Acoustic Markers of Prominence Influence Infants' and Adults' Segmentation of Speech Sequences

Ricardo A. H. Bion, Silvia Benavides-Varela and Marina Nespor

Language and Speech published online 10 January 2011
DOI: 10.1177/0023830910388018

Article

Language *and* Speech

Acoustic Markers of Prominence Influence Infants' and Adults' Segmentation of Speech Sequences

Language *and* Speech
XX(X) 1–18

DOI: 10.1177/0023830910388018
las.sagepub.com

Ricardo A. H. Bion[1, 2], Silvia Benavides-Varela[2], Marina Nespor[2, 3, 4]

Abstract

Two experiments investigated the way acoustic markers of prominence influence the grouping of speech sequences by adults and 7-month-old infants. In the first experiment, adults were familiarized with and asked to memorize sequences of adjacent syllables that alternated in either pitch or duration. During the test phase, participants heard pairs of syllables with constant pitch and duration and were asked whether the syllables had appeared adjacently during familiarization. Adults were better at remembering pairs of syllables that during familiarization had short syllables preceding long syllables, or high-pitched syllables preceding low-pitched syllables. In the second experiment, infants were familiarized and tested with similar stimuli as in the first experiment, and their preference for pairs of syllables was accessed using the head-turn preference paradigm. When familiarized with syllables alternating in pitch, infants showed a preference to listen to pairs of syllables that had high pitch in the first syllable. However, no preference was found when the familiarization stream alternated in duration. It is proposed that these perceptual biases help infants and adults find linguistic units in the continuous speech stream. While the bias for grouping based on pitch appears early in development, biases for durational grouping might rely on more extensive linguistic experience.

Keywords

language development, memory, perceptual biases, speech segmentation

Introduction

When learning a language, infants are faced with the task of finding words and phrases, and of discovering the relation between these units in an apparently continuous stream. In order to get the

[1]Stanford University
[2]International School for Advanced Studies ISAS-SISSA, Trieste
[3]University of Milan-Bicocca
[4]Accademia dei Lincei

Corresponding author:
Ricardo A. H. Bion, Department of Psychology, Stanford University, Jordan Hall, Building 420, Stanford, CA 94305, USA
Email: ricardoh@stanford.edu

analysis started, predispositions for paying attention to specific acoustic cues, tendencies to generalize, and constraints against overgeneralization must be present at a very early age. Discovering the way these abilities develop from infancy to adulthood is one of the main challenges for language acquisition research.

Several studies indicate that infants' first linguistic preferences and discrimination abilities are guided by the prosodic and rhythmic properties of speech. Some days after birth, infants rely on these properties to distinguish between different languages (Nazzi, Bertoncini, and Mehler, 1998; Ramus, Hauser, Miller, Morris, and Mehler, 2000). They also appear to exploit them to identify their native language and their mother's voice (De Casper and Fifer, 1980). Within their first months of life, infants make use of prosody to find words, clauses, and phrases in fluent speech (Christophe, Gout, Peperkamp, and Morgan, 2003; Gout, Christophe, and Morgan, 2004; Jusczyk, Cutler, and Redanz, 1993; Jusczyk et al., 1992). The correlations between linguistic units and prosodic cues (e.g., Cooper and Paccia-Cooper, 1980; Goldman-Eisler, 1972), and infants' early sensitivity to the latter, have led to the proposal that acoustic information, such as variations in duration and pitch, might help infants break into language (for a review, Morgan and Demuth, 1996; Weissenborn and Höhle, 2001a, 2001b).

A central issue in the field of language acquisition concerns the extent to which these perceptual biases are innate or emerge from linguistic experience. One of the earliest prosodic biases observed is a preference for trochaic over iambic words in early infancy. English-learning 9-month-olds prefer to listen to lists of trochees over lists of iambs (Jusczyk et al., 1993), and German-learning infants show the same preference as early as 6 months of age (Höhle, Bijeljac-Babicb, Weissenborn, and Nazzi, 2009). At a similar point in development, Dutch, German, and English learning infants have a trochaic bias in the segmentation of words from fluent speech (Cutler, Mehler, Norris, and Segui, 1986; Cutler and Norris, 1988; Höhle, 2002; Houston, Jusczyk, Kuijpers, Coolen, and Cutler, 2000).

Three different explanations for infants' early trochaic biases are possible. One of the possibilities is that the trochaic bias is innate and universal (Allen and Hawkins, 1978), either being present at birth, or emerging between 6 and 9 months of age, when it is observed in different languages (Höhle et al., 2009; Jusczyk et al., 1993). According to this proposal, mistakes in segmentation will occur in case infants are learning a language that does not follow a trochaic pattern at the word level. A second possibility is that the trochaic bias emerges after generalizations are extracted over isolated words or over words segmented using distributional analyses (Thiessen and Saffran, 2003). This proposal is supported by experiments that show that when prosodic and distributional cues are contrasted, 7-month-old infants prefer statistically coherent words at the segmental level over trochaic words (Thiessen and Saffran, 2003). A third possibility is that trochaic biases only emerge for infants learning stress-based languages such as German, Dutch, or English, and should thus not be present in cases where infants are learning syllable-timed languages such as French or Italian (Nazzi, Iakimova, Bertoncini, Frédonie, and Alcantara, 2006).

This third proposal, known as the rhythmic-activation proposal, is supported by two different pieces of evidence. The first supporting evidence is that infants are sensitive to the rhythmic properties of languages from birth onwards (Nazzi, Bertoncini, and Mehler, 1998; Ramus et al., 2000), and therefore have the perceptual mechanisms necessary to trigger the trochaic bias. The second source of evidence comes from cross-linguistic studies that show that infants learning French, a syllabic language, segment words using a syllabic strategy, while infants learning English, a stress-based language, rely on a trochaic bias instead (Nazzi et al., 2006). Additional evidence comes from the fact that infants learning German prefer trochaic over iambic words at 6 months of age, while this bias is not present in French-learning infants of the same age (Höhle et al., 2009). In

addition, studies using event-related potentials find different neural responses to trochaic and iambic words in French- and German-learning infants as early as 4 months of age (Friederici, Friedrich, and Christophe, 2007).

However, these studies fail to acknowledge that stress is marked by different acoustic cues in different languages, and that infants' grouping biases might rely on acoustic cues rather than on adults' perception of the position of lexical stress. Most of the acoustic and perceptual descriptions of word stress focused on Germanic stress-timed languages, and indicated that word stress is marked mainly by changes in pitch and duration, and less reliably by changes in intensity (Cutler, 2005). It is generally agreed that in English, none of these acoustic cues separately is crucial, but that all are evaluated simultaneously when judging lexical stress (Lieberman, 1960). However, in other languages, these cues have been shown to be orthogonal. In Welsh, strong effects of duration, but inconsistent effects of pitch, are found in the perception of lexical stress (Williams, 1985). In Thai and other tone languages, stress is signaled uniquely by duration (Potisuk, Gandour, and Harper, 1996), while in Polish stress is signaled uniquely by pitch (Dogil, 1999). Even when the same acoustic cues are used, their exact weighting might change across languages (Bleakley, 1973; Llisterri, Machuca, de la Mota, Riera, and Rios, 2003) and even across speakers of the same language (Nakatani and Aston, 1978). Therefore, talking about a trochaic universal bias without taking into account the way stress is realized at the acoustic levels across languages is problematic. This fact is aggravated by the fact that in many cross-linguistic studies, stimuli from one single language are used to test perceptual biases in several different languages (Friederici et al., 2007; Höhle et al., 2009).

Different acoustic realizations of stress are also present at other linguistic levels. French, for example, does not have consistent stress at the lexical level, but is an iambic language at the level of the phonological phrase, with stress marked mainly by increased duration (Dupoux, Pallier, Sebastián-Gallés, and Mehler, 1997). Even within languages of the same rhythmic class, different acoustic cues are used to signal clause boundaries. In English, for example, stress is marked by higher resets in pitch than in Dutch, which relies more heavily on pauses (Johnson and Seidl, 2008). Interestingly, Johnson and Seidl (2008) found that Dutch- and English-learning infants rely on different acoustic cues to segment clauses (Johnson and Seidl, 2008). However, in their conclusions, the authors point out that two explanations are possible for this pattern of results: on the one hand, it is possible that the infants weight acoustic cues differently; on the other hand, it is possible that infants use a universally similar cue weighting, but that clause boundaries are not as saliently marked by pitch and duration in Dutch. A similar acknowledgement should be made by studies discussing infants' early trochaic biases, with special emphasis on cross-linguistic variation and on the influence of acoustic cues on different levels of linguistic processing, such as the word and the phonological phrase level, on infants' early biases.

Referring back to perceptual biases for grouping, studies with adults show that iambic and trochaic biases can be switched depending on the acoustic cue being manipulated. In short, humans segment sequences of sounds that differ only in duration with the longest element in final position (iambs) and sequences of sounds that differ only in intensity with the more intense element in initial position (trochees). This bias was initially proposed to explain the grouping of musical sequences (Bolton, 1894; Cooper and Meyer, 1960; Woodrow, 1951), but was later expanded to account for regularities in speech production and biases in speech perception.

The first formulations of the iambic-trochaic law for language focused on the word level and on the effects of duration and intensity on the grouping of syllables. In speech production, this law was proposed to account for the location of word secondary stress crosslinguistically. Specifically, if secondary stresses are initial in the phonological constituent known as the metrical foot, then they are manifested mainly with increased intensity; if they are final, they are manifested mainly

with increased duration (Hayes, 1995). The existence of the iambic-trochaic grouping principle is confirmed by experimental evidence. Hay and Diehl (2007), for example, investigated grouping preferences for linguistic and non-linguistic stimuli by French and English speakers. Listeners were presented a sequence of repeating syllables or tones varying in either intensity or duration, and had to decide whether the sequence was formed by trochaic or iambic sub-sequences. Results from both modalities (speech and non-speech) and from both languages (English and French) corroborated the iambic-trochaic law, supporting previous claims that it must be guided by general principles of the human auditory system (Bolton, 1894; Cooper and Meyer, 1960; Woodrow, 1951).

The most recent extensions of the iambic-trochaic law focused on the phrasal level and on the role of duration and pitch. In speech production, it has been proposed that the location of stress in phrasal constituents differs depending on whether it is initial or final, i.e. on the first or on the last word of the phonological phrase (Nespor and Vogel, 2008). It has then been proposed that the physical realization of phonological phrase stress also varies depending on its location: increased duration characterizes iambic patterns, while greater intensity and higher pitch characterize trochaic patterns (Nespor, Shukla, van de Vijver, Avesani, Schraudolf, and Donati, 2008). These findings, however, should be replicated with infant-directed speech. Still, from this body of studies, it is possible to conclude that in speech production, units with higher pitch or intensity mark the beginning of speech sequences, whereas units having long duration mark the end of speech sequences. However, it was not investigated whether pitch would also lead to trochaic grouping in speech perception as previously shown for intensity (Hay and Diehl, 2007).

Interestingly, Nespor et al. (2008) propose that the different manifestation of prominence within phonological phrases might be exploited to acquire word order (cf. also Nespor, Guasti, and Christophe, 1996). Languages with initial prominence (trochaic prominence marked by pitch) within phonological phrases have syntactic heads that follow their complements, and languages with final prominence (iambic prominence marked by duration) within phonological phrases have syntactic heads that precede their complements (Nespor and Vogel, 2008). Broadly, head-direction refers to the ordering of complements with respect to their head: for example, objects with respect to either verbs or adpositions, or nominal complements with respect to nouns. In head-initial languages, like English and French, for example, verbs precede objects, while in head-final languages, like Turkish and Japanese, verbs follow objects. In addition, in head-initial languages, function words tend to precede content words, while in head-final languages function words tend to follow them. Importantly, Nespor et al. (2008) also found that in German, a language in which the object can either follow or precede the verb or the adposition depending on syntactic factors, the prominence of the phonological phrase is differently realized: mainly through duration, as in French, when the object is final and mainly through pitch, as in Turkish, when the object is initial. The iambic-trochaic law could thus also be of aid to infants exposed to a language with mixed word order.

This theoretical proposal became more appealing when Christophe, Guasti, Nespor, and van Ooyen (2003) showed that infants are sensitive to these prosodic differences. Infants of 6 to 12 weeks of age can discriminate between phrases with different orders of heads and complements; specifically, they can distinguish French (head-initial) from Turkish (head-final) phrases. Importantly, the sentences in this study were resynthesized to remove phonemic information, and were also matched on all factors that are thought to influence prosody, differing only in their stress pattern within phonological phrases. Namely, French follows the prominence final pattern of head-initial languages, and Turkish follows the prominence initial pattern of head-final languages.

In addition to evidence across languages, infants were also shown to be sensitive to prosodic cues that signal the order of heads and complements within a single language. Bion, Höhle, and Schmitz (2007) found that 14-month-old German infants distinguish between head-initial and

head-final German phrases, but that this ability is no longer present when the prosodic information is not isomorphic with the ordering of words within phrases. In these studies, however, the exact acoustic cues on which infants were relying for discriminating phrases within the same language (Bion et al., 2007) and across languages (Christophe et al., 2003) were not manipulated, and infants' preferences could be guided by variations in intensity, pitch, or duration. In music perception, 4.5-month-old infants were shown to rely on longer duration and lower pitch to perceive boundaries of tonal groups (Jusczyk and Krumhansl, 1993). However, once more, pitch and duration were present simultaneously in the stream, and it is not possible to determine their use in isolation.

In the present study, we aimed at investigating this issue further, focusing on how acoustic markers of prominence influence the grouping of linguistic units by infants and adults. The current study is different from the previous studies discussed so far in several respects. First, the previous studies with adults used a direct measurement of perceptual grouping, presenting a sequence of syllables or tones to the listeners, and then asking whether participants had heard iambic or trochaic sequences (Hay and Diehl, 2007). Since subjects heard the same syllable or tone repeated several times, having to decide on an internal ordering might be considered an artificial task. In the present study, we aim to investigate perceptual grouping indirectly, by focusing on whether linguistic units congruent with the iambic-trochaic law are better remembered than incongruent units, in the case of adults, or preferred over incongruent units, in the case of infants. Studies on event segmentation in the visual domain show that event segmentation is automatic and is best observed by implicit behavioral and neurophysiological measures. Often, direct measures, such as asking subjects about event boundaries, can influence the perceptual processes they attempt to measure (Zacks and Swallow, 2007). Studies on visual segmentation indicate that memory for events is a robust implicit measure of event segmentation (Kurby and Zacks, 2008; Swallow and Zacks, 2008; Swallow, Zacks, and Abrams, 2009). An additional advantage of using an indirect measure with adults lies in the possibility of better comparing the results from the adult and infant experiments. In the case of the infant experiment, it is, in fact, obviously not possible to ask participants for conscious judgments of event boundaries.

Another fundamental difference between our study and previous ones lies in the acoustic markers of prominence under investigation. Specifically, while previous studies focused on the effect of intensity in rhythmic grouping, the present study focuses on pitch instead. It is hard to dissociate the effect of pitch from the effect of intensity, since sounds with higher pitch tend to be perceived as having higher intensity. However, it has been shown that, at the onset of lexical recognition, infants are sensitive to variations in pitch but ignore variations in intensity (Singh, White, and Morgan, 2008). This strategy is particularly profitable, as amplitude variations are highly influenced by the listening environment, such as the distance or the presence of a barrier between speaker and listener. In contrast, pitch is modulated solely by the speaker. While most languages use pitch to encode linguistic information, no language is known to rely only on intensity to signal linguistic contrasts. These pieces of evidence taken together appear to indicate that intensity might be a very poor candidate to guide infants' early perceptual biases.

The biggest difference between our and previous studies, however, lies in the populations under investigation. To our knowledge, this is the first study to use a similar set of familiarization and test stimuli to test both infants and adults. The developmental nature of our study can bring insights into the amount of linguistic exposure necessary for the emergence of the iambic-trochaic biases. Specifically, two studies were carried out: the first investigates the role of alternations in pitch or duration on memory of speech sequences by Italian adults, while the second investigates the effect of these same markers of prominence on the listening preferences of Italian 7-month-old infants.

We predict that adult native speakers will group the sequences as duration-final and pitch-initial. It is an open question how perceptual grouping will work with infants. The rhythmic-activation proposal (Nazzi et al., 2006, not to be confused with Nespor and colleagues' rhythmic-activation principle, 1996) would predict neither an iambic nor a trochaic bias in our infants, since Italian is a syllable-based language and the trochaic bias should be present only in infants learning stress-timed languages. In contrast, proposals that claim that the trochaic bias is universal, and do not take into consideration the role of different acoustic cues (Allen and Hawkins, 1978), would predict trochaic biases both for streams varying in pitch and for streams varying in duration. Alternatively, it is possible that the iambic-trochaic bias is present early in development. In this case, infants should segment streams alternating in pitch as formed by trochees, and streams alternating in duration as formed by iambs. Last, it is possible that perceptual biases relying on pitch and on duration emerge at different points in development.

2 Experiment 1

2.1 Subjects

Participants were 45 Italian adults, 15 in each experimental condition (duration: 7 males, mean age = 23.8, range = 21–29; pitch: 5 males, mean age = 22.4, range = 19–26; control: 8 males, mean age = 24.1, range = 22–27).

2.2 Stimuli

This experiment consisted of three different conditions. In the first condition (pitch condition), the syllables in the familiarization stream alternated in pitch; in the second condition (duration condition), syllables alternated in duration; and in the third condition (control condition), all syllables had the same pitch and duration (Figure 1).

Specifically, in the pitch condition, a stream with 10 adjacent syllables (/pa su tu ke ma vi bu go ne du/), separated by an interval of 100 ms, was concatenated and repeated 20 times. The first two syllables had their pitch set to 180 Hz and their phoneme duration set to 180 ms. Starting from the third syllable, the pitch of the central portion of the vowel of the odd syllables was increased in 20 Hz steps, until it reached and remained at 400 Hz. Therefore, the third syllable had a pitch of 200 Hz (180 + 20), the fourth 180 Hz, the fifth 220 Hz (200 + 20), the sixth 180 Hz, until the odd syllable peaked and remained at 400 Hz. This gradual increase in pitch was implemented to avoid participants forming bisyllabic groupings from the onset of the familiarization, as previous studies showed that rhythmic judgments might be based on the pattern of the first two stimuli of a sequence (Hay and Diehl, 2007).

In the duration condition, the familiarization stream was composed of the same stream of syllables and acoustic cues as in the pitch condition. In this condition, however, instead of an increase in pitch, the odd syllables had prominence marked by increasing the duration of their vowel, in 20 ms steps, up to 400 ms. In the control condition, the same syllables were used, but no marker of prominence was added. The values for pitch and duration were extrapolated from recordings made at SISSA Language and Development lab. These values were based on acoustic measurements of productions of child-directed iambs and trochees, are within the range of values used in previous experiments (Hay and Diehl, 2007), and fall within the range of values naturally found in speech (Crystal and House, 1990). The three familiarization streams were synthesized with the female voice of the MBROLA Italian database IT4 (Dutoit, Pagel, Pierret, Bataille, and Vreken, 1996).

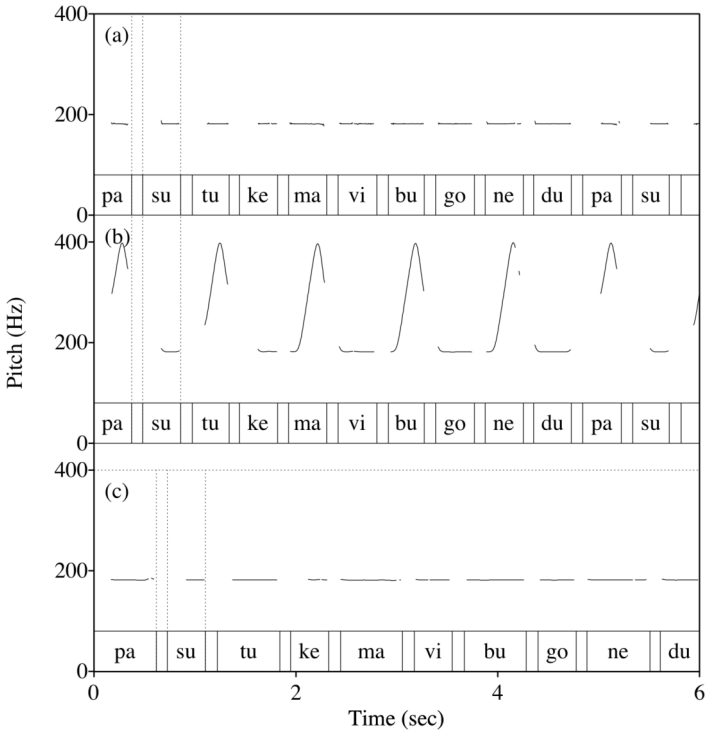

Figure 1. A segment from the middle of the familiarization stream with the three conditions of Experiment 1, (a) control condition, with phoneme duration of 180 ms and pitch of 180 Hz, (b) pitch condition, same as flat condition but with the pitch of the odd syllables peaking at 400 Hz at the midpoint of its vowel, (c) duration condition, same as control condition, but with the vowel of the odd syllables lasting 400 ms

For the test, three different sets of stimuli were synthesized by pairing syllables present in the familiarization. The first set of stimuli was created by pairing syllables that occurred adjacently and had prominence in their first syllable during familiarization (i.e., pa-su, tu-ke, ma-vi, bu-go, ne-du); the second set of stimuli was created by pairing syllables that occurred adjacently and had prominence on their second syllable during familiarization (i.e., su-tu, ke-ma, vi-bu, go-ne, du-pa); and the third by pairing non-adjacent syllables (i.e., pa-vi, vi-pa, su-bu, bu-su, tu-go, go-tu, ke-ne, ne-ke, ma-du, du-ma). The pairs of syllables that were synthesized were all spondees with pitch set to 180 Hz and phoneme duration set to 180 ms, and were the same for all participants. In this way, differences in performance could be attributed only to the differences in the familiarization streams.

2.3 Procedure

Participants were tested in a silent room and listened to the stimuli through headphones. After listening to one of the three possible familiarization streams, they were presented with the test stimuli, and were asked to judge whether they had heard these pairs of syllables adjacently during the familiarization phase or not. Each pair of syllables was presented twice, and a different random

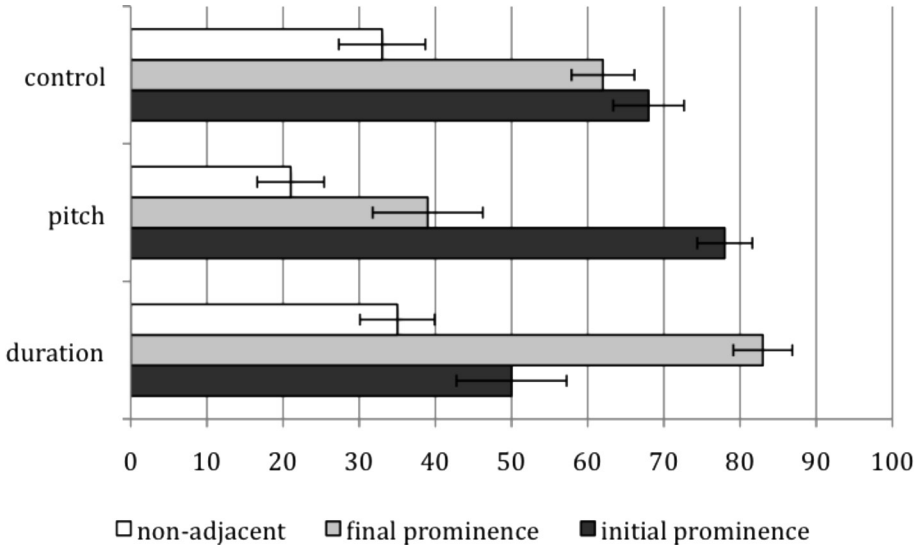

Figure 2. Percentage correct and standard error for the three experimental groups on the three sets of stimuli

order was created for each participant. The experiment was controlled by PsyScope X (Cohen, MacWhinney, Flatt, and Provost, 1993).

2.4 Results

In this experiment, participants in the pitch and in the duration conditions remembered different sets of stimuli during the test phase. An ANOVA indicates an interaction between familiarization (pitch, duration, and control) and condition (iamb, trochee, non-word) to be statistically significant, $F(1, 28) = 31.74$, $p < 0.001$, with no main effect of group, $F(1, 28) = 2.246$, $p = 0.145$, or test, $F(1, 28) = 0.225$, $p = 0.639$ (Figure 2). Specifically, participants in the pitch group were better at remembering pairs of syllables that had prominence in the first syllable during the familiarization than pairs of syllables that had prominence in their second syllables, $t(14) = 4.64$, $p < 0.001$. In contrast, participants in the duration group better remembered pairs of syllables with prominence in the second syllable, $t(14) = -3.426$, $p = 0.004$.

In order to confirm that these results were not caused by a particular property of the syllables chosen, the performance of the participants in the control group was analyzed. In this control condition, in which no acoustic marker of prominence was present during familiarization, no preference was observed for either set of stimuli, $t(14) = 0.84$, $p = 0.427$. In all three experimental groups, participants recalled pairs of syllables that occurred adjacently in familiarization significantly more often than pairs of non-adjacent syllables, $p < 0.05$.

These results strongly support our prediction that pairs of syllables alternating in pitch will be grouped with initial prominence (trochees), while sequences alternating in duration will be grouped with final prominence (iambs). Markers of prominence indicate the chunks over which the stream should be processed. In this experiment, instead of trying to memorize ten different syllables, participants were guided to memorize five adjacent pairs of syllables that followed the principles of

the iambic-trochaic law. This pre-processing of the stimuli might diminish cognitive load and be highly profitable in speech processing and language acquisition (Miller, 1956).

In order to investigate whether the same pattern of perceptual grouping would also be present earlier in development, Experiment 2 focused on 7-month-old infants' listening preferences.

3 Experiment 2

3.1 Subjects

Participants were 48 Italian infants, 16 in each experimental condition (duration condition: age range 7m 05d–8m 02d, mean age 7m 17d; pitch condition: age range 7m 01d–8m 03d, mean age 7m 16d; control condition: age range 7m 01d–8m 05d, mean age 7m 17d). Seventeen additional infants were tested but not included in the analyses (duration condition: 7 infants rejected, 5 fussy, and 2 for accumulated rejected trials;[1] pitch condition: 6 infants rejected, 4 fussy, 2 for accumulated rejected trials; control condition: 4 infants rejected, 3 fussy, 1 experimental error).

3.2 Stimuli

As in Experiment 1, three different conditions were investigated. Again, the difference between the conditions lay in the acoustic cue used to signal prominence during the familiarization phase. In the pitch condition, syllables alternated in pitch; in the duration condition, they alternated in duration; and in the control condition, all syllables had the same pitch and duration.

The same acoustic markers of prominence and stimuli from Experiment 1 were used in this experiment. However, instead of 10 adjacent syllables, the familiarization consisted of 6 adjacent syllables (/pa su tu ke ma vi/). The familiarization stream was three minutes long.

For the test, six different pairs of adjacent syllables were synthesized, comprising two different sets of stimuli. The first set of stimuli was created by pairing syllables that had prominence in their first syllable during the familiarization (i.e., pa-su, tu-ke, ma-vi), while the second set of stimuli was created by pairing syllables that had prominence in their second syllable (i.e., vi-pa, su-tu, ke-ma). In each trial, a pair of syllables (e.g., pa-su or su-tu) was repeated for 14 seconds, with an interval of 200 ms between pairs of syllables. This resulted in a total of six different trials, each repeated twice, resulting in a total of 12 test trials.

In order to gain statistical power for the comparison of the two main conditions (iambs vs. trochees), pairs of non-adjacent syllables were not included in this second experiment. The test stimuli were synthesized with the acoustic parameters from Experiment 1, they were spondees with all the syllables having the same pitch (180 Hz) and phoneme duration (180 ms). The same pairs of syllables were used in the three experimental conditions, thus any difference between the experiments can only be due to the effect of the different streams infants were exposed to during the familiarization phase.

3.3 Procedure

Throughout the experiment, the infant was seated on the lap of the caregiver on a chair in the center of a booth. In front of the infant, a green lamp was fixed. On the left and right sides of the booth, yellow lamps were fixed at the same height as the green lamp. Loudspeakers were placed behind the yellow lamps.

The experimenter was located outside the testing booth. By looking at movements of the infant projected onto a video screen, the experimenter could control the presentation of the acoustic stimuli, and the flashing of the yellow and green lamps. Both the experimenter and the caregiver were "deaf" in regards to the stimuli being played in the testing booth.

The design of the experiment followed a modification of the head-turn preference paradigm (Saffran, Johnson, Aslin, and Newport, 1999). During the familiarization phase, the infant listened to one of the three familiarization streams from the experiment (i.e., prominence marked by pitch, prominence marked by duration, and a control stream). While the infant listened to the stream, the lights on the sides or on the center of the testing booth flashed, contingently upon the infant's looking behavior, and independently of the sound being played.

During the test phase, the infant was confronted with 12 trials (each trial with each of the 6 words repeated for 14 seconds). The trials started with the center light flashing, in order to attract the infant's attention. Once the infant looked at the central light, it was extinguished, and the light on either the left or right side of the booth started flashing. When infants fixated on the side light (defined as a 30º head turn towards the light), the sound started playing from the speaker behind the corresponding light. If the infant looked away for more than two seconds, the trial was stopped, and another trial started. In cases where the infant held his/her head towards the presentation side constantly, the entire trial was presented. The maximum looking time per trial could be the sum of several head turns, or a single head turn. The experiment lasted around 4 minutes.

The side and the order of the presentation of the stimuli were counterbalanced within each of the three experimental conditions. Infants were videotaped during the experiment, and their looking behavior was coded off-line by "blind" coders.

3.4 Results

The total looking times for the words with initial and final prominence were calculated for each infant. Infants familiarized with the stream varying in pitch showed during the test phase a preference to listen to pairs of syllables that during familiarization had high pitch in the first syllable, $t(15) = 2.584$, $p = 0.021$. When infants were familiarized to the stream varying in duration, no preference for sequences with longer syllables in initial or final position was found, $t(15) = 1.417$, $p = 0.177$. In the control condition, in which all syllables had the same pitch and duration, no preference for any type of syllable was found either, $t(15) = -1.064$, $p = 0.304$. A repeated measures general linear model was fitted to the data, with prominence during familiarization (syllable initial and syllable final) as a within subject factor and condition (pitch, duration, and flat) as a between subject factor. The model reveals no main effect of condition, $F(2, 45) = 2.092$, $p = 0.135$, no main effect of prominence, $F(1, 45) = 2.276$, $p = 0.138$, but a significant interaction between prominence and condition, $F(2, 45) = 3.811$, $p = 0.030$ (Figure 3). Therefore, infants' preference for different sequences of syllables during the test phase (prominence) was influenced by the stream they heard during familiarization (condition).

In short, the results showed that a bias to group linguistic stimuli trochaically when prominence is marked by pitch is already present in pre-verbal infants. When infants were familiarized with a sequence alternating in pitch, during the test phase they showed a preference to listen to linguistic units that were previously heard with initial prominence. In other words, they preferred sequences that during familiarization were coherent with the claims of the 'iambic-trochaic' law.

In contrast, infants who were familiarized with the sequence alternating in duration or with constant duration and pitch did not show any preference. Importantly, the stimuli during the test phase were presented with no prominence, thus any difference in infants' preferences can only be due to the different familiarization phases.

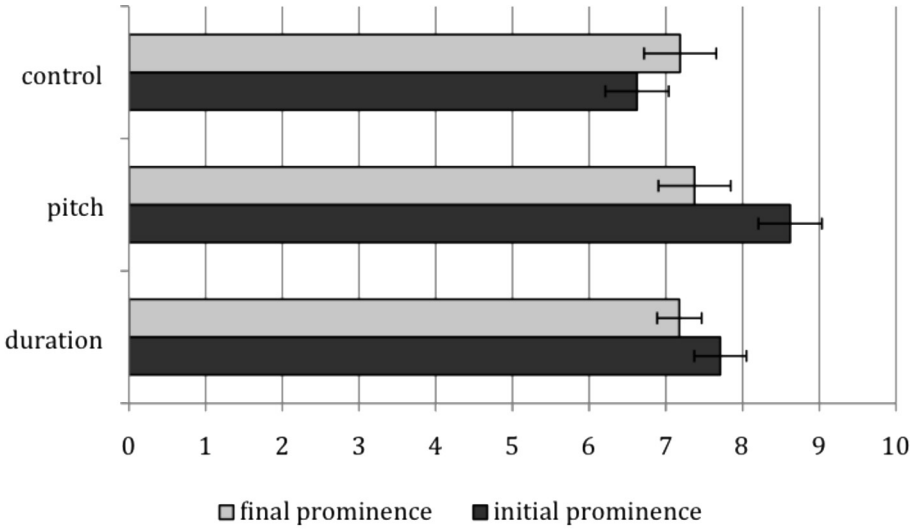

Figure 3. Mean looking times (seconds) and standard errors for the three experimental groups on the two sets of stimuli

4 Discussion and conclusions

In the two studies presented here, a similar set of stimuli was used to investigate adults' and infants' perceptual biases. Specifically, our studies focused on how prominence marked by either pitch or duration influenced infants' and adults' segmentation of speech sequences. The findings from the adult experiment were consistent with our predictions. Adults followed the principles of the iambic-trochaic law; they were better at remembering linguistic units that ended in a long element, or that started with a high-pitched element. These results add to the body of literature on the iambic-trochaic law, providing evidence that Italian speakers, like French and English speakers (Hay and Diehl, 2007), segment auditory streams following the iambic-trochaic law. This is also the first study to show that pitch, in addition to intensity, induces trochaic grouping in adults.

The findings from the infant study, in contrast, were not entirely consistent with the iambic-trochaic law. Infants showed trochaic biases for the stream varying in pitch, but showed no preferences for the stream varying in duration. It seems, therefore, that perceptual grouping based on pitch and perceptual grouping based on duration follow different developmental paths. Seven-month-old infants already show perceptual grouping similar to adults' for sequences alternating in pitch, but not for sequences alternating in duration.

Interestingly, these results support those of several other recent studies on the iambic-trochaic law. Studies with adults indicate that perceptual grouping by duration, instead of being guided uniquely by innate perceptual principles, might also be influenced by the properties of one's native language (Iversen, Patel, and Ohgushi, 2008). Iversen and colleagues (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, English listeners tended to place the long stimuli in final position (grouping the sequences as iambs), while the Japanese on average did not show a preference for either iambs or trochees.

A similar pattern of results was found by Yoshida and colleagues (2010) with Japanese- and English-learning infants. In this study, infants' listening preferences for sequences of alternating long and short tones (trochees) or short and long tones (iambs) were investigated. No listening preferences were found for either the Japanese- or the English-learning participants at 5 months of age. However, English-learning 8-month-olds preferred to listen to sequences of iambs, but the Japanese infants did not show a preference either way – similar to the findings reported by Iversen et al. (2008) with adults. These results are in agreement with our findings that segmentation biases based on duration emerge from language experience.

In yet another study, similar conclusions were reached (Hay and Saffran, 2008). In this study, the iambic-trochaic bias was contrasted with distributional cues, and 9-month-old infants relied more strongly on prosody (trochees signaled by intensity and iambs signaled by duration) than on transitional probabilities when segmenting a continuous speech stream or a sequence of tones. However, at 6 months of age, trochaic biases using intensity are present, but no bias for segmentation based on duration is observed. This extensive set of studies supports our hypothesis that grouping biases based on duration emerge from language experience and are not yet present at 7 months of age, but emerge around 9 months of age.

However, we found trochaic biases for the pitch condition, and these results have implications for the different proposals put forward to explain early trochaic biases in infancy. The rhythmic-activation proposal claims that trochaic biases should be present in infants learning stress-timed languages such as English or Dutch, while no bias should be present in infants learning syllable-timed languages such as French or Italian (Nazzi et al., 2006). The existence of a trochaic bias in our pitch condition for Italian infants seems to be problematic for this proposal. It is possible therefore that the lack of a trochaic bias in French-learning infants might be due to the acoustic characteristics of the stimuli used in the different experiments (Höhle et al., 2009), in which both duration and pitch co-occur. Those cues often co-occur in natural languages, but owing to their different weights in different languages, descriptions of infants' perceptual biases should be more specific at the acoustic level.

Our results also do not support a proposal for innate trochaic bias without taking into account the acoustic features of the stimuli (Allen and Hawkins, 1978), or the existence of an innate iambic-trochaic bias. This is due to the fact that trochaic biases were observed in our experiment only for the stream alternating in pitch, while no iambic bias was present for the stream alternating in duration. It therefore appears that different developmental paths are present for grouping biases based on duration and pitch, with the former emerging later in development.

Developmental changes in the weighting of pitch and duration have been observed in other domains of linguistic processing. For instance, Seidl and colleagues (Seidl and Cristià, 2008) found that 4-month-old infants weight acoustic cues to signal phrasal boundaries differently from 6-month-olds. The authors found that at 4 months of age, pitch, duration, and intensity are used in conjunction for the segmentation of clauses, while at 6 months of age, pitch is used as the main cue for segmentation (Seidl, 2007). Importantly, these developmental changes were not captured by previous studies that failed to manipulate pitch and duration separately (Nazzi, Kemler Nelson, Jusczyk, and Jusczyk, 2000). Acoustic cues are also weighted differently in child-directed speech according to the age of the infant being talked to (Kitamura and Burnham, 2003), and the presence of different acoustic cues in infant-directed speech also influences infants' listening preferences differently throughout development (Panneton, Kitamura, Mattock, and Burnham, 2006). For instance, slower speech enhances 4-month-olds' attention to infant-directed speech, while it does not influence 8-month-olds' listening preferences (Panneton et al., 2006). Language experience also affects the weighting of cues such as duration at the word level, with sensitivity to vowel

duration being highly dependent on the phonological weight of this cue in the infants' native language (Dietrich, Swingley, and Werker, 2007; Mugitani, Pons, Fais, Werker, and Amano, 2008). Further studies should investigate how attention to acoustic cues at different levels of linguistic processing changes throughout infancy, and how acoustic information such as pitch and duration at the segmental, word stress, clausal, and affective and pragmatic level are kept apart or interact to influence infants' early perceptual biases. Interestingly, the different weighting of different acoustic cues throughout infancy might in itself signal the linguistic units that warrant infants' attention at their current developmental stage.

In addition to different weighting of acoustic cues, it is also possible that perceptual grouping reflects infants' sensory abilities. Previous studies have shown that 7-month-old infants show pitch perception abilities similar to adults', being able to discriminate harmonic complexes from pitch categories differing by only 20% (e.g., 160 vs. 200 Hz, Clarkson and Clifton, 1985). Other studies seem to indicate that infants' abilities to discriminate variations in pitch are superior to adults' in some cases (Trainor and Trehub, 1992). It seems that the ability to detect variations of pitch is already present at birth (Nazzi, Floccia, and Bertoncini, 1998). The ability to discriminate changes in duration, however, seems to only be reached between the ages of 8 and 10 years (Elfenbein, Small, and Davis, 1993).

The question that remains is whether this early bias for grouping based on pitch observed in infants can still help them to learn about language-specific word order. We believe that it can: if the prominent element of phrases in the language is marked by pitch, infants can assume that the language is head-final; otherwise, that it is head-initial. This logic might be enough for infants to discover the ordering of words within phrases in their language, and adding another acoustic cue to signal word order might appear redundant. This proposal should be appealing to both generative and statistical approaches to language acquisition. Generative approaches might see this correlation either as a way to set universal switches during language acquisition or as a way to deviate from the default value just in case a signal indicates that it is necessary to do so. In a parameter-setting perspective, it is possible to argue that the parameter for head-directionality is initially set to head-initial, only changed to head-final in cases where prominence in the language is marked by duration. Statistical approaches, in contrast, might see this correlation as a way to organize the input into linguistically relevant units, or as a probabilistic cue aiding the acquisition of syntax. This bias might be directed towards universal acoustic cues, or it might instead be a general ability to develop sensitivity to the particular cues in the language being learned.

These perceptual biases can in turn be combined with other cues in the learning of language-specific word order. The unreliability of prosodic cues is not a problem if prosodic information is supplemented with other types of phonological cues and with distributional information (Christiansen, Allen, and Seidenberg, 1998; Hirsh-Pasek, Tucker, and Golinkoff, 1996). Recent studies have shown that infants' perceptual grouping is also influenced by the frequency of occurrence of the alternating elements in a stream, which also correlates with head-directionality within and across languages (Gervain, Nespor, Mazuka, Horie, and Mehler, 2008). All these cues, in combination with different perceptual biases, such as attention to adjacent repetitions, and biases for paying attention to the edges of linguistic constituents (Endress, Dehaene-Lambertz, and Mehler, 2007; Endress and Mehler, 2005; Gervain, Macagno, Cogoi, Peña, and Mehler, 2008) might be used in conjunction to discover language-specific word order regularities.

Independently of its use for syntactic bootstrapping, these biases for grouping of linguistic stimuli might be beneficial to both infants and adults. Humans have been shown to have a limited processing capacity (around seven chunks, Miller, 1956; or even fewer, Cowan, 2001), thus segmenting our sequences as three pairs of syllables instead of six individual syllables might decrease

the processing load. Specifically, interpreting the sequence /pa/ /su/ /tu/ /ke/ /ma/ /vi/ as /pasu/ /tuke/ /mavi/ might improve memorization in a similar way as does segmenting the sequence BBC-NBCCNN as three names of broadcasters (BBC, NBC, CNN). It has indeed been found that variations in pitch induce perceptual grouping of long speech sequences, such as phone numbers, and enhance memory performance (Frankish, 1995). Even pre-verbal untrained infants use chunking to increase their memory (Feigenson and Halberda, 2008). Importantly, these iambic-trochaic biases allow acoustic events to be segmented instantly without any previous experience with the stream, differently from classical statistical computations, which are highly experience dependent (Saffran, Aslin, and Newport, 1996). Once these events are found, further structural regularities can start to be extracted from the signal (Peña, Bonatti, Nespor, and Mehler, 2002).

In our pitch condition, infants are clearly segmenting the stream as three bisyllabic words, instead of into six individual syllables. In the duration condition, however, it is not clear what are the exact units infants are extracting from the stream. Infants might be extracting syllables, as a syllable-based segmentation strategy would propose, or bigger chunks of two, three, up to even six syllables. They are not segmenting the stream into three adjacent bisyllabic units, as this would probably result in a listening preference when those units are paired against units spanning a word boundary. A series of separate experiments would be necessary to distinguish the consistent units, if any, infants might be extracting from the stream. Our experiments show an asymmetry, at 7 months of age, in the segmentation of a speech stream when it alternates in duration or pitch, whereas adults show trochaic biases in a stream alternating in pitch, and iambic biases in a stream alternating in duration. Infants, in contrast, show perceptual biases only for bisyllabic segmentation in a stream alternating in pitch. The exact nature of the extraction of chunks from the stream alternating in duration cannot be identified at present, as it cannot be predicted from the iambic-trochaic law alone.

This study is the first step in investigating how different acoustic cues influence iambic and trochaic biases in very young infants. Further studies should use different syllable durations and pitch contours in the infant experiment. Specifically, while the increase in duration present in the adult experiment was enough to trigger iambic segmentation, longer durations than those we used might be necessary for the iambic bias to emerge in infants. In addition, it is not clear up to what extent variability at the segmental level influences the reliance on segmentation cues on the suprasegmental level. Most studies testing the iambic-trochaic law have focused on repetitions of a single isolated word, syllable, or tone. Hay and Diehl (2007), for instance, tested adults on several repetitions of the syllable /ga/ with variations in duration or intensity. Studies with infants used simplified streams as well. Trainor and Adams (2000) have used a single tone with variation in either intensity or duration in their stimuli, Yoshida et al. (2010) used a single tone with variation in duration, and Höhle and colleagues (2009) used several repetitions of the word /gaba/ with stress in the first or last syllable. In all these experiments, despite the simplicity of the stream, segmentation biases were observed. The hypothesis that less variation makes infants more sensitive to prosodic regularities is supported by studies showing that newborns are able to discriminate stress patterns when vowels do not change, but fail to do so when they vary (Sansavini, Bertoncini, and Giovanelli, 1997; Van Ooijen, Bertoncini, Sansavini, and Mehler, 1997).

Alternatively, it is possible that variability at the segmental level could help infants pay attention to prosodic cues. It has been shown that increased variability in a speech stream makes infants better at learning invariant structures, such as non-adjacent rules (Gómez, 2002). However, in our experiment, infants do not have to learn rules, but rather memorize the syllables in a speech stream. In other words, infants have to pay attention to both words and prosodic cues in order to extract

smaller sized units. Most studies with infants at the same age as our group use no more than three different words in their familiarization stream (Saffran et al., 1996), exactly the number of words we used in our experiment. Therefore, considering infants' limited short-term memory (Feigenson and Halberda, 2008), increasing the number of words in our stream might result in failed segmentation, even if the invariant prosodic cues become more salient.

The present study adds to a current debate on the origins of trochaic biases observed in infants from different language backgrounds. It is pointed out that trochaic and iambic biases are highly dependent on the acoustic cues present in the speech stream, with iambic biases emerging from prominence marked by duration and trochaic biases emerging from prominence marked by pitch. The exact emergence of these two biases, however, was shown to follow different developmental pathways. Grouping biases based on pitch seem to emerge earlier in development than grouping biases based on duration. It is possible that perceptual grouping based on duration emerges with language experience, or instead depends on perceptual maturation, or attention to particular linguistic cues. This study sheds light on how acoustic markers of prominence might influence segmentation and memory for speech sequences by infants and adults. The exact way different perceptual mechanisms and biases interact and develop in early infancy, and help infants discover words, phrases, and the relation among these units in the apparently continuous speech stream is still an open question.

Acknowledgements

This work was supported by a grant to the second author from MICIT (Ministerio de Ciencia y Tecnología), CONICIT (Consejo Nacional para Investigaciones Científicas y Tecnológicas) of Costa Rica. We thank Marijana Sjekloća for help with recruiting and testing participants.

Note

1 A trial was rejected if looking time was less than 1 second. An infant was rejected if he had more than three rejected trials.

References

Allen, G., and Hawkins, S. (1978). The development of phonological rhythm. In A. Bell and J. Hooper (Eds.), *Syllables and segments* (pp. 173–185). Amsterdsam: North-Holland.

Bion, R. A. H., Höhle, B., and Schmitz, M. (2007). The role of prosody on the perception of word-order differences by 14-month-old German infants. *Proceedings of ICPhS 2007*, 1537–1540.

Bleakley, D. (1973). The effect of fundamental frequency variations on the perception of stress in German. *Phonetica*, 28, 42–59.

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

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

Christophe, A., Gout, A., Peperkamp, S., and Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, 31, 585–598.

Christophe, A., Guasti, M. T., Nespor, M., and Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-complement parameter. *Developmental Science*, 6, 213–222.

Clarkson, M. G., and Clifton, R. K. (1985). Infant pitch perception: Evidence for responding to pitch categories and the missing fundamental. *Journal of the Acoustical Society of America*, 77, 1521–1528.

Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, 25(2), 257–271.

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

Cooper, W. E., and Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.
Crystal, T., and House, A. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, 88, 101–112.
Cutler, A. (2005). Lexical stress. In D. B. Pisoni and R. E. Remez (Eds.), *The handbook of speech perception* (pp. 264–289). Oxford: Blackwell.
Cutler, A., Mehler, J., Norris, D., and Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4), 385–400.
Cutler, A., and Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 113–121.
De Casper, A. J., and Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208, 1174–1176.
Dietrich, C., Swingley, D., and Werker, J. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences of the USA*, 104, 454–464.
Dogil, G. (1999). The phonetic manifestation of word stress in Lithuanian, Polish and German and Spanish. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 273–311). Berlin: Mouton de Gruyter.
Dupoux, E., Pallier, C., Sebastián-Gallés, N., and Mehler, J. (1997). A distressing "deafness" in French. *Journal of Memory and Language*, 36, 406–442.
Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and Vreken, O. (1996). The Mbrola project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In H. T. Bunell and W. Isardi (Eds.), *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1393–1396). Wilmington, DE: Alfred I. duPont Institute.
Elfenbein, J. L., Small, A. M., and Davis, J. M. (1993). Developmental patterns of duration discrimination. *Journal of Speech and Hearing Research*, 36, 842–849.
Endress, A., Dehaene-Lambertz, G., and Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, 105(3), 577–614.
Endress, A., and Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General*, 134(3), 406–419.
Feigenson, L., and 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.
Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology*, 9(7), S5–S22.
Friederici, A., Friedrich, M., and Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, 17, 1208–1211.
Gervain, J., Macagno, F., Cogoi, S., Peña, M., and Mehler, J. (2008). The neonate brain detects speech structure. *PNAS*, 105(37), 14222–14227.
Gervain, J., Nespor, M., Mazuka, R., Horie, R., and Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese–Italian crosslinguistic study. *Cognitive Psychology*, 57, 56–74.
Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, 15, 103–113.
Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, 5, 431–436.
Gout, A., Christophe, A., and Morgan, J. (2004). Phonological phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory and Language*, 51, 547–567.
Hay, J., and Diehl, R. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics*, 69(1), 113–122.
Hay, J., and Saffran, J. (2008). Perceptual constraints on segmentation of speech and non-speech. Paper presented at the 33rd BUCLD, October 31–November 2, 2008, Boston, MA.
Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago, IL: University of Chicago Press.

Hirsh-Pasek, K., Tucker, M., and Golinkoff, R. (1996). Dynamic systems theory: Reinterpreting "Prosodic Bootstrapping" and its role in language acquisition. In J. Morgan and K. Demuth (Eds.), *Signal to syntax* (pp. 453–465). Mahwah, NJ: Erlbaum.
Höhle, B. (2002). *Der Einstieg in die Grammatik: Die Rolle der Phonologie/Syntax-Schnittstelle für Sprachverarbeitung und Spracherwerb*. Freie Universität Berlin.
Höhle, B., Bijeljac-Babicb, R., Weissenborn, J., and 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.
Houston, D., Jusczyk, P., Kuijpers, C., Coolen, R., and Cutler, A. (2000). Crosslanguage word segmentation by 9-month-olds. *Psychonomic Bulletin* and *Review*, 7, 504–509.
Iversen, J., Patel, A., and Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, 124(4), 2263–2271.
Johnson, E., and Seidl, A. (2008). A cross-linguistic perspective on the detection of juncture in speech. *Infancy*, 13(5), 440–455.
Jusczyk, P., Cutler, A., and Redanz, L. (1993). Preference for the predominant stress patterns of English words. *Child Development*, 64, 675–687.
Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., and Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24, 252–293.
Jusczyk, P., and Krumhansl, C. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3), 627–640.
Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, 4(1), 85–110.
Kurby, C., and Zacks, J. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12, 72–79.
Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, 33, 451–454.
Llisterri, J., Machuca, M., de la Mota, C., Riera, M., and Rios, A. (2003). The perception of lexical stress in Spanish. In M.-J. Solé, D. Recasens, and J. Romero (Eds.), *Proceedings of the XVth International Conference of Phonetic Sciences* (Vol. 2, pp. 2023–2026). Barcelona: Causal Productions.
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.
Morgan, J. L., and Demuth, K. D. (Eds.). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
Mugitani, R., Pons, F., Fais, L., Werker, J., and Amano, S. (2008). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology*, 45(1), 236–247.
Nakatani, L. H., and Aston, C. H. (1978). Perceiving stress patterns of words in sentences. *Journal of the Acoustical Society of America*, 63(supplement 1), S55(A).
Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1–11.
Nazzi, T., Floccia, C., and Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, 21(4), 779–784.
Nazzi, T., Iakimova, I., Bertoncini, J., Frédonie, S., and Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic difference. *Journal of Memory and Language*, 54, 283–299.
Nazzi, T., Kemler Nelson, D., Jusczyk, P., and Jusczyk, A. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, 1(1), 123–147.
Nespor, M., Guasti, M. T., and 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., and Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio*, No. 2, 139–168.

Nespor, M., and Vogel, I. (2008). *Prosodic phonology*. Berlin: Mouton de Gruyter. (1st edition 1986. Dordrecht: Foris.)
Panneton, R., Kitamura, C., Mattock, K., and Burnham, D. (2006). Slow speech enhances younger but not older infants' perception of vocal emotion. *Research in Human Development*, 3(1), 7–19.
Peña, M., Bonatti, L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593), 604–607.
Potisuk, S., Gandour, J., and Harper, M. P. (1996). Acoustic correlates of stress in Thai. *Phonetica*, 53, 200–220.
Ramus, F., Hauser, M., Miller, C., Morris, D., and Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288, 340–351.
Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science*, 274, 1926–1928.
Saffran, J., Johnson, E., Aslin, R., and Newport, E. (1999). Statistical learning of tonal sequences by human infants and adults. *Cognition*, 70, 27–52.
Sansavini, A., Bertoncini, J., and Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, 33, 3–11.
Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, 57, 24–48.
Seidl, A., and Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, 11(4), 596–606.
Singh, L., White, K. S., and 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.
Swallow, K., and Zacks, J. (2008). Sequences learned without awareness can orient attention during the perception of human activity. *Psychonomic Bulletin & Review*, 15(1), 116–122.
Swallow, K., Zacks, J., and Abrams, R. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, 138, 236–257.
Thiessen, E., and Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, 39, 706–716.
Trainor, L., and Adams, B. (2000). Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception and Psychophysics*, 62(2), 333–340.
Trainor, L. J., and Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 394–402.
Van Ooijen, B., Bertoncini, J., Sansavini, A., and Mehler, J. (1997). Do weak syllables count for newborns? *Journal of the Acoustical Society of America*, 102, 3735–3741.
Weissenborn, J., and Höhle, B. (Eds.). (2001a). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (Vol. 1). Amsterdam: John Benjamins.
Weissenborn, J., and Höhle, B. (Eds.). (2001b). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (Vol. 2). Amsterdam: John Benjamins.
Williams, B. (1985). Pitch and duration in Welsh stress perception: The implications for intonation. *Journal of Phonetics*, 13, 381–406.
Woodrow, H. (1951). Time perception. In 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., and Werker, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese–English cross-linguistic study. *Cognition*, 115(2), 356–361.
Zacks, J., and Swallow, K. (2007). Event segmentation. *Current Directions in Psychological Science*, 16, 80–84.