, and may thus be manipulated by changing the brightness of the stimulus (Blakemore, Adler, & Pointon, 1993). In the auditory modality, pitch represents the perceived fundamental frequency of a sound over time. Accordingly, objects moving periodically at a higher temporal frequency (cycles per time unit) produce higher pitched sounds. For example, in speech, higher pitched sounds are produced by vibrating the vocal folds at a higher frequency (i.e., more vibrations per time unit); in a guitar, higher pitched chords vibrate at higher frequencies than lower pitched chords; in a tuning fork, higher pitched sounds are produced by forks that vibrate at higher frequencies. Therefore, we choose to define the perception of pitch in the visual modality as the processing of temporal frequency of visual stimuli. Temporal frequency refers to the number of times per time unit at which an image changes its physical properties, such as its spatial configuration. For example, movies at the cinema are played at the frame rate of 24 frames per second (a temporal frequency of 24 Hz). If the same movie was played at twice this frame rate, the temporal frequency would increase to 48 Hz, and objects would be perceived as moving faster. The inverse would happen if one decreased the number of frames per second.

This working definition of visual pitch is undoubtedly controversial, and we are open to other possible correlates such as hue and spatial frequency. However, temporal frequency seems to constitute a good candidate equivalent of "pitch" in vision, and it is the perceptual correlate that we decided to first investigate in this study. In the visual domain, changes in temporal frequency are perceived as changes in velocity (Reisbeck & Gegenfurtner, 1999; Thompson, 1982; Thompson & Stone, 1997), and a proposed correlate of pitch in sign language is the peak velocity of the sign's movement (Wilbur, 1999). We controlled temporal rather than spatial frequency of visual events as an auditory analogue of pitch, because spatial frequency refers to how often components of visual events occur per unit of space. As speech evolves in time, in absence of previous data, in the present study, we decided to manipulate temporal rather than spatial frequency.

On the basis of these working definitions, in three experiments, we investigated whether changes in either duration, temporal frequency, or intensity are exploited in the segmentation and grouping of visual events in a similar way to that in which changes in duration, pitch, or intensity are exploited in the auditory modality.

We have three hypotheses:

> *Hypothesis 1:* Visual sequences alternating in temporal frequency (sequences in which visual events alternate in higher and lower rates of change) should be grouped as trochees (the element with highest temporal frequency will be perceived as beginning a group).
>
> *Hypothesis 2:* Visual sequences alternating in duration (sequences in which short and long visual events alternate) should be grouped as iambs (the longest visual event will be perceived as marking the end of a visual sequence).
>
> *Hypothesis 3:* Visual sequences alternating in intensity (sequences in which bright and dim visual events alternate) should be grouped as trochees (the element with highest intensity will be perceived as beginning a group).

The visual sequence used during the familiarization phase of our experiments consisted of a series of 10 visual events (a triangle event, a circle event, a pentagon event, a star event, and so on) repeated several times in the same order. Each of the 10 visual events had a shape associated with it, and it consisted of 15 identical exemplars of this shape presented simultaneously in random positions on the computer screen (e.g., 15 triangles or 15 circles randomly changing their location on the computer screen). The visual events were presented changing their location at either a high (25 Hz) or a low (10 Hz) temporal frequency (the equivalent of auditory pitch), for either a long (800 ms) or a short (320 ms) time (the equivalent of auditory duration), and with either high (60–80 μW/cm^2) or low (5–15 μW/cm^2) brightness (our visual equivalent to auditory intensity). During test, participants saw pairs of single shapes (i.e., two of the 10 shapes associated with events during familiarization, such as a star and a square) without information about pitch, duration, or intensity at which they appeared during the familiarization and were asked whether these shapes occurred adjacently and in the same order during the familiarization. Studies on visual segmentation indicate that memory for visual events is a robust implicit measure of segmentation of continuous visual sequences (Kurby & Zacks, 2008; Swallow & Zacks, 2008; Swallow et al., 2009).

Experiment 1: Grouping of Visual Events Alternating in Temporal Frequency

Method

Participants. Twenty participants (10 women, 10 men) were tested. Their ages ranged from 19 to 28 years. They were all native speakers of Italian with either normal or corrected-to-normal vision, and received monetary compensation for their participation.

Stimuli. We created the visual sequence for the familiarization phase by concatenating 10 visual events and repeating them in the same order for 3 min. Each visual event was composed of an arrangement of 15 identical exemplars of a particular shape (e.g., 15 stars or 15 squares) randomly changing their location on the computer screen. In the first experiment, all visual events were presented for 800 ms (constant duration), with a brightness of 60–80 μW/cm^2 (constant intensity), but the shapes within each of the 10 visual events changed their location on the computer screen at either a high (25 Hz) or a low (10 Hz) temporal frequency (alternation in pitch). The high temporal frequency was attained by randomly changing the location of the shapes every 40 ms (25 Hz), whereas the low rate was achieved by changing the position of the shapes every 100 ms (10 Hz; see Figure 1A and supplemental materials, Video 1).

Visual sequences were presented in the center (in a 270 × 230-mm^2 area) of a 21-inch (53.34-cm), 1024 × 768-pixel resolution screen. Participants were seated at a distance of 60–70 cm from the screen, allowing the visual events to be presented in the fovea (i.e., visual events subtended between 1.6° and 1.9° of visual angle in height and between 1.5° and 1.6° in width).

The sequence of shapes in the visual events consisted of a square, a four-pointed star, a cross, an arc, a triangle, a circle, a pentagon, a five-pointed star, a rectangle, and a diamond. This sequence was identical for all participants. Visual events alternated in high and low temporal frequency, and the temporal frequency of the visual events was counterbalanced across participants. All visual events were presented in silver color against a black background. Because consecutive visual events alternated in temporal frequency, every switch in a visual event (and in shape) matched a change in temporal frequency.

For the test phase, we prepared a series of 20 static images composed of different pairings of the 10 shapes presented during familiarization. For each static image, one of the shapes of the pair was presented on the left side of the computer screen, and the other

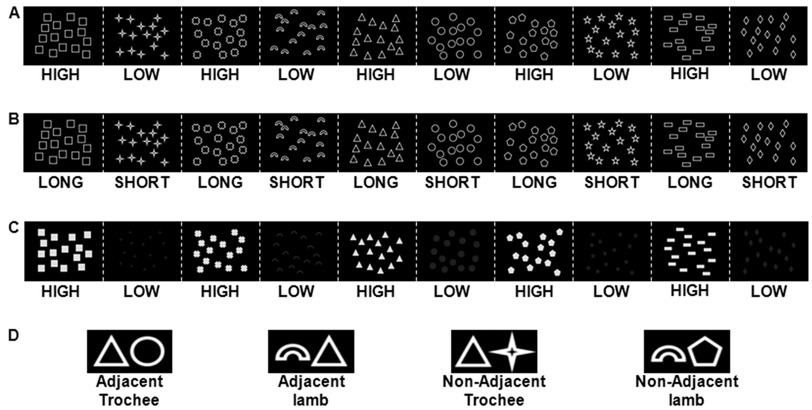

Figure 1. Shows the fixed order of visual events within the visual sequences used in all experiments. In Experiment 1, visual events alternated in temporal frequency (A); in Experiment 2, they alternated in duration (B); and in Experiment 3, they alternated in intensity (C). In each experiment, the stimulus properties (temporal frequency, duration, intensity) of specific visual events were counterbalanced between participants (e.g., the star shape was presented at 25 Hz for half the participants and at 10 Hz for the other half). Figure 1D shows an example of the test items used in Experiments 1–3.

shape was presented on the right side of the screen (see Figure 1D). As participants were facing the test items projected on the screen, the shape that appeared on the left side of the screen was projected on the right visual field of the participants, and vice versa. We assumed that when asked to judge their temporal order, Western participants would interpret the image on the left of the screen as temporally preceding the image on the right (Boroditsky, 2001).

The 20 combinations of shapes were subdivided into four types of test items: (a) five adjacent iamb test items consisting of pairs of shapes that appeared in adjacent visual events during familiarization, with a low temporal frequency visual event on the right and a high temporal frequency visual event on the left right side of the screen; (b) five adjacent trochee test items consisting of pairs of shapes that appeared adjacently during familiarization, with the shape of the high temporal frequency visual event on the right and the low temporal frequency visual event on the left side of the screen; (c) five nonadjacent iamb test items consisting of pairs of shapes that never occurred adjacently during familiarization, but the shape on the right side of the screen appeared with a low temporal frequency, whereas the shape on the left of the screen appeared with a high temporal frequency; and (d) five nonadjacent trochee test items consisting of two shapes that never appeared in adjacent visual events during familiarization, but the shape on the right side of the screen appeared with high temporal frequency, whereas the shape on the left side of the screen appeared with a low temporal frequency during familiarization. Nonadjacent Test Items c and d are designed (a) to evaluate whether participants segmented items into pairs during the familiarization phase or only encoded the elements that mark the beginning or the end of a sequence, and (b) to test for generalizations (i.e., whether participants generalize the trochaic or iambic pattern to items that did not occur adjacently during familiarization). Previous studies, in fact, have shown that systematic patterns of speech units can be learned after being familiarized to different exemplars (Gómez & Gerken, 2000). In our task, we used nonadjacent test items with an iambic and/or a trochaic pattern. Any preference for them could reflect the discovery of the systematic structure of the tokens and not only the memory for exemplars that occurred during familiarization.

The test images were static and did not contain any information about temporal frequency, duration, or intensity. In short, the design of the test was a 2 $\times$ 2 design, with Occurrence (adjacent and nonadjacent; i.e., shapes that occurred or not adjacently in visual events that were presented during familiarization) and Temporal Frequency (trochees and iambs; i.e., pairs of shapes in which the shape on the left side of the screen was presented with a high temporal frequency during familiarization and the shape on the right side was presented with a low temporal frequency, and vice versa) as within-subject factors (for an example of each type of the test items see Figure 1D).

Procedure. All experiments were carried out in a soundproof room that was dimly lit (0.05–0.1 $\mu W/cm^2$). Written instructions were presented on a computer screen informing the participants that the experiment would have two phases: a familiarization phase and a test phase. During the familiarization phase, participants were instructed to look attentively at the sequence of visual events and were asked to memorize the precise order in which they occurred, because at the end of the familiarization phase they would be asked to judge whether the order of a series of test items, composed of pairs of shapes, appeared in an order identical to that observed during familiarization.

During the test phase, participants were instructed to give their response, as accurately and fast as possible, by pressing a *yes* button if they thought that the pair of shapes had appeared in the same order and immediately adjacent during familiarization and by pressing a *no* button otherwise. Each of the 20 pairs of shapes (five per condition) was presented twice during test.

The shapes of the test items presented in visual events with high and low temporal frequency were counterbalanced across participants, and the presentation of the test stimuli was randomized across participants.

We expected that if the ITL plays a role in visual grouping, participants' memory for the pairs of shapes in the test items would be influenced by the temporal frequency of the presentation of the visual events during the familiarization. During familiarization, the visual sequences should thus be segmented into chunks, with the high temporal frequency visual events preferentially located in initial position (at the left side from the point of view of the participants). For example, we expected participants to segment the familiarization sequence [square (high), bright star (low), cross (high), arc (low), triangle (high), circle (low)] into pairs of trochaic visual events [square (high), bright star (low)], [cross (high), arc (low)], and [triangle (high), circle (low)]. In the test phase, adjacent trochaic test items (high–low temporal frequency pairs of visual events) should thus be better recognized than adjacent iambic test items (low–high temporal frequency pairs of visual events). That is, participants should be better at recognizing that the visual events [cross (left), arc (right)] occurred adjacently during familiarization as compared with the visual events [arc (left), triangle (right)], which also occurred adjacently during familiarization, but in which the first element had low temporal frequency and the second high temporal frequency. We did not predict significant differences for items that did not occur adjacently during the familiarization phase. That is, participants should be equally accurate in rejecting the pairs [square (left), circle (right)] and [circle (left), square (right)].

Results and Discussion

To estimate participants' memory for the iamb and trochee test items in both the adjacent and the nonadjacent conditions, we computed the percentage of "yes" responses for each of these four conditions (adjacent iambs, nonadjacent iambs, adjacent trochees, and nonadjacent trochees). Participants were asked to respond "yes" when they thought that a pair of shapes occurred adjacently and in the same order during familiarization and to respond "no" when they thought that a pair of shapes did not occur adjacently during familiarization. A correct response would be to respond "yes" for adjacent iambs and trochees and to respond "no" for nonadjacent iambs and trochees. Therefore, for adjacent iambs and trochees, the percentage of "yes" responses corresponds to the percentage of correct recognition, and for nonadjacent iambs and trochees it constitutes the percentage of incorrectly judging as adjacent during familiarization items that were not. The statistics using percentage of "yes" responses should lead to identical conclusion to those reached by using accuracy as our dependent measure ("yes" responses for adjacent items and "no" responses for nonadjacent items). The mean percentage of "yes" responses for Experiment 1 is shown in Figure 2.

We submitted participants' percentage of "yes" responses to a repeated measures analysis of variance with Temporal Frequency (iambs and trochees) and Occurrence (adjacent and nonadjacent) as within-subject factors. We expected participants to give significantly greater percentage of "yes" responses to items that occurred adjacently during familiarization as compared with items that did not occur adjacently during familiarization (indicating that they remembered the order of the events during familiarization), and within the items that occurred adjacently during familiarization, we expected participants to give a significantly greater percentage of "yes" responses to trochees than to iambs, as predicted by our hypothesis to extend the ITL to the visual modality. We did not predict significant differences in the percentage of "yes" responses for trochees and iambs that did not occur adjacently during familiarization. However, if participants give a significantly greater percentage of "yes" responses for nonadjacent trochaic test items, that may reflect that they have been able to generalize the trochaic pattern to items that did not occur adjacently during familiarization.

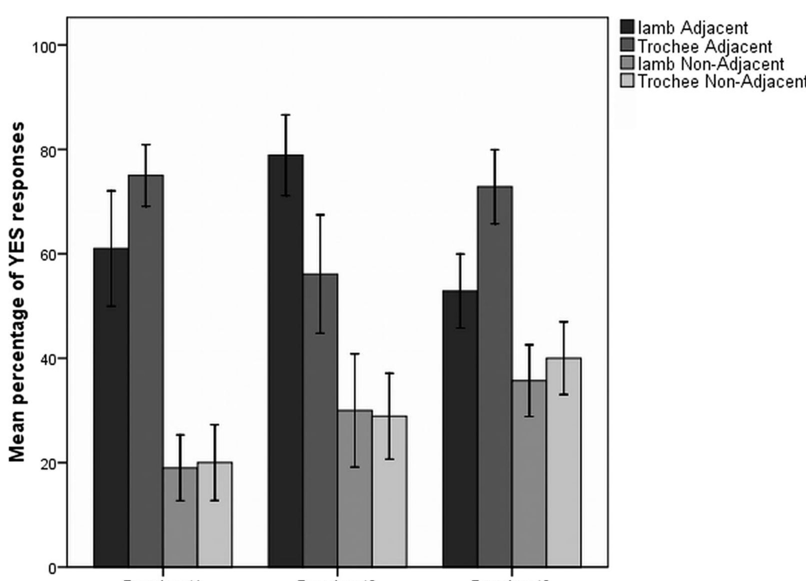

Figure 2. Illustrates the percentage of "yes" responses per experiment and type of test items. Experiments are indicated in the *x*-axis. Vertical lines indicate 2 standard errors of the mean. Although participants showed trochaic biases in Experiments 1 (temporal frequency) and 3 (intensity) and iambic biases in Experiment 2 (duration), no differences between iambs and trochees were found for the nonadjacent items.

We found a main effect of Temporal Frequency, $F(1, 19) = 5.3$, $p = .034$, $\eta^2 = .22$, because, overall, participants gave more "yes" responses to trochaic than to iambic test items. We also found a main effect of Occurrence, $F(1, 19) = 83.9$, $p < .001$, $\eta^2 = .81$, because, as expected, participants gave more "yes" responses to adjacent than to nonadjacent tests items. Crucially, we also found a significant interaction of Temporal Frequency × Occurrence. For adjacent test items only, participants gave more "yes" responses to trochees than to iambs, $F(1, 19) = 6.9$, $p = .017$, $\eta^2 = .27$. In short, when the familiarization phase alternated in temporal frequency, participants were more accurate in recognizing that two visual events occurred adjacently during familiarization when they formed a trochaic group than when they formed an iambic group (as indicated by their percentage of "yes" responses). No difference in the percentage of "yes" responses was found for the iambic and the trochaic test items that did not occur adjacently during familiarization.

Our results are thus consistent with an extension of the ITL to the processing of visual sequences with visual events alternating in temporal frequency. Participants were better at remembering pairs of visual events that begin with a visual event displayed at higher temporal frequency, indicating that they segmented the sequence into trochees. No differences were found for visual events that did not occur adjacently during familiarization, suggesting that participants were actually segmenting the sequence instead of preferentially encoding the test items that started or ended a sequence.

Experiment 2: Grouping of Visual Events Alternating in Duration

In the auditory domain, it has been shown that the distribution of the duration of the auditory events also influences the way in which auditory sequences are segmented and grouped. In Experiment 2, we explored the effect in segmentation and grouping of visual events alternating in duration.

Method

Participants. Eighteen participants (10 women, eight men) were tested. Their ages ranged from 19 to 25 years. They were all native speakers of Italian with either normal or corrected-to-normal vision, and received monetary compensation for their participation.

Stimuli. We used the same sequences as in Experiment 1; however, in Experiment 2, the visual events alternated in duration (i.e., 320 or 800 ms), whereas temporal frequency and intensity remained constant (25 Hz and 60–80 $\mu W/cm^2$, respectively; see Figure 1B and supplemental materials, Video 2). To ensure that the visual events were distinguishable in duration, we set the individual events at either 320 ms or 800 ms. Because the adjacent visual events alternated in duration, every switch in visual event (and in shape) matched a change in duration.

For the test phase, we used the same test items that we used in Experiment 1 (see Figure 1D).

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

We expected that if the ITL influences the segmentation of visual sequences alternating in duration, visual sequences should be grouped into chunks in which long visual events are preferentially located in final position (iambs). Therefore, in the test phase, participants should have a greater percentage of "yes" responses for adjacent than nonadjacent test items (indicating that they remembered the order of events during familiarization); moreover, in adjacent test items, participants should present a greater percentage of "yes" responses for the iambic than for the trochaic test items. We did not predict significant differences for nonadjacent test items.

Results and Discussion

The results of Experiment 2 are shown in Figure 2. We submitted participants' percentage of "yes" responses to a repeated measures analysis of variance with Duration (iamb and trochees) and Occurrence (adjacent and nonadjacent) as within-subject factors.

We found a main effect of Duration, $F(1, 17) = 8.8$, $p = .009$, $\eta^2 = .34$, because participants gave more "yes" responses to iambic than to trochaic test items, and a main effect of Occurrence, $F(1, 17) = 40.1$, $p < .001$, $\eta^2 = .70$, because participants correctly gave more "yes" responses to adjacent test items than to nonadjacent test items, Crucially, we also found a significant interaction of Duration × Occurrence, $F(1, 17) = 5.3$, $p = .030$, $\eta^2 = .24$. For adjacent test items only, participants gave more "yes" responses to iambs than to trochees, $F(1, 17) = 10.7$, $p = .004$, $\eta^2 = .39$. In short, when the familiarization phase alternated in duration, participants were more accurate in recognizing that two visual events occurred adjacently during familiarization when they formed an iambic group than when they formed a trochaic group.

Our results are thus consistent with the extension of the ITL to the processing of visual sequences with visual events alternating in duration. Participants were better at remembering the pairs of visual events that ended with a long event, indicating that they segmented the sequence into iambs. No differences were found for visual events that did not occur adjacently during familiarization.

Experiment 3: Grouping of Visual Events Alternating in Intensity

Because the ITL predicts that intensity also plays a role in the segmentation of auditory sequences, in Experiment 3, we explored the role of intensity in the segmentation and grouping of visual sequences.

Method

Participants. Fourteen participants (eight women, six men) were tested. Their ages ranged from 18 to 28 years. They were all native speakers of Italian with either normal or corrected-to-normal vision, and received monetary compensation for their participation.

Stimuli. We used the same visual sequences as in Experiment 1 and 2; however, in Experiment 3, the visual events alternated in intensity instead of temporal frequency or duration. All the adjacent visual events were presented at constant duration (i.e., 800 ms) and constant temporal frequency (i.e., 25 Hz); however, intensity alternated from low (5–15 $\mu W/cm^2$) to high (60–80 $\mu W/cm^2$) in the adjacent visual events (see Figure 1C and supplemental materials, Video 3). Because each visual event was displayed only at one of the two possible intensities, every switch in visual event matched a change in intensity. Test items were identical to those used in Experiments 1 and 2.

Procedure. The procedure was identical to that of the previous experiments.

We expected that the alternation in intensity of the visual events would influence the segmentation of the visual sequence. During familiarization, participants should segment the sequence of visual events into chunks, with the bright visual events in initial position (i.e., trochees). Therefore, in the test phase, participants should give a significantly greater percentage of "yes" responses for adjacent trochaic than for iambic test items. As in the previous experiments, we did not expect differences in percentage of "yes" responses for nonadjacent test items.

Results and Discussion

The results of Experiment 3 are shown in Figure 2. We submitted participants' percentage of "yes" responses to a repeated measures analysis of variance with Intensity (iambs and trochees) and Occurrence (adjacent and nonadjacent) as within-subject factors. We found a main effect of Occurrence, $F(1, 13) = 61.5$, $p < .001$, $\eta^2 = .80$, because participants gave more "yes" responses to adjacent than to nonadjacent tests items. No effect of Intensity was found. However, we found a significant interaction of Intensity × Occurrence, because only for adjacent test items, participants gave more "yes" responses to trochees than to iambs, $F(1, 13) = 11.1$, $p = .006$, $\eta^2 = .46$. In short, when the familiarization phase alternated in intensity, participants were more accurate in recognizing that two visual events occurred adjacently during familiarization when they formed a trochaic group than when they formed an iambic group.

Our results are thus consistent with an extension of the ITL to the processing of visual sequences with visual events alternating in intensity. Participants were better at remembering the pairs of visual events that started with a bright visual event, indicating that they segmented the sequence into trochees. No differences were found for visual events that did not occur adjacently during familiarization.

General Discussion

In order to make sense of a constantly changing world, it is necessary to segment it into meaningful units. In the auditory domain, segmentation is used in speech to find units such as words or phrases. Once speech is segmented, it is possible to discover regularities such as phonotactics, the ordering of words within phrases, as well as nonadjacent dependencies (Peña, Bonatti, Nespor, & Mehler, 2002). In the visual domain, segmentation is used in sign language to find units such as individual signs or phrases. Of course, segmentation is also used in both the auditory and the visual modalities for nonlinguistic purposes, such as parsing events in the environment.

The ITL was originally formulated to account for the grouping of auditory stimuli, first in music and then in speech. The ITL establishes that sequences of sounds that alternate mainly in duration are grouped with the most prominent element in final position (an iambic pattern), whereas sequences of sounds that alternate mainly in pitch or intensity are grouped with the most prominent element in initial position (a trochaic pattern). In three experiments, we investigated whether this law, besides being general in the acoustic modality, also organizes grouping in the visual modality.

In the experiments we presented, our plan was to concentrate on sequences of visual stimuli that differ either in duration, temporal frequency (a possible visual analogue of pitch), or intensity, to investigate whether these three cues determine different grouping of visual sequences. The results of our three experiments are in agreement with the prediction that both higher temporal frequency and higher intensity begin a group (trochees), whereas longer duration ends a group (iambs). Interestingly, although the task at hand did not directly address these cues (participants were instructed to memorize the order of visual events in a sequence and were not asked to pay attention to the changes in the temporal frequency, duration, or intensity of the shapes), these perceptual cues still influenced participants' segmentation of a visual sequence.

This automatic segmentation influenced by subtle cues might very well have been beneficial to our participants. Humans have a limited processing capacity of around seven chunks (Miller, 1956), or even less (Cowan, 2001). Thus seeing our visual sequences as composed of pairs of visual events, instead of individual stimuli, might decrease the processing load. Specifically, interpreting the sequence [star, circle, square, diamond] as [star, circle] and [square, diamond] might improve memorization in a similar way as segmenting the visual sequence BBCNBCCNN into three names of broadcasters (BBC, NBC, CNN). Even preverbal untrained infants use chunking to increase their visual memory (Feigenson & Halberda, 2008).

Interestingly, these perceptual biases might potentially allow visual and acoustic events to be grouped without extensive experience of the stream, in contrast with classical statistical computations of transition probabilities, which are highly experience dependent (Saffran et al., 1996). Although transition probabilities by definition need to be computed over longer sequences of items, and the length of the sequence determines the reliability of these computations, the ITL can operate (at least in theory) on very short streams. For example, the familiarization streams of Hay and Diehl (2007) lasted only 11 s, and this short familiarization was enough for participants to group stimuli as iambs or trochees. Future studies should investigate how the visual ITL and statistical computations interact in the segmentation of visual events.

The present work also offers a contribution to the discussion about the nature of the mechanisms of language acquisition, specifically to their specific or general nature. We have shown that the ITL, which influences grouping in the linguistic domain (Hayes, 1995; Nespor et al., 2008), might not be specific to the acoustic modality. If the ITL could guide the infant into language acquisition by signaling the relative order of heads and complements in oral languages, as proposed in Nespor et al. (2008), the fact that it appears not be restricted to the acoustic modality suggests that it also might well guide infants in the acquisition of sign languages. The ITL may thus constitute yet another example of the generality of language acquisition mechanisms. The specificity of language acquisition might thus lie not only in mechanisms that are specific to language but also in the way different general perceptual mechanisms and computational biases interact throughout development.

This claim is in agreement with the fact that several computational and rule-learning mechanisms, initially thought to apply only to language, were later shown to be both domain general (e.g., Saffran, Johnson, Aslin, & Newport, 1999) and modality general (e.g., Kirkham, Slemmer, & Johnson, 2002). Interestingly, however, these apparently general mechanisms, when used to process linguistic stimuli, are influenced by several factors, including linguistic properties (e.g., prosody, Shukla, Nespor, & Mehler, 2007; the categorical distinction between vowels and consonants, Bonatti, Peña, Nespor, & Mehler, 2005, 2007; Toro, Nespor, Mehler, & Bonatti, 2008; Toro, Shukla, Nespor, & Endress, 2008; coarticulation, Fernandes, Ventura, & Kolinsky, 2007), information-processing capacities (e.g., attention, Toro, Sinnett, & Soto-Faraco, 2005), and language experience (e.g., phonotactic constraints, Finn & Hudson Kam, 2008). It is not clear whether similar computational specificities are also present when the ITL is used to group visual sequences instead of linguistic stimuli. Therefore, showing that the iambic–trochaic bias is modality general does not imply that it does not acquire a special status when applied to the segmentation of speech.

Interestingly, recent studies appear to indicate that perceptual grouping guided by the iambic–trochaic bias, instead of being guided uniquely by innate perceptual principles, might be influenced by the properties of one's native language as well (Iversen, Patel, & Ohgushi, 2008). Iversen et al. (2008) compared the perception of a sequence of tones alternating in either intensity or duration by Japanese and English listeners. The two groups of listeners did not differ in their grouping of tones alternating in intensity, always placing the prominent element in initial position (trochee). However, the two groups differed in their grouping of stimuli alternating in duration. Specifically, Japanese listeners showed no consistent segmentation bias, whereas English listeners tended to place the long element in final position (grouping the sequences as iambs). A similar pattern of results was also observed with 5-month-old Japanese and English learning infants (Yoshida et al., 2010). It remains to be clarified whether speakers of languages other than Italian (e.g., Japanese) have the same biases to segment our visual sequences they showed when segmenting auditory sequences.

Future studies should, in addition, focus on how information from the visual and auditory modalities are integrated in speech segmentation. Some proposals of speech perception claim that speech sounds are perceived as articulatory gestures (e.g., Best & McRoberts, 2003; Browman & Goldstein, 1992), and some broader proposals of speech perception argue that sounds are perceived in terms of audible source attributes (e.g., Gaver, 1993). For speech, these models propose that when one hears a high-pitched speech sound, one perceives vocal folds vibrating at a high temporal frequency. In this framework, information about the visual and the auditory world could potentially work together to give rise to the ITL, and be used in conjunction in order to segment visual and auditory sequences. In a similar direction, it should be explored how other properties of movement (in addition to changes in temporal frequency) might influence the segmentation of visual sequences and whether the visual ITL would apply to a broader array of visual events.

Given the relevance of the ITL at the phrasal level for language acquisition in oral languages (Bion et al., 2011; Nespor et al., 2008), research in the visual modality should investigate whether the ITL also characterizes prominence at the phrasal level in sign languages. It is clear that sign languages have temporal regularities due to both the nature of human movements and their own specific properties, which could be exploited in segmentation. As for duration, body movements seem to have a natural tendency to have final lengthening due to muscular relaxation at the end of a planned movement. It is feasible that the parallel of vocal pitch and intensity in human movement is the natural tendency to decrease the temporal frequency and the intensity of a movement as it is ceasing. Some of these properties have in fact been described in different sign language prosodies (Nespor & Sandler, 1999; Wilbur, 2000). If we were to find that the ITL indeed generally characterizes the prosody of sign languages, the experiments we presented would constitute an encouragement to experimentally test whether there are biases in the perception of sign languages to group signs as predicted by the ITL.

References

Best, C. C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and speech, 46,* 183–216. doi:10.1177/00238309030460020701

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech, 54,* 123–140. doi:10.1177/0023830910388018

Bion, R. A. H., Höhle, B., & Schmitz, M. (2007). The role of prosody on the perception of word-order differences by 14-month-old German infants. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1537–1540). Saarbrücken, Germany: International Congress of Phonetic Sciences.

Blakemore, C., Adler, K., & Pointon, M. (1993). *Vision: Coding and efficiency.* Cambridge, England: Cambridge University Press.

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology, 6,* 145–238. doi:10.2307/1410948

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16,* 451–459. doi:10.1111/j.0956-7976.2005.01565.x

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychological Science, 18,* 924–925. doi:10.1111/j.1467-9280.2007.02002.x

Boroditsky, L. (2001). Does language shape thought? English and Mandarin speakers' conceptions of time. *Cognitive Psychology, 43,* 1–22. doi:10.1006/cogp.2001.0748

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49,* 155–180.

Christophe, A., Nespor, M., Guasti, M. T., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-direction parameter. *Developmental Science, 6,* 211–220. doi:10.1111/1467-7687.00273

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

Corina, D. P., & Knapp, H. P. (2008). Signed language and human action processing: Evidence for functional constraints on the human mirror-neuron system. In G. F. Eden & D. L. Flowers (Eds.), *Annals of the New York Academy of Sciences* (Vol. 1145, pp. 100–112). New York, NY: Wiley. doi:10.1196/annals.1416.023

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185. doi:10.1017/S0140525X01003922

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 9926–9930. doi:10.1073/pnas.0709884105

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception & Psychophysics, 69,* 856–864. doi:10.3758/BF03193922

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition, 108,* 477–499. doi:10.1016/j.cognition.2008.04.002

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 15822–15826. doi:10.1073/pnas.232472899

Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology, 5,* 1–29. doi:10.1207/s15326969eco0501_1

Gleitman, L. R., & Wanner, E. (1982). The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge, England: Cambridge University Press.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4,* 178–186. doi:10.1016/S1364-6613(00)01467-4

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition, 34,* 1221–1235. doi:10.3758/BF03193267

Hay, J. S. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics, 69,* 113–122. doi:10.3758/BF03194458

Hayes, B. (1995). *Metrical stress theory: Principles and case studies.* Chicago, IL: University of Chicago Press.

Hess, R. F., & Snowden R. J. (1992). Temporal properties of human visual filters: Number, shapes and spatial covariation. *Vision Research, 32,* 47–59. doi:10.1016/0042-6989(92)90112-V

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,* 2263–2271. doi:10.1121/1.2973189

Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 627–640. doi:10.1037/0096-1523.19.3.627

Kelly, D. H., & Burbeck, C. A. (1984). Critical problems in spatial vision. *Critical Reviews in Biomedical Engineering, 10,* 125–177.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83,* B35–B42. doi:10.1016/S0010-0277(02)00004-5

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1,* 70–73. doi:10.1111/j.1467-9280.1990.tb00070.x

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12,* 72–79. doi:10.1016/j.tics.2007.11.004

Lamb, T. D. (1991). The role of photoreceptors in light-adaptation and dark-adaptation of the visual system. In C. Blakemore (Ed.), *Vision coding and efficiency* (pp. 161–168). Cambridge, England: Cambridge University. doi:10.1017/CBO9780511626197.017

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283,* 77–80. doi:10.1126/science.283.5398.77

McGurk, H., & MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748. doi:10.1038/264746a0

McNeill, D. (2005). *Gesture and thought.* Chicago, IL: University of Chicago Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97. doi:10.1037/h0043158

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, Germany: Akademie Verlag.

Nespor, M., & Sandler, W. (1999). Prosody in Israeli sign language. *Language and Speech, 42,* 143–176. doi:10.1177/00238309990420020201

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, 7,* 1–20.

Nespor, M., & Vogel, I. (2008). *Prosodic phonology* (2nd ed.). Berlin, Germany: de Gruyter.

Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298,* 604–607. doi:10.1126/science.1072901

Poizner, H., & Kegl, J. (1993). Neural disorders of the linguistic use of space and movement. In P. Tallal, A. M. Galaburda, R. R. Llinás, & C. von Euler (Eds.), *Annals of the New York Academy of Sciences* (Vol. 682, 192–213). New York, NY: Wiley. doi:10.1111/j.1749-6632.1993.tb22969.x

Reisbeck, T. E., & Gegenfurtner, K. R. (1999). Velocity tuned mechanisms in human motion processing. *Vision Research, 39,* 3267–3285. doi:10.1016/S0042-6989(99)00017-6

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science, 274,* 1926–1928. doi:10.1126/science.274.5294.1926

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. doi:10.1016/S0010-0277(98)00075-4

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition, 105,* 669–680. doi:10.1016/j.cognition.2006.11.004

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology, 54,* 1–32. doi:10.1016/j.cogpsych.2006.04.002

Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development, 4,* 157–178. doi:10.1080/15475440801922131

Spillmann, L. (2006). From perceptive fields to Gestalt. *Progress in Brain Research, 155,* 67–92. doi:10.1016/S0079-6123(06)55005-8

Swallow, K. M., & Zacks J. M. (2008). Sequences learned without awareness can orient attention during the perception of human activity. *Psychonomic Bulletin & Review, 15,* 116–122. doi:10.3758/PBR.15.1.116

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General, 138,* 236–257. doi:10.1037/a0015631

Thompson, P. G. (1982). Perceived rate of movement depends on contrast. *Vision Research, 22,* 377–380. doi:10.1016/0042-6989(82)90153-5

Thompson, P., & Stone, L. S. (1997). Contrast affects flicker and speed perception differently. *Vision Research, 37,* 1255–1260. doi:10.1016/S0042-6989(96)00302-1

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science, 19,* 137–144. doi:10.1111/j.1467-9280.2008.02059.x

Toro, J. M., Shukla, M., Nespor, M., & Endress, A. D. (2008). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception & Psychophysics, 70,* 1515–1525. doi:10.3758/PP.70.8.1515

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97,* B25–B34. doi:10.1016/j.cognition.2005.01.006

Wilbur, R. B. (1999). Stress in ASL: Empirical evidence and linguistic issues. *Language & Speech, 42,* 229–250. doi:10.1177/00238309990420020501

Wilbur, R. B. (2000). Phonological and prosodic layering of nonmanuals in American Sign Language. In K. Emmorey & H. Lane (Eds.), *The signs* *of language revisited: An anthology to honor Ursula Bellugi and Edward Klima* (pp. 215–244). Hillsdale, NJ: Erlbaum.

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

Wright, D. B., & Wareham, G. (2005). Mixing sound and vision: The interaction of auditory and visual information for earwitnesses of a crime scene. *Legal and Criminological Psychology, 10,* 103–108. doi:10.1348/135532504X15240

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werner, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition, 115,* 356–361. doi:10.1016/j.cognition.2010.01.005

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28,* 979–1008. doi:10.1207/s15516709cog2806_5

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition, 112,* 201–216. doi:10.1016/j.cognition.2009.03.007

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science, 16,* 80–84. doi:10.1111/j.1467-8721.2007.00480.x

Received July 8, 2010
Revision received March 21, 2011
Accepted March 23, 2011 ■