Journal of Memory and Language 66 (2012) 285–306

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Can prosody be used to discover hierarchical structure in continuous speech?

Alan Langus [a,*], Erika Marchetto [b], Ricardo Augusto Hoffmann Bion [c], Marina Nespor [d]

[a] SISSA/ISAS – International School for Advanced Studies, Cognitive Neuroscience Sector, Language, Cognition and Development Laboratory, Trieste, Italy
[b] Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP), Ecole Normale Supérieure, Paris, France
[c] Stanford University, Psychology Department, Center for Infant Studies, USA
[d] University of Milano-Bicocca, Psychology Department, Milan, Italy

ARTICLE INFO

Article history:
Received 22 July 2010
revision received 16 September 2011
Available online 19 October 2011

Keywords:
Prosody
Prosodic hierarchy
Phonological Phrase
Intonational Phrase
Artificial grammar learning
Language acquisition

ABSTRACT

We tested whether adult listeners can simultaneously keep track of variations in pitch and syllable duration in order to segment continuous speech into phrases and group these phrases into sentences. The speech stream was constructed so that prosodic cues signaled hierarchical structures (i.e., phrases embedded within sentences) and non-adjacent relations (i.e., AxB rules within phrases), while transitional probabilities between syllables favored adjacent dependencies that straddled phrase and sentence boundaries. In Experiments 1 and 2, participants hierarchically segmented the stream and learned the grammar used to generate the phrases when prosodic cues were consistent with their native language. In Experiment 3, participants segmented the stream based on transitional probabilities when no prosodic cues were present and all syllables had the same pitch and duration. In Experiment 4, participants were able to exploit non-native prosody in order to learn hierarchical relations and non-adjacent dependencies. These results suggest that prosodic cues such as pitch declination and final lengthening provide a stronger basis for learning than transitional probabilities even when both are unfamiliar and not wholly consistent with native language informational structure.

Introduction

In order to understand language, it is necessary to process its hierarchical structure. Sentences often contain more than one phrase, phrases more than one word, words more than one morpheme. Adult speakers apply generative rules at each level of this hierarchy, thus producing from a finite number of morphemes and words, an indefinite number of phrases and sentences. However, it is not clear how language learners manage to keep track of these different levels of linguistic processing when interpreting spoken language. In this paper, we investigate the possibility that at least part of the human ability to hierarchically organize words into phrases, and phrases into sentences, can be acquired from prosody, that is, by tuning to the acoustic properties of the speech signal.

The prosody of speech is characterized by changes in duration, intensity and pitch (Cutler, Dahan, & van Donselaar, 1997; Lehiste, 1970). Speakers can intentionally manipulate these acoustic cues to convey information about their emotional states (e.g. irony or sarcasm), to define the type of statement they are making (e.g. a question or a statement), and to highlight certain elements over others (e.g. by contrasting them). These acoustic cues can also be combined in different ways across languages to contrast the meaning of words (e.g. word stress is phonemic in Italian, e.g. *méta* 'aim', *metà* 'half'; phoneme length is phonemic in Estonian, e.g. *ma* 'I', *maa* 'land'; pitch is phonemic in tonal languages like Mandarin, e.g. *mà* 'mumbler', *má* 'hemp').

* Corresponding author. Address: SISSA/ISAS – International School for Advanced Studies, via Beirut 2-4, 34014 Trieste, Italy. Fax: +39 040 3787 615.
E-mail address: alanlangus@gmail.com (A. Langus).

0749-596X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2011.09.004

Importantly, prosody also contains information about the syntactic structure of languages. Because syntactic structure is automatically mapped onto prosodic structure during speech production, variations in duration, intensity and pitch, are also systematically related to the hierarchical structure of syntax (Nespor & Vogel, 1986; Selkirk, 1984; Nespor et al., 2008). There is not a one-to-one mapping between prosody and syntax, but at least some aspects of syntactic information are deducible from prosodic contours.

The prosodic hierarchy

Just like syntax, phrasal prosody is structured hierarchically, and its constituents go from the prosodic word up to the utterance (e.g. Beckman & Pierrehumbert, 1986; Hayes, 1989; Nespor & Vogel, 1986; Selkirk, 1984). The different levels of the prosodic hierarchy are organized so that lower levels are exhaustively contained into higher ones (e.g. Selkirk, 1984). This is best exemplified by considering the two prosodic constituents most relevant for the present paper: the Phonological Phrase and the Intonational Phrase. The Phonological Phrase extends from the left edge of a phrase to the right edge of its head in head-complement languages; and from the left edge of a head to the left edge of its phrase in complement-head languages (Nespor & Vogel, 1986). The constituent that immediately dominates the Phonological Phrase is the Intonational Phrase – a more variable constituent as to its domain – that is coextensive with intonation contours, thus accounting for natural break points in speech (Pierrehumbert & Hirschberg, 1990). While the number of Phonological Phrases contained in an Intonational Phrase may vary, Phonological Phrases never straddle Intonational Phrase boundaries: Phonological Phrases are exhaustively contained in Intonational Phrases.

While syntactic and prosodic constituent edges often coincide – e.g. all phrasal syntactic boundaries are signaled by prosodic boundaries – there is no one-to-one correspondence between the two (c.f. Steedmna, 1990; Hirst, 1993; Inkelas & Zec, 1990; Shattuck-Hufnagel & Turk, 1996; see Cutler et al., 1997 for an overview). On the one hand, the prosodic hierarchy is flatter than the syntactic hierarchy (Nespor & Vogel, 1986; Selkirk, 1984), in that there are fewer different levels in prosody than there are in syntax (Cutler et al., 1997), and prosodic structure is not recursive (Nespor & Vogel, 1986; Selkirk, 1984). On the other hand, while Phonological Phrases signal syntactic phrases consistently, their boundaries are also found within syntactic phrases. One well-known case is that of optional Phonological Phrase restructuring, where the complement or modifier of a head can be included in its same Phonological Phrase if non-branching. There is thus an asymmetry, with respect to syntax, in [I export rabbits] vs. [I export] [Australian rabbits]. In addition, while both edges of a sentence are always aligned with the edges of an Intonational Phrase, an Intonational Phrase boundary can occasionally be found within a sentence, and its location is less predictable, e.g. in case a sentence is very long and uttered at a normal rate of speech. For example a sentence as *All the children of the many friends of Guinevere will get together to sing in the choir for the end of the year* may be restructured into two Intonational Phrases in at least two different places – e.g. with a break either after *Guinevere*, or after *together*.

However, because syntactic structure is automatically mapped onto prosodic structure during speech production (Nespor & Vogel, 1986), many prosodic cues do signal syntactic boundaries. The boundaries of major prosodic units are associated with acoustic cues like final lengthening and pitch reset and decline (c.f. Beckman & Pierrehumbert, 1986; Cooper & Paccia-Cooper, 1980; Klatt, 1976; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). These cues are organized so that different levels of the prosodic hierarchy use at least partially different prosodic cues. For example, among the most common cues for Phonological Phrase boundaries is final lengthening. The strongest cues for Intonational Phrase-boundaries are pitch resetting at the left edge, and the declining pitch contour at the right edge (Pierrehumbert, 1979; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Wightman et al., 1992). Importantly, the largest variations in pitch and duration, typical of boundaries of prosodic constituents, most often coincide with edges of syntactic constituents cross-linguistically (Cooper & Sorensen, 1981; O'Shaughnessy, 1979; Vaissiere, 1974, 1975).

While speakers automatically map syntactic structure onto prosodic contours, there is also evidence that listeners are sensitive to the prosodic cues that signal syntactic constituents. It has in fact been found that listeners can locate major syntactic boundaries in the speech stream by relying on prosody alone (Collier & 't Hart, 1975; de Rooij, 1975, 1976; Collier, de Pijper, & Sanderman, 1993; Schafer, Speer, Warren, & White, 2000; Speer, Warren, & Schafer, 2011). Final lengthening in Phonological Phrases (Lehiste, 1973; Lehiste, Olive, & Streeter, 1976; Nooteboom, Brokx, & de Rooij, 1978; Scott, 1982; Vaissière, 1983) and the declining pitch contour that characterizes Intonational Phrases (Beach, 1991; Cooper & Sorensen, 1977; Streeter, 1978; Wightman et al., 1992) are the most reliable cues for segmenting continuous speech into syntactic constituents cross-linguistically (Cutler et al., 1997; Fernald & McRoberts, 1995). For example, adults use final lengthening to segment artificial speech streams (Bagou, Fougeron, & Frauenfelder, 2002) and can use both Phonological Phrase and Prosodic Word boundaries to constrain lexical access (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Marslen-Wilson & Tyler, 1980; Millotte, Rene, Wales, & Christophe, 2008). Furthermore, syntactic ambiguities can be disambiguated by prosodic boundaries (e.g., [bad] [boys and girls] vs. [bad boys] [and girls]) (Lehiste, 1974; Lehiste et al., 1976; Price et al., 1991; Streeter, 1978), clearly showing that syntax imposes constraints on prosodic structure (c.f. Shattuck-Hufnagel & Turk, 1996).

While prosodic cues thus appear to signal syntactic constituency in fluent speech, there is, to our knowledge, no experimental evidence that participants can simultaneously keep track of distinct prosodic cues from different levels of the prosodic hierarchy. Therefore, the first question the experiments in this paper address is whether listeners view prosody as hierarchically structured, and assign the different cues – i.e. duration and pitch declination

– to specific levels of the prosodic hierarchy – Phonological Phrase and Intonational Phrase, respectively.

The two roles of prosody

The possibility that language learners see prosodic cues as hierarchically structured poses the question of how they make use of them. Experimental evidence from language acquisition suggests that prosody has two primary roles in breaking the continuous speech stream.

On the one hand, infants use prosodic cues to segment speech (c.f. Jusczyk, 1998). Infants can discriminate pitch change by 1–2 months of age (Kuhl & Miller, 1982; Morse, 1972). By 4.5 months, infants prefer passages with artificial pauses inserted at clause boundaries rather than other places in the sentence (Hirsh-Pasek et al., 1987; Jusczyk, Hohne, & Mandel, 1995; Kemler Nelson et al., 1995; Morgan, Swingley, & Miritai, 1993). At 6 months, infants are able to use prosodic information consistent with clausal units (Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000), and also demonstrate some sensitivity to prosodic information consistent with phrasal units (Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). At 9 months, infants show a preference for passages with pauses coincident with phrase boundaries over passages where the pauses are inserted elsewhere in the sentence (Jusczyk et al., 1992). By 13 months of age, infants can use Phonological Phrase boundaries to constrain lexical access (Gout, Christophe, & Morgan, 2004). In sum, the sensitivity to cues carried by prosody appears to emerge within the first year of life.

On the other hand, there is also some evidence that language learners may be able to use variation in pitch and duration for grouping speech sequences into prosodic constituents, and thus likely also into syntactic constituents. One way in which pitch and duration guide listeners in grouping sound sequences is expressed by the Iambic-Trochaic law. The Iambic-Trochaic law (ITL) was first proposed to account for auditory perception in music (Bolton, 1894, Woodrow, 1951, Cooper & Meyer, 1960): while intensity mainly characterizes prominence in trochaic groupings (i.e. sequences with the most intense element in initial position), duration mainly characterizes prominence in iambic groupings (i.e. sequences with the longest element in final position). For language, the ITL was first proposed to account for the different realization of word internal secondary stresses. Within the metrical (or prosodic) constituents known as feet, elements that alternate mainly in duration are grouped iambically (weak–strong i.e. in this specific case, short–long), and elements that alternate mainly in intensity are grouped trochaically (strong–weak i.e. in this specific case, high–low) (Hay & Diehl, 2007; Hayes, 1995).

Following the proposal of Nespor et al. (2008) that in language, trochaic grouping is signaled also by pitch, Bion, Benavides, and Nespor (2011) showed that 7-month-old Italian infants segment sequences of syllables alternating in pitch into trochaic words with high pitch on the first syllable. The trochaic preference for sequences alternating in high and low pitch has also been found with English infants (Jusczyk, Cutler, & Redanz, 1993; Thiessen & Saffran, 2003); and both types of preference – depending on the familiarization stream – was found with 7-month-old English bilinguals with Japanese, Hindi, Punjabi, Korean or Farsi as the other language (Gervain & Werker, 2008). While age and linguistic environment appear to play a crucial role in the development of these grouping preferences (c.f. Yoshida et al., 2010), infants ability to use pitch and duration cues for segmenting and consequently discovering constituents from continuous speech appears to emerge during the first year of life. As discussed in Bion et al. (2011) and in Peña, Bion, and Nespor (2011), perceptual grouping based on pitch and duration follow different developmental trajectories. Grouping biases based on pitch seem to emerge very early in development, possibly reflecting universal abilities, while perceptual grouping based on duration either emerges with language experience or depends on perceptual maturation.

Research on the ITL suggests that prosody may help infants in language acquisition by signaling syntactic relations in speech. Nespor et al. (2008) argue that the ITL might serve as a cue to language-specific word order because Phonological Phrase prominence is signaled mainly through pitch and intensity in complement-head languages (e.g. Japanese and Turkish), where it is in initial position, and mainly through duration in head-complement languages (e.g. English and Italian), where it is in final position. That is, in a pair of words consisting of a complement and a head, the prominence always falls on the complement, but it is differently realized mainly by an increase in pitch or in duration depending on whether the complement precedes or follows its head. Thus, although prosodic cues to the actual syntactic constituents are to some extent probabilistic – e.g. given the possibility of restructuring mentioned above – prosodic cues also offer more abstractly non probabilistic information about the syntactic structure of a language: an iambic rhythm in phrasal stresses signals that complements follow heads; a trochaic rhythm signals that heads follow complements (Nespor et al., 2008).

As already mentioned, there is no one-to-one correspondence between prosodic structure and syntactic structure, but researchers assume that infants rely on prosodic cues in order to understand some hierarchical syntactic relations in the speech stream (see Gerken, 1996a, 1996b for a discussion). For example, Gerken, Jusczyk, and Mandel (1994) presented 9-month-old infants sentences where: (A) prosodic boundaries cued major syntactic boundaries (e.g. in *My neighbor/never walks his dog*, where a Phonological Phrase boundary signals the major syntactic boundary between the Subject Noun Phrase and the Verb Phrase), and (B) prosodic boundaries fail to cue major syntactic boundaries (e.g. in *He never walks his dog*, where there is no Phonological Phrase boundary between the pronominal Subject and the Verb Phrase even though this corresponds to a major syntactic boundary). Importantly, Gerken et al. (1994) inserted pauses either after the Subject (a major syntactic boundary) or after the verb (a minor syntactic boundary). Nine-month-old infants showed longer looking times to sentences where pauses were inserted at major syntactic boundaries that were signaled by prosodic boundaries (sentences in A), compared to sentences where pauses were inserted at minor syntactic boundaries.

However, there were no differences when the two syntactic boundaries were not signaled by prosodic boundaries (sentences in B). This suggests that as old as 9-months, infants assign boundaries according to prosodic cues, and fail to detect certain major syntactic boundaries. This tendency to follow prosodic constituents over syntactic ones, which is also observed in children's language production, appears to continue well into the third year of life (cf. Demuth, 1995; Fee, 1995; Gerken, 1994; Gerken, 1996a, 1996b). Thus these results suggest that, in early acquisition, the prosodic hierarchy may be an important cue for understanding the hierarchical structure of the speech stream – even if the prosodic grouping principles do not perfectly predict the underlying syntactic structure.

On the one hand, prosody signals breaks in the speech stream, providing the listener with the edges of individual constituents. For example, phrasal prosodic constituents can be exhaustively parsed into a sequence of non-overlapping words (e.g., Nespor & Vogel, 1986; Selkirk, 1984, 1996; Shattuck-Hufnagel & Turk, 1996). Since phrasal prosodic constituent boundaries are also word boundaries, they can also be used for discovering words (Millotte, Frauenfelder, & Christophe, 2007; Shukla, Nespor, & Mehler, 2007). On the other hand, because prosody uses at least partially different cues to signal breaks at different levels of the prosodic hierarchy, it also provides information about how the segmented units relate to each other. For example, the declining pitch contour does not only signal where an Intonational Phrase begins and ends, but it also groups together the Phonological Phrases that it contains. In theory, thus, prosody can play a crucial role in language acquisition both for finding words to build a lexicon and for discovering at least part of the syntactic structure according to which words are arranged into sentences.

In previous studies, prosody has been seen as a tool for finding individual constituents in the speech stream (e.g., either words or phrases) and is often neglected as a viable cue for bootstrapping into the hierarchical syntactic structure (i.e., finding both words and phrases or both phrases and sentences). There is, to our knowledge, no evidence that participants can use prosody for understanding the hierarchical structural relations between different prosodic constituents, and thus also possibly have a cue to the structural relations between morpho-syntactic constituents, i.e. words, phrases and sentences in the speech stream. Therefore, the second question the experiments in this paper address is whether participants are capable of using prosodic cues from different levels of the prosodic hierarchy to segment continuous speech hierarchically. Even though throughout the paper we will refer to the units as phrases and sentences, it is important to note, that these are arbitrary terms for two hierarchical levels of structure, since we do not know how our participants are treating the elements.

Mechanisms for extracting the hierarchical structure from the speech stream

There is some evidence that infants approach the speech stream as if expecting it to be hierarchically structured. Adult participants (Saffran, Newport, & Aslin, 1996) as well as 8-month-old infants (Saffran, Aslin, & Newport, 1996) are able to discover nonsense words from a continuous artificial speech stream by calculating Transitional Probabilities (TPs) between adjacent syllables. Additionally English learning 8-month-old infants can find bisyllabic words defined by TPs from naturally spoken Italian sentences (Pelucchi, Hay, & Saffran, 2009) and 17-month-old infants can map meaning to statistically defined syllable sequences (Graf Estes, Evans, Alibali, & Saffran, 2007). This led Saffran and Wilson (2003) to investigate whether 12-month-old infants can engage into two statistical learning tasks to discover simple multi-level structure in the speech stream. Infants in that study were familiarized for 2 min with a recurring sequence of 10 words that conformed to a finite state grammar. In that sequence, the TPs were always 1.0 within words and .25 at word boundaries. In the test-phase, infants listened to grammatical and ungrammatical syllable sequences and their preferential looking-times showed that they could: (1) use adjacent TPs to discover words and (2) consequently also discover the relations between the segmented words within sentence-like strings.

While the findings of Saffran and Wilson (2003) show that statistical computations can be used for discovering syntactic-like structures from simple artificial speech, statistical computations alone might be insufficient to discover hierarchical structures in natural languages. Yang (2004) argued that because monosyllabic words have no word-internal TPs, they are invisible to statistical computations that compare TPs between adjacent syllables and assign segmental breaks between syllables where TPs drop. Interestingly, eliminating this problem by taking into account also the relative frequency of words (e.g. Aslin, Saffran, & Newport, 1998), creates exactly the opposite problem where the high frequency monosyllabic words such as *be* may led the child to incorrectly parse *behave* into two separate words: *be* and *have* (Johnson & Jusczyk, 2001; Peters, 1985). Additionally, the apparent ease with which infants find words by computing TPs between syllables in artificial speech streams breaks down when the words are of different length. Thus, while 8-month-infants are capable of discovering bisyllabic and trisyllabic words when the length of the words is homogenous in the familiarization stream (either only bisyllabic or only trisyllabic words in the familiarization stream) (Houston, Santelman, & Jusczyk, 2004), they fail to find words when their length varies (familiarization stream contained both bisyllabic and trisyllabic words) (Johnson & Jusczyk, 2003a, 2003b; Johnson & Tyler, 2010). Though, this failure might be caused by artificial stimuli, as with natural language stimuli infants could find words with variable length by relying on TPs (Hay, Pelucchi, Graf Estes, & Saffran, 2011; Pelucchi et al., 2009). Finally, considering the statistical relations between syntactic constituents, the size of the lexicon and the possible combinations in which words can be arranged, suggests that while the differences between within-word TPs and TPs at word boundaries may differ sufficiently to discover possible word candidates, the differences between TPs at different word boundaries are bound to be considerably smaller than the differences between within-word TPs and TPs at word boundaries (c.f. Saffran, Newport et al., 1996). Thus while statistical

computations may play a role in discovering possible word-candidates in continuous speech, they are bound to be extremely inefficient for discovering the structural grouping principles between words.

There is experimental evidence concerning the relative strengths of TPs and prosody as cues for finding possible word-candidates in continuous speech. Johnson and Jusczyk (2001) used the Head-Turn Preference Procedure to investigate the interaction of prosody and TPs by familiarizing 8-month-old English-learning infants with sequences of syllables that contained TP cues, as well as word-level speech cues (word stress and coarticulation) signaling different word boundaries. In the test phase, infants' looking-times showed a novelty-preference for statistical words over part-words signaled by prosodic cues, indicating that stress and coarticulation count more than statistical regularities in segmenting speech stimuli. Similar results have been found also with 11-month-infants (Johnson & Seidl, 2009), who weigh prosodic cues (i.e. word stress) more heavily than statistical regularities. However, the study of Thiessen and Saffran (2003) suggests that infants may start using word-stress only after they have segmented the words using statistics. In that study 9-month-old (Experiment 1) and 7-month-old (Experiment 2) English-learning infants were familiarized with four bisyllabic words in a random order for 2 min. Infants' looking times showed that 9-month-old infants used stress cues to segment words and ignored TPs, whereas 7-month-old infants relied more on TPs than on stress cues. These results have been taken to suggest that infants start tackling the speech stream by using statistical regularities, discover the stress pattern from statistically segmented words, and only then begin relying more heavily on prosody than on statistics. Not only is stress location language-specific (Hyman, 1977), but the location of the stressed syllable also differs within a language (e.g. while English is predominantly a stress initial language, some words such as "guitar" have stress on the final syllable). This indicates that the specific stress patterns of a language have to be acquired (e.g. Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003) and that a segmentation strategy based on lexical stress alone (e.g. Cutler & Carter, 1987; Cutler & Norris, 1988) will not guarantee the discovery of words (e.g. Dupoux, Pallier, Sebastian, & Mehler, 1997; Sebastian, Dupoux, Segui, & Mehler, 1992).

Instead, experimental evidence suggests that prosodic boundary cues may be more reliable cues for segmentation than lexical stress. For example, Intonational Phrase (Kjelgaard & Speer, 1999; Warren, Grabe, & Nolan, 1995), Phonological Phrase (Gout et al., 2004; Millotte et al., 2008) as well as Phonological Word boundaries (Endress & Hauser, 2010; Millotte et al., 2007) constrain lexical access, suggesting that prosodic boundary cues might universally signal constituents at different levels of the prosodic hierarchy. Thus, Shukla et al. (2007) used Intonational Phrase boundaries to investigate the relative strength of TPs and prosody. Italian-speaking adult participants were familiarized for 8 min with an artificial speech-stream that contained statistically defined words (word-internal adjacent TP = 1.0) that occurred either within 10-syllable-long Intonational Phrases (defined by pitch decline) or straddled Intonational Phrase boundaries. In the test-phase participants were asked to discriminate between words that occurred in the familiarization phase and words that did not. The results show that participants recognized the words only when they occurred within the Intonational Phrase boundaries, but not when they straddled them. Importantly, similar results were found with 6-month-old infants who could map meaning to bisyllabic sequences defined by TPs only when the syllable sequences were aligned with Intonational Phrase boundaries and not when the syllable sequences straddled the Intonational Phrase boundaries (Shukla, White, & Aslin, 2011). This suggests that prosodic boundary cues (i.e. the declining pitch contour) are used as filters to suppress possible statistically well-formed word-like sequences that occur across Intonational Phrase boundaries. Prosodic boundary cues appear to be considerably more consistent than lexical stress both within and across languages. It is not known, however, whether prosodic boundary cues (such as pitch declination and final lengthening) signal different levels of the prosodic hierarchy on-line.

It is also known that besides calculating TPs between syllables, infants as young as 7-months can also learn simple structural regularities of the kind ABA or ABB (e.g. "gatiga" and "nalili") from as brief exposure as 2 min (Marcus, Vijayan, Bandi Rao, & Vishton, 1999), and the neonate brain appears to distinguish structural sequences such as ABB (e.g. "mubaba,") from structureless sequences such as ABC (e.g. "mubage") already during the first days of life (Gervain, Macagno, Cogoi, Peña, & Mehler, 2008). Kovács and Endress (in preparation) thus investigated with a modified head-turn preference procedure (see Gervain, Nespor, Mazuka, Horie, & Mehler, 2008) whether 7-month-old infants can learn hierarchically embedded structures that are based on identity relations at two different levels. They familiarized infants with a stream of syllables that contained words formed by syllable repetitions ("abb" or "aba", where each letter corresponds to a syllable), and sentences that were formed by repetition of these trisyllabic words ("AAB" or "ABB", where each letter corresponds to a word). In the test-phase, infants' looking times showed that they were able to discriminate novel syllable sequences adhering to the repetition rules from illegal syllable sequences both at the word and at the sentence level. This is one more piece of evidence that suggests that infants do approach the speech signal as if expecting it to be organized on multiple-structural levels.

Because the majority of studies investigating grammar-like rule learning have used segmented artificial streams, Peña, Bonatti, Nespor, and Mehler (2002) investigated whether a continuous speech stream also allows the extraction of structural generalizations. They familiarized adult participants with a syllable stream composed of a concatenation of trisyllabic nonsense words. In each word, the first syllable predicted the last syllable with certainty, whereas the middle syllables varied. To identify words and rules, participants could thus not rely on adjacent TPs, but had instead to compute TPs between nonadjacent syllables. Participants were asked which words they had heard during familiarization in a dual-forced choice task.

The results demonstrate that participants could compute non-adjacent TPs, but only for segmenting the speech stream and not for generalizing the dependency between the first and the last syllable of words. After a 10 min long familiarization, participants chose words that occurred during familiarization over part-words that occurred during familiarization but violated the word-boundaries signaled by TPs. However, they did not prefer novel-rule words that had a novel middle syllable over part-words that actually occurred during the familiarization but violated the word-boundaries signaled by TPs. Participants failed to generalize the long-distance dependencies even after a 30 min long familiarization. Only when words were separated by subliminal pauses (25 ms), could participants generalize the dependency between the first and the last syllable by choosing rule-words that had a novel middle syllable over part-words that occurred during familiarization but violated word-boundaries (see however Perruchet, Tyler, Galland, & Peereman, 2004 for criticism; and Bonatti, Peña, Nespor, & Mehler, 2006 for a reply). On the basis of these findings Peña et al. (2002) argued that structural generalizations can be only drawn from a segmented speech stream and that the subliminal pauses that allowed rule-generalization mimicked the cues provided by prosodic constituency (i.e. Nespor & Vogel, 1986; Selkirk, 1984; cf. Bonatti et al., 2006 for a thorough discussion). The necessity of segmental cues in the form of pauses for discovering and generalizing long-distance regularities is also found in 12-month-old infants (Marchetto & Bonatti, in preparation a, in preparation b).

However, in natural languages words are not systematically separated by subliminal pauses (Perruchet et al., 2004). In fact, pauses have been found to be unreliable cues for word segmentation in natural speech (for a discussion about pauses as segmentation cues cf. Fernald & McRoberts, 1995). It remains unknown whether real prosodic cues facilitate grammar-like rule learning, and whether they do so at every level of the prosodic hierarchy. Thus the third question the experiments in this paper address is whether rule-generalization is facilitated also with more realistic prosodic cues than silences (like duration and pitch) that correspond to actual cues present in the speech stream. Importantly, because in natural language structural regularities are organized hierarchically, we ask whether listeners can generalize hierarchical structural regularities by using cues from different levels of the prosodic hierarchy.

Can prosody be used for discovering hierarchical structure in continuous speech?

The prosody of fluent speech signals syntactic constituency – a property that may facilitate the acquisition of hierarchical structures from the speech stream. We investigate the sensitivity to prosodic structure in four artificial grammar experiments where participants were first familiarized with an artificially synthesized speech stream that contained prosodic cues that signal constituents at different levels of the prosodic hierarchy, and then tested for learning grammar-like regularities on a dual forced-choice task. We used the two prosodic cues that are most reliable for syntactic processing (as discussed above): duration and pitch. We implemented these cues onto two distinct levels of the prosodic hierarchy: duration as final lengthening in Phonological Phrase (henceforth referred to as phrases), and a declining pitch contour spanning the Intonational Phrases (henceforth referred to as sentences). The prosody was artificially synthesized over an imaginary language composed of phrases and sentences that contained long-distance dependencies where the first syllable predicted with certainty the last syllable of any given constituent (c.f. Peña et al., 2002). Furthermore, the structure of the familiarization streams was created so that adjacent transitional probabilities straddled prosodic boundaries, and non-adjacent transitional probabilities were contained in prosodic constituents. While a direct test for the generalization of non-adjacent long-distance dependencies does not fall within the scope of the present paper, we did investigate both the interaction of prosody and statistical computations, and the role of prosody in the extraction of generalizations from continuous speech.

Experiment 1: Rule learning through prosodic cues from different levels of the prosodic hierarchy

In Experiment 1 we investigated four questions: (1) Can listeners keep track of prosodic cues from different levels of the prosodic hierarchy? (2) Do listeners perceive prosody as organized hierarchically? (3) Can listeners use hierarchically structured prosody to acquire hierarchically organized rule-like regularities? and (4) Does prosody provide stronger cues for constituent boundaries than do TPs? Thus, we used a familiarization paradigm to test whether listeners rely on duration to group syllables into phrases, while simultaneously relying on pitch declination to group phrases into sentences, and consequently generalize structural regularities on both levels. The familiarization stream consisted of phrases that followed long-distance dependency rules structurally identical to those used in Peña et al. (2002). However, in order to instantiate rules also at a higher level – let us call it the sentence level – we did not change the order of the phrases (contrary to Peña et al., 2002), but paired each two subsequent phrases into a sentence. This resulted in adjacent transitional probabilities across phrase and sentence boundaries of 1.0. Since according to the statistical learning literature, high TPs are perceived within (rather than across) constituent boundaries (Saffran, Aslin et al., 1996; Saffran, Newport et al., 1996) and adjacent TPs are easier to compute than non-adjacent TPs (Gebhart, Newport, & Aslin, 2009; Newport & Aslin, 2004; Peña et al., 2002), participants were not expected to find phrases and sentences when tracking transitional probabilities (see Experiment 3). Instead, to signal phrase and sentence level constituents, in the familiarization stream we used two prosodic cues from two distinct levels of the prosodic hierarchy: (A) final lengthening that mimicked Phonological Phrases and was instantiated over the final vowel of each phrase; and (B) pitch declination that mimicked Intonational Phrases, and was instantiated over sentences. If prosodic cues from different levels of the prosodic hierarchy are stronger cues for

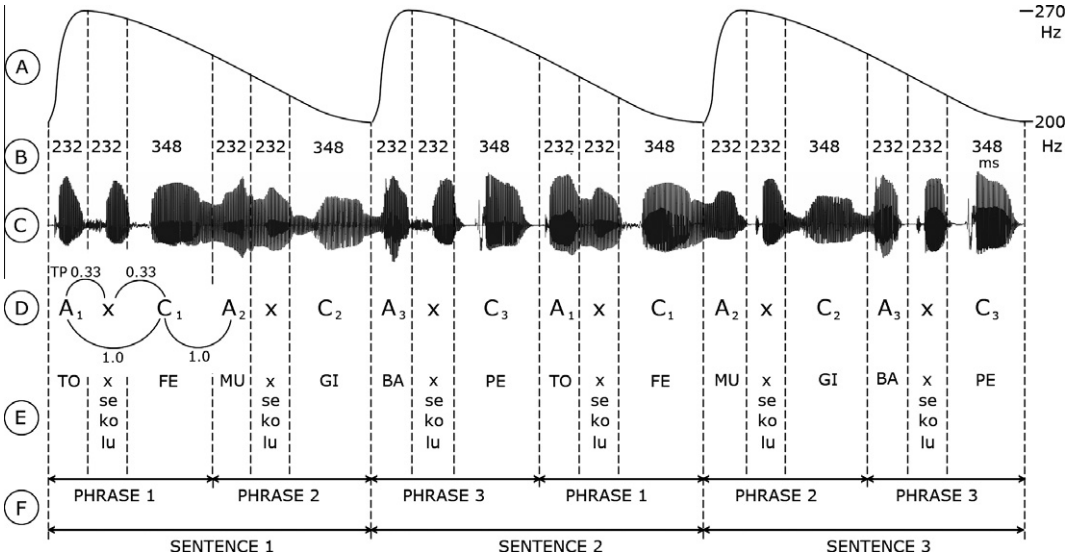

Fig. 1. Prosody and structure of the habitation streams: (A) declining pitch contour; (B) syllable length; (C) the speech signal; (D) the long-distance dependency rules; (E) the actual syllables used; and (F) how syllables formed phrases and sentences. The streams were structurally identical in all four conditions. In the no-prosody conditions (not depicted) the pitch was constant at 200 Hz and the syllable length was 232 ms.

segmentation than TPs, then participants should be able to use both final lengthening and pitch declination to extract the rules on the phrase- as well as on the sentence-levels.

Method

Participants

We recruited 28 native speakers of Italian (13 females, mean age 20.1, range 19–25 years) from the subject pool of SISSA – International School of Advanced Studies (Trieste, Italy). Participants reported no auditory, vision, or language related problems. They received a monetary compensation.

Materials

Fig. 1 shows the syllabic and prosodic structure of the familiarization stream. The syllabic structure of the familiarization stream contained trisyllabic phrases that formed two-phrase sentences (see Fig. 1D–F). In order to increase surface variation in the familiarization stream, both phrases and sentences followed long-distance dependency rules, where the first syllable of each constituent predicted the last syllable with certainty at both phrase and sentence level (Fig. 1D). We used three long-distance dependency rules (AxC) on the phrase level (TO_x_FE; MU_x_GI; BA_x_PE). In all these rules, the first syllable (A) always predicted the third syllable (C) with a probability of 1.0. The middle syllable (x) varied between three different syllables that were the same for all three rules (se; ko; lu) (Fig. 1E). Two consecutive phrases formed a sentence. Because the phrases were repeated in the same order throughout the familiarization stream, there were exactly three long-distance dependency rules on the sentence level (TO_x_FEMU_x_GI; BA_x_PETO_x_FE; MU_x_GIBA_x_PE) (Fig. 1F). In all these rules the first syllable of the first phrase (A_1) always predicted the final syllable of the second phrase (C_2) with a probability of 1.0 and the last syllable of the first phrase (C_1) always predicted the first syllable of the second phrase (A_2) with a probability of 1.0. In the familiarization stream each of the three long-distance dependency rules that formed the phrases (AxC) was repeated 60 times (a total of 180 phrase repetitions) and each of the three long-distance dependency rules that formed the sentences ($A_1 \ldots C_2$) was repeated 30 times (a total of 90 sentence repetitions).

The familiarization stream included prosodic cues for Phonological Phrases (final lengthening) and for Intonational Phrases (declining pitch contour) (see Fig. 1A–C). The final lengthening was instantiated by increasing the duration of the final vowel of each phrase by 50%, resulting in a phoneme length of 232 ms.[1] All the other phonemes were 116 ms long. The declining pitch contour started from a baseline of 200 Hz with a rapid initial ascent that peaked at 270 Hz on the center of the vowel of the first syllable of the first phrase of the sentence and then declined to 200 Hz at the center of the vowel of the last syllable of the second phrase of the sentence. In between these points, pitch was interpolated and then smoothed quadratically (four semitones). These parameters fall within the range used in previous studies (c.f. Bion et al., 2011), and are within the limits of

[1] Final lengthening was instantiated over the final vowel of each phrase (and not over the whole final syllable that consistent of a consonant and a vowel) because pilot experiments showed that participants did not notice lengthening when it was instantiated over the whole syllable. This is in line with the finding that in English consonants tend to be longer in word-initial position than in word-final position (Klatt, 1974; Oller, 1973; Umeda, 1977). Thus because words are exhaustively contained in Phonological Phrases (e.g. Selkirk, 1984; Nespor & Vogel, 1986), also the consonants at the end of Phonological Phrases must be shorter than at the beginning of Phonological Phrases. This may suggest that participants failed to perceive final lengthening over the final syllable because lengthening the consonants was in conflict with the expectation of finding longer consonants at the beginning of the phrases.

pitch and syllable durations in natural speech (c.f. Shukla et al., 2007). The familiarization stream was 2 min 26 s long. In order to prevent participants from finding the phrases and sentences simply by noticing the first or the final phrase of the familiarization stream, the initial and final 10 s of the file were ramped up and down in amplitude to remove onset and offset cues.

The materials for testing rules on the phrase-level consisted of nine novel rule-phrases and 18 part phrases (for a full list of test materials see Appendix A). The novel rule-phrases had the same AxC long-distance dependency but a middle syllable (x) that had not occurred in this position before (mu; gi; ba). Thus the novel rule-phrases contained the non-adjacent long-distance dependency of 1.0 but broke the adjacent dependency of 1.0. The part-phrases were present in the familiarization phase but violated the prosodically signaled phrase boundaries (Fig. 1D: x_C_1A_2; C_1A_2_x; x_C_2A_3; C_2A_3_x; x_C_3A_1; C_3A_1_x). Thus the part phrases contained the adjacent dependency of 1.0 but broke the non-adjacent long-distance dependency of 1.0. The test materials for testing rules on the sentence-level consisted of nine novel rule-sentences and nine part-sentences (see Appendix A). The novel rule-sentences had the same long-distance dependency ($A_1 \ldots C_2$) as the familiarization sentence, but had a middle syllable (x) that had not occurred in this position before (mu; gi; ba). The part-sentences paired two rule-phrases that occurred during the familiarization but that did not form a sentence in the familiarization phase (Fig. 1D: A_1_x_C_1A_3_x_C_3; A_2_x_C_2A_1_x_C_1; A_3_x_C_3A_2_x_C_2). All the phonemes in the test items were 116 ms long (test phrases were 696 ms and the test sentences were 1392 ms long). We used prosodically flat test items, in order not to bias the choice of the participants. Thus, the phrases and sentences heard during test are acoustically different from those heard during familiarization.

All the stimuli of this experiment as well as of the following experiments were synthesized with PRAAT (Boersma, 2001) and MBROLA (Dutoit, Pagel, Pierret, Bataille, & Van Der Vreken, 1996). We used the French female diphone database (fr2). We chose the French diphone database since pilot studies showed that artificial speech synthesized using this database resulted in speech that was perceived by Italian adults better than with other similar databases, including the Italian diphone database (notice that the diphone database does not encode sentential prosody).

Procedure

Participants were seated in front of a 15-in. PowerBook G5 Apple Macintosh computer that was equipped with SONY MDR-XB700 headphones. The experiment was designed and run on PsyScope X experimental software (version B54).

The experiment followed a between-subjects design. Participants were randomly assigned to one of two testing conditions: (1) phrase-level rule testing with prosody; or (2) sentence-level rule testing with prosody. Participants were told that they would listen to a continuous stream of an imaginary language that has its own words and rules that do not exist in any language of the world (familiarization phase). Participants were instructed to listen carefully because in the second part of the experiment (test phase), they would be asked to discriminate between pairs of sounds in which one is a phrase/sentence of the imaginary language and the other is not (test phase).

In order to guarantee that participants had understood the task, before the beginning of the familiarization phase, they were asked to complete a training session of the dual forced choice task. Subjects were presented with 10 pairs of syllables: the syllable [zo] coupled with another syllable ([pwo], [pje], [pwe], [ze], [za]). The syllables of each pair were separated by a pause. Participants had to indicate on the computer keyboard whether the first or the second syllable they heard was /zo/. In this pre-training session, participants were given feedback about their performance (correct/incorrect) after each trial and could only continue to the familiarization phase after having correctly responded on 10 trials.

During the familiarization phase participants listened to the familiarization stream that was identical for both conditions (see Materials section). Following the familiarization phase, participants were tested on a dual forced choice task that consisted of 36 pairs of sequences. Participants tested for rules on the phrase-level (condition 1) were presented with pairs of syllable sequences in which one sequence was always a novel rule-phrase and the other a part-phrase. Participants tested for rules on the sentence-level (condition 2) were presented with pairs of syllable sequences in which one sequence was always a novel rule-sentence and the other a part-sentence. The presentation order of novel rule- and part-sequences (which was presented first and which second in the dual forced choice task) was balanced across trials. Before the beginning of the test phase, participants were told that they had to respond with the first idea that came to their mind to the question: "which of the two sequences is most likely a phrase/sentence in the imaginary language". Participants were told that the phrases/sentences did not necessarily appear during the familiarization phase, and that they had to give a response before they could continue to the next trial.

Results

Fig. 2A presents the percentage of correctly chosen novel rule-phrases. Participants in condition 1 chose novel

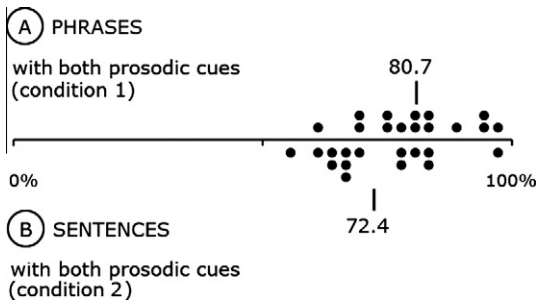

Fig. 2. Participants' responses in Experiment 1: (A) the average percentage of correctly chosen novel rule-phrases over part-phrases on the phrase level rule learning with prosody (condition 1); (B) the average percentage of correctly chosen novel rule-sentences over part-sentences on the sentence level rule learning with prosody (condition 2).

rule-phrases over part-phrases on average in 80.7% of the cases (*t*-test against chance with equal variance not assumed: $t(13) = 11.670$, $p < .001$). Fig. 2B presents the percentage of correctly chosen novel rule-sentences. Participants in condition 2 chose novel rule-sentences over part-sentences on average 72.4% of the cases (*t*-test against chance with equal variance not assumed: $t(13) = 7.474$, $p < .001$). Participants tested for rules on the phrase-level (condition 1) chose correct novel rule-phrases significantly more than participants tested for rules on the sentence-level (condition 2) chose novel rule-sentences (*t*-test: $t(26) = 1.923$, $p = .046$).

Discussion

The results suggest that participants used the prosodic cues to learn rules for both phrases and sentences. On the one hand, participants chose significantly more rule-phrases with novel surface structure (novel middle syllables that they had not heard in this position before) than part-phrases they had actually heard during the familiarization but that violated the prosodic boundaries (condition 1). Were they not relying on final lengthening that signaled the phrase-boundary, we would expect them to have preferred part-phrases that actually occurred in the familiarization stream and that contained the adjacent transitional probability of 1.0. On the other hand, participants chose significantly more rule-sentences that had a novel surface structure (novel middle syllables that they had not heard in these positions before) over part-sentences that contained phrases they had actually heard during the familiarization phase, but that did not conform to the rule-sentence (condition 2).

Importantly, participants in both conditions were familiarized with the same stream that contained both prosodic cues and they were not informed beforehand whether they would be queried for phrases or sentences. Thus, on the one hand, if participants were not processing the statistical or prosodic information in familiarization stream, their performance in choosing novel rule-phrases and rule-sentences over part-phrases and part-sentences in the test phase would have been at chance level. On the other hand, if participants paid attention to the syllabic structure of the familiarization stream, but failed to use the prosodic cues, we would have expected them to segment the stream into constituents that contained the adjacent TP of 1.0 both at phrasal and at sentence levels, and consequently not be able to discover the non-adjacent long-distance dependencies that formed rule-phrases and rule-sentences (for a control with a prosodically flat familiarization stream see Experiment 3). This suggests that in order to perform above chance on both phrase-level and sentence-level rule testing, listeners were able to keep track of prosodic cues from different levels of the prosodic hierarchy.

It is also important to note that these results suggest that participants appeared to use prosodic cues to group syllables into constituents rather than simply use them to segment the familiarization stream. If participants were only segmenting the familiarization stream, they should have found only the constituents that fall between any two segmental cues in the speech stream regardless of what these cues are. The corresponding constituents in our familiarization stream where phrases. However participants also found sentences, suggesting that they did use final lengthening and pitch declination to group syllables into constituents on two different levels. This idea is supported by the finding that, significant differences emerged also between phrase-level (condition 1) and sentence-level (condition 2) rule learning. There are three possible explanations for this. It is possible that participants segmented and grouped syllables together according to their respective prosodic cues online. Alternatively, the differences between sentences and phrases may have emerged because there were double as many instances of phrases as there were sentences. The relative difficulty of finding the sentence-level rules could also be increased by the fact that final lengthening is sometimes seen as a stronger prosodic cue than a declining pitch contour. This view is supported by evidence that final lengthening appears to be a more consistent cue to segmentation than the declining pitch contour (de Rooij, 1976; Beach, 1991; Streeter, 1978; for a discussion see Fernald & McRoberts, 1995). Since, in the experimental stimuli the sentences were twice as long as the phrases, participants may have performed better on the phrases because they require less memory capacity.

However, there is another possibility that could explain participants' poorer performance on sentences than on phrases. In the familiarization phase the TPs between phrases were always 1.0 (that is, the order of phrases did not change). In the test-phase, participants in condition 2 had to choose between novel rule-sentences that conformed to this order and part-sentences that violated the order of phrases (the phrases that formed part-sentences occurred in the familiarization stream, but the part-sentences themselves did not). If participants found phrases by relying on final lengthening, it might have been enough to remember the order of all the phrases and not process the pitch-declination at all. Participants thus might have performed worse on sentences because they did not perceive pitch declination: they had to find the phrases first and then the order between them (as the order between the phrases did not vary and part-sentences contained phrases in a scrambled order, participants could identify sentences from part-sentences simply by relying on the order of phrases). In order to distinguish between these two alternatives, according to which participants were using either both prosodic cues, or only final lengthening, we carried out Experiment 2. Importantly, because the TPs of the familiarization stream predicted phrase/sentence boundaries differently from prosodic boundaries (high adjacent TPs of 1.0 fall within constituents and not at constituent boundaries), we carried out Experiment 3 where participants were familiarized with prosodically flat streams.

Experiment 2: Rule learning through single prosodic cues from different levels of the prosodic hierarchy

In Experiment 2 we attempted to determine whether participants were indeed using both prosodic cues (final lengthening and pitch declination) to extract rules from

the familiarization streams. We kept the syllable structure of the familiarization streams used in Experiment 1, but familiarized participants with either only final lengthening as a cue to phrases or with only pitch declination as a cue to sentences. We expected participants to learn phrase-level rules only when they had final lengthening as a cue to phrases and to learn sentence-level rules only when they had pitch declination as a cue to sentences: (1) participants who were familiarized with prosodic cues to phrases (final lengthening) were expected to choose novel rule-phrases over part-phrases that actually occurred in the familiarization stream; (2) participants who were familiarized with prosodic cues to phrases (final lengthening) were not expected to choose novel rule-sentences over part-sentences; (3) participants who were familiarized with prosodic cues to sentences (pitch declination) were expected to choose novel rule-sentences over part-sentences; and (4) participants who were familiarized with prosodic cues to sentences (pitch declination) were not expected to choose novel rule-phrases over part-phrases. Importantly, if participants in Experiment 1 relied only on final lengthening (and not on pitch declination), we expected them to fail on rules instantiated at the phrase-level because the only prosodic cue they had available was pitch declination.

Method

Participants

We recruited 56 native speakers of Italian, mean age 20.1, range 19–26 years, 28 females, from the subject pool of SISSA – International School of Advanced Studies (Trieste, Italy). None of the participants had taken part in Experiment 1. Participants reported no auditory, vision or language related problems. They received a monetary compensation.

Materials

The stimuli of Experiment 2 consisted of two familiarization streams. The syllabic structure of the streams was identical to the one used in Experiment 1 (see the Materials section of Experiment 1). The crucial difference with respect to Experiment 1 was that the individual prosodic cues (pitch declination and final lengthening) were separated into two familiarization streams. One stream contained prosodic cues only for phrases (final lengthening). Final lengthening was identical to that used in Experiment 1. The resulting familiarization stream was 2 min and 26 s long. The other stream contained prosodic cues only for sentences (declining pitch contour). Pitch declination was identical to that used in Experiment 1. The resulting familiarization stream was 2 min 5 s long (the second familiarization stream was shorter because there was no final lengthening, but it contained the same number of instances of phrases and sentences as the first familiarization stream). The materials for the test phase were identical to those of Experiment 1. The synthesis of the stimuli was identical to that in Experiment 1.

Procedure

The experimental setup and the procedure of Experiment 2 were identical to that of Experiment 1, except that the familiarization stream varied either only in duration or only in pitch (instead of varying in both pitch and duration). Therefore, this experiment comprises four conditions – two familiarization conditions (stream varying in pitch or stream varying in duration) and two test conditions (investigating listeners segmentation of phrases or listeners segmentation of sentences): (1) rule testing with prosodic cues for only phrases (familiarization contained only final lengthening) with test on part-phrases against novel rule-phrases; (2) rule testing with prosodic cues for only phrases (familiarization contained only final lengthening) and test on part-sentences vs. novel rule-sentences; (3) rule testing with prosodic cues for only sentences (familiarization contained only pitch declination) and test on part-sentences vs. novel rule-sentences; and (4) rule testing with prosodic cues for only sentences (familiarization contained only pitch declination) and test on part-phrases vs. novel rule-phrases.

Results

Fig. 3A presents the percentage of correctly chosen novel rule-phrases following the familiarization with final lengthening only. Participants in condition 1 chose novel rule-phrases over part-phrases on average 82.3 % of the cases (*t*-test against chance with equal variance not

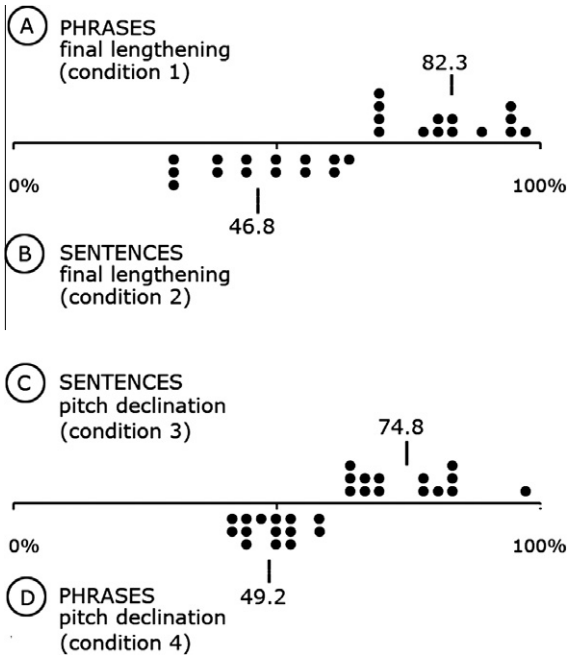

Fig. 3. Participants' responses in Experiment 2: (A) the average percentage of correctly chosen novel rule-phrases over part-phrases at the phrase-level rule testing with final lengthening as a cue (condition 1); (B) the average percentage of correctly chosen rule-sentences over part-sentences on the sentence level rule learning with final lengthening as a cue (condition 2); (C) the average percentage of correctly chosen novel rule-sentences over part-sentences at the sentence-level rule testing with pitch declination as a cue (condition 3); (D) the average percentage of correctly chosen novel rule-phrases over part-phrases at the phrase-level rule testing with pitch declination as a cue (condition 4).

assumed: $t(13) = 7.221$, $p < .001$). Fig. 3B presents the percentage of correctly chosen novel rule-sentences following the familiarization with final lengthening only. Participants in condition 2 did not significantly choose novel rule-sentences over part-sentences (46.8%: t-test against chance with equal variance not assumed: $t(13) = 10.348$, $p = .66$). Fig. 3C presents the percentage of correctly chosen novel rule-sentences following the familiarization with pitch declination only. Participants in condition 3 chose novel rule-sentences over part-sentences on average 74.8 % of the cases (t-test against chance with equal variance not assumed: $t(13) = 10.644$, $p < .001$). Fig. 3D presents the percentage of correctly chosen novel rule-phrases over part-phrases following the familiarization with pitch declination only. Participants in condition 4 did not significantly chose novel rule-phrases over part-phrases (49.2%: t-test against chance with equal variance not assumed: $t(13) = 7.342$, $p = .67$). Participants in phrase-level rule testing (condition 1) chose correct novel rule-phrases significantly more than participants in sentence-level rule testing (condition 3) chose rule-sentences (t-test: $t(26) = 4.237$, $p = .045$).

No significant differences emerged between Experiment 1 and 2. Participants who were familiarized with both prosodic cues (final lengthening and pitch declination) and tested on phrase-level rule testing (Experiment 1 condition 1) did not perform significantly better than participants who were familiarized only with final lengthening and tested on phrase-level rule testing (Experiment 2 condition 1) (t-test: $t(26) = 3.020$, $p = .231$). Also participants who were familiarized with both prosodic cues and tested on sentence-level rule testing (Experiment 1 condition 2) did not perform significantly better than participants who were familiarized only with pitch declination and tested on sentence-level rules (Experiment 2 condition 3) (t-test: $t(26) = 4.121$, $p = .134$).

Discussion

The results suggest that participants in Experiment 2, just like participants in Experiment 1, used prosodic cues to learn rules for both phrases and sentences. On the one hand, participants familiarized with an artificial speech stream that contained prosodic cues for phrases (condition 1), chose significantly more rule-phrases with novel surface structure (novel middle syllables that they had not heard in this position before) than part-phrases they had actually heard during the familiarization but that violated the prosodic boundaries. Were they not relying on final lengthening that signaled the phrase-boundary, we would expect them to have preferred part-phrases that actually occurred in the familiarization stream that contained the adjacent TP of 1.0. On the other hand, participants familiarized with an artificial speech stream that contained prosodic cues for sentences (condition 3), chose significantly more rule-sentences that had a novel surface structure (novel middle syllables that they had not heard in these positions before) over part-sentences that they had actually heard during the familiarization but that violated the prosodic boundaries. Were they not relying on pitch declination that signaled the beginning and the end of sentences, we would expect them to have preferred part-sentences that had actually occurred in the familiarization stream.

Importantly, we can also rule out the possibility that participants in Experiment 1 were relying only on final lengthening to learn phrase-level rules as well as sentence-level rules (and were not relying on pitch declination at all, because the order of phrases did not change in the familiarization stream). Participants, who were familiarized with only final lengthening as a cue to phrases, did not choose significantly more novel rule-sentences over part-sentences (condition 2). This suggests that while final lengthening alone enabled participants to generalize the phrase-level rules, it was not sufficient to learn the rules for sentences. Additionally, participants, who were familiarized with only pitch declination as a cue to sentences, did not choose significantly more novel rule-phrases over part-phrases (condition 4). This suggests that while pitch declination alone enabled participants to generalize the sentence-level rules, it was not sufficient to learn the rules at the phrasal-level. Because the syllabic structure of the familiarization streams in Experiment 1 and 2 were identical, participants must have relied on the individual cues: on final lengthening for phrase-level rules and on pitch declination for sentence-level rules.

Our results also suggest that in Experiment 1 participants were not choosing significantly less novel rule-sentences than novel rule-phrases because they only relied on final lengthening. The findings of Experiment 1 left open the possibility that participants were performing better on the phrasal level than on the sentence level simply because they only used final lengthening to find phrases and then, because the order of the phrases did not vary, discovered the sentences. If this were the case, we would have expected participants who were familiarized only with final lengthening to perform like participants in Experiment 1. In fact, participants in Experiment 2 did choose more novel rule-phrases than novel rule-sentences (conditions 1 and 3). However, participants who were familiarized with final lengthening (Experiment 2 condition 2) did not choose novel rule-sentences significantly over part-sentences, suggesting that final lengthening alone was insufficient to discover rules on both phrasal- as well as sentence-level. The differences between phrase-level and sentence-level rules in Experiment 1 and 2 must thus have emerged because final lengthening is a stronger cue to phrase boundaries than pitch declination is to sentence boundaries, because in the familiarization phase there were twice as many instances of phrases as there were sentences, or because sentences were longer and thus potentially more demanding on memory.

Interestingly, the comparisons between the findings of Experiment 1 and Experiment 2 also suggest that the strength of a prosodic boundary is not the sum of the strength of individual prosodic cues. In the familiarization phase of Experiment 1 the end of each sentence was marked by both final lengthening and pitch decline. In contrast, in Experiment 2 participants familiarized with one cue only, in that sentences were only marked by pitch declination. However, there were no significant differences between participants' performance in learning sentence-level

rules in Experiment 1 (condition 2) and in Experiment 2 (condition 3). Our results thus suggest that pitch declination and final lengthening are not additive in signaling prosodic boundaries. Instead, participants seem to have assigned the individual cues to specific levels in the prosodic hierarchy and used them separately to discover rules at the phrase-level and sentence-level.

While the findings of Experiment 2 showed that participants did indeed rely on both prosodic cues (final lengthening and pitch declination) to learn the rules at the phrase-level and at the sentence-level, both Experiment 1 and Experiment 2 relied on the assumption that if participants did not learn the rules they would be performing at chance. However, this need not be the case as the familiarization streams also contained statistical regularities between syllables that could be used for segmentation (Aslin et al., 1998; Saffran, Aslin et al., 1996; Saffran, Newport et al., 1996). In order to investigate how participants would treat the familiarization streams when these are stripped of prosodic cues, to establish the magnitude of the rule learning effects, and to see how statistical computations interact with prosodic cues from different levels of the prosodic hierarchy, we carried out Experiment 3.

Experiment 3: Rule learning without prosodic cues

The statistics carried out on the results of Experiments 1 and 2 relied on the assumption that if participants failed to use the prosodic cues and consequently did not extract any kind of regularities from the speech stream, their performance should have been at chance level. However, the syllabic structure of the familiarization stream, that was the same in both previous experiments, contained transitional probabilities that could have influenced participants' performance. Listeners have been shown to assign word-boundaries in a sequence of syllables where the TPs between syllables drop, rather than where they increase (Aslin et al., 1998; Saffran, Aslin et al., 1996; Saffran, Newport et al., 1996). In our familiarization streams, the TPs between phrases were always 1.0 and the TPs within phrases were always 0.3. This means that, if participants use TPs for segmenting the familiarization stream, they should prefer part-phrases that included the high TP and assign the phrase-boundaries either before or after the middle (x) syllable. To test for this possibility, we stripped the familiarizations streams of prosodic cues and tested two more groups of participants for rules at the phrase-level and at the sentence-level with familiarization streams that had flat prosody.

Method

Participants

We recruited 28 native speakers of Italian, mean age 22.3, range 20–26 years, 14 females, from the subject pool of SISSA, International School of Advanced Studies (Trieste, Italy). None of the participants had taken part in Experiments 1 and 2. Participants reported no auditory, vision or language related problems. They received a monetary compensation.

Materials

The syllabic structure of the familiarization stream used in Experiment 3 was identical to the ones used in Experiment 1 and 2 (see the Materials section of Experiment 1). The crucial difference between this and previous familiarization streams was that it was prosodically flat, that is the pitch was kept constant at 200 Hz and all phonemes were 116 ms long. The resulting familiarization stream was 2 min 5 s long. The materials for the test phase were identical to those used in Experiment 1 and Experiment 2. The synthesis of the stimuli was identical to those in Experiment 1 and Experiment 2.

Procedure

The experimental setup and the procedure of the experiment were identical to that of Experiment 1, except that participants listened to a familiarization stream that was prosodically flat. The experiment followed a between-subjects design and participants were randomly assigned to one of two conditions: (1) phrase-level rule testing without prosody; or (2) sentence-level rule testing without prosody.

Results

Fig. 4A presents the percentage of correctly chosen novel rule-phrases. Participants who were familiarized with the flat stream (condition 1) chose novel rule-phrases over part-phrases on average 41.3% of the cases (*t*-test against chance with equal variance not assumed: $t(13) = -2.150$, $p = .051$). Fig. 4B presents the percentage of correctly chosen novel rule-sentences. Participants who were familiarized with the flat stream (condition 2) chose novel rule-sentences over part-sentences on average 39.7% of the cases (*t*-test against chance with equal variance not assumed: $t(13) = -4.642$, $p < .001$). The difference between participants who had to choose between novel rule-phrases and part-phrases (condition 1) or novel rule-sentences and part-sentences (condition 2) was not significant (*t*-test: $t(26) = .346$, $p = .733$).

When we compare the results of Experiment 1 and 3 we see that participants who were familiarized with the stream containing both prosodic cues (final lengthening and pitch declination) chose significantly more novel rule-phrases than participants who were familiarized with

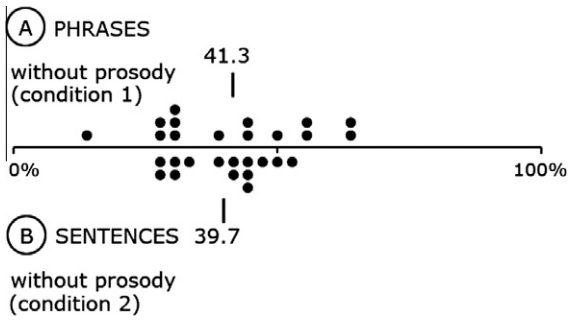

Fig. 4. Participants' responses in Experiment 3: (A) the average percentage of correctly chosen rule-phrases over part-phrases on the phrase level rule learning without prosody (condition 1); (B) the average percentage of correctly chosen rule-sentences over part-sentences on the sentence level rule learning without prosody (condition 2).

the flat stream (*t*-test: $t(26) = 8.070$, $p < .001$). Participants who were familiarized with the stream containing both prosodic cues chose significantly more also novel rule-sentences than participants who were familiarized with the flat stream (*t*-test: $t(26) = 8.769$, $p < .001$). We obtain the same results also when we compare the results of Experiment 2 and 3. Participants who were familiarized with the stream containing final lengthening chose significantly more novel rule-phrases than participants who were familiarized with the flat stream (*t*-test: $t(26) = 8.324$, $p < .001$). Participants who were familiarized with the stream containing pitch declination chose significantly more novel rule-sentences than participants who were familiarized with the flat stream (*t*-test: $t(26) = 7.343$, $p < .001$).

Discussion

These results, obtained with prosodically flat familiarization streams, demonstrate that when the familiarization streams were stripped of prosodic cues for phrases (final lengthening) and for sentences (pitch declination), participants could no longer learn the rules for either the phrases or the sentences that they were able to learn in Experiment 1 and 2. Participants tested for rules on the phrase-level (condition 1) preferred part-phrases to novel rule-phrases. Participants tested for rules on the sentence-level (condition 2) preferred part-sentences to novel rule-sentences. This means that when prosodic cues for phrase and sentence boundaries were no longer available, participants failed to generalize the long-distance dependency rules and preferred instead syllable sequences that they heard during the familiarization phase. This also means that the generalizations participants made in Experiment 1 and 2 were not simply due to the specific characteristics of the familiarization streams (i.e. the specific syllable combinations used).

The results also suggest that participants were sensitive to transitional probabilities (TPs) between adjacent syllables. Listeners have been found to assign constituent boundaries in a sequence of syllables where the TPs between syllables drop, rather than where they increase (Aslin et al., 1998; Saffran, Aslin et al., 1996; Saffran, Newport et al., 1996). In our familiarization streams, the TPs between phrases were always 1.0 and the TPs within phrases were always 0.3. This means that, if participants were using TPs for segmenting the familiarization stream, they should have preferred part-phrases that included the high TP and assign the constituent boundaries either before or after the middle (x) syllable. The results show that this is the case: participants chose part-phrases over novel rule-phrases (condition 1). These results suggest that when prosodic information is not available, participants use statistical information for segmenting the speech stream. Instead, when we compare the results of Experiments 1 and 2 to the results of Experiment 3, we observe that prosody reversed participants' preference from statistically better-defined part-phrases and part-sentences to prosodically defined rule-phrases and rule-sentences. Final lengthening and pitch declination must be powerful cues to assign a constituent boundary where statistical computations could not assign a boundary (TP 1.0). These results suggest that prosody overrides statistical computations on both the Intonational and the Phonological Phrase level.

Experiment 4: Rule learning with non-native prosody

The finding that participants preferred prosodic cues over TPs in Experiments 1 and 2 is remarkable because the syllabic structure of the familiarization streams contained TPs that were 0.3 within phrases and 1.0 between phrases and sentences. The average TPs between nouns and verbs in English have been estimated to fall within the range of .00037–.00989 (Frisson, Rayner, & Pickering, 2005; McDonald & Shillcock, 2003), suggesting that the structure of our familiarization streams strongly favored TPs over prosodic cues. Because the TPs in our familiarization streams were about 100 times stronger than in natural languages and because the calculation of TPs is an automatic process (Saffran, Aslin et al., 1996; Shukla et al., 2007), our findings suggest that prosodic cues are considerably more important than statistical cues in finding constituents and their relations in continuous speech.

However, it is also possible that participants followed prosodic cues over statistical cues because the former were more familiar than the latter. Pitch declination within Intonational Phrases and final lengthening within Phonological Phrases are typical of the prosody of many languages, including Italian (Nespor & Vogel, 1986), and their magnitude in the familiarization streams was considerably more similar to Italian prosody (Shukla et al., 2007) than the TPs were to the statistical regularities between constituents in natural languages (McDonald & Shillcock, 2003).

In order to test whether Italian-speaking participants were indeed relying on prosodic cues only because these were more familiar, we designed Experiment 4. In this experiment we aligned two individual prosodic cues (initial high pitch and final lengthening) with the phrases and the sentences in a way that while not typical of participants' native language, is typical of other languages and contrary to the familiarization streams used in Experiments 1 and 2. In the streams of Experiment 4, initial high pitch signaled phrase boundaries and final lengthening sentence boundaries. Notice that while initial high pitch in Phonological Phrases is not typical to Italian it is a natural property of the prosody of complement–head languages (Nespor et al., 2008). Also final lengthening in Intonational Phrase final position is found in many with complement-head languages, e.g. Dutch (Cambier-Langeveld, Nespor, & Van Heuven, 1997). Thus while the prosodic structure of Experiment 4 was non-native for Italian-speaking participants, it does occur in natural complement-head languages. If prosodic cues are indeed stronger than TPs in signaling constituents and their relations in continuous speech, we would expect participants' performance in Experiment 4 to be similar to that of Experiment 1 and 2 (favor constituents according to prosodic cues). However, if participants found constituents and their relations in Experiment 1 and 2 only because the prosody of the familiarization streams was similar to the prosody of Italian, we would expect them to perform like participants in Experiment 3 (favor constituents according to TPs).

Method

Participants

We recruited 28 native speakers of Italian, mean age 23.9, range 21–29 years, 11 females, from the subject pool of SISSA – International School of Advanced Studies (Trieste, Italy). None of the participants had taken part in Experiments 1, 2 and 3. Participants reported no auditory, vision or language related problems. And they were not speakers of a head-complement language as an L2. They received a monetary compensation.

Materials

The stimuli of Experiment 4 consisted of a familiarization stream that had the same syllabic structure as the familiarization streams used in Experiment 1, 2 and 3 (see the Materials section of Experiment 1). The crucial difference with respect to previous experiments was that we switched the individual prosodic cues so that pitch-declination signaled phrase boundaries (as opposed to sentence boundaries) and final lengthening sentence boundaries (as opposed to phrase boundaries) – i.e. the individual prosodic cues signaled exactly the opposite constituents from those of Experiment 1 and 2. We synthesized a pitch peak over the first syllable of every phrase by increasing the pitch from 200 Hz to 270 Hz by the middle of the first syllable (the rest of the phrase was kept constant at 200 Hz). Final lengthening was instantiated over the vowel of the last syllable of each sentence resulting in a syllable twice as long as the original one. Final lengthening thus occurred at the end of every second phrase. The resulting familiarization stream was 2 min and 16 s long. The materials for the test phase were identical to those of Experiment 1. The synthesis of the stimuli was identical to that in Experiment 1.

Procedure

The experimental setup and the procedure of Experiment 4 were identical to those of Experiment 1. That is, Experiment 4 comprises two conditions: (1) rule-testing on the phrase level, and (2) rule-testing on sentence level.

Results

Fig. 5A presents the percentage of correctly chosen novel rule-phrases. Participants in condition 1 chose novel rule-phrases over part-phrases on average in 85.3% of the cases (*t*-test against chance with equal variance not assumed: $t(13) = 13.966$, $p < .001$). Fig. 5B presents the percentage of correctly chosen novel rule-sentences. Participants in condition 2 chose novel rule-sentences over part-sentences on average 59.5% of the cases (*t*-test against chance with equal variance not assumed: $t(13) = 5.805$, $p < .001$). Both results are significant. However, participants tested for rules on the phrase-level (condition 1) chose correct novel rule-phrases significantly more than participants who were tested for rules on the sentence-level (condition 2) chose novel rule-sentences (*t*-test: $t(26) = 8.525$, $p < .001$).

When we compare the results of Experiment 3 and 4, we see that participants who were familiarized with the stream containing non-native prosodic cues chose significantly more novel rule-phrases than participants who were familiarized with the flat stream (*t*-test: $t(26) = 9.189$, $p < .001$). Participants who were familiarized with the stream containing non-native prosodic cues chose significantly more also novel rule-sentences than participants who were familiarized with the flat stream (*t*-test: $t(26) = 7.182$, $p < .001$). No significant differences emerged in rule learning on phrasal level between Experiment 1, 2 and 4. Participants who were tested for rules at the phrase-level in Experiment 4 (condition 1) did not perform significantly better than participants who were tested for rules at the phrase-level in Experiment 1 (condition 1) (*t*-test: $t(26) = 1.158$, $p = .258$) or in Experiment 2 (condition 1) (*t*-test: $t(26) = .746$, $p = .463$). However, participants who were tested on sentence-level rules in Experiment 4 (condition 2) performed significantly worse than participants who were tested on sentence-level rules in Experiment 1 (condition 2) (*t*-test: $t(26) = 3.774$, $p = .001$) and in Experiment 2 (condition 3) (*t*-test: $t(26) = 4.851$, $p < .001$).

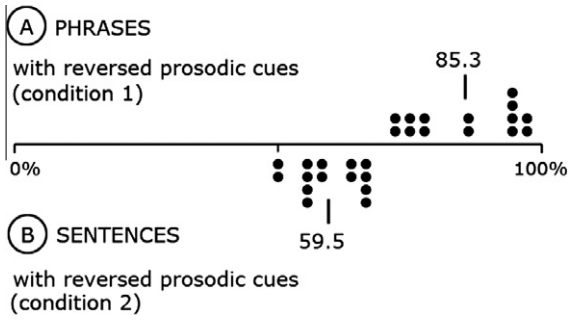

Fig. 5. Participants' responses in Experiment 4: (A) the average percentage of correctly chosen rule-phrases over part-phrases on the phrase level rule learning with reversed prosodic cues (condition 1); (B) the average percentage of correctly chosen rule-sentences over part-sentences on the sentence level rule learning with reversed prosodic cues (condition 2).

Discussion

The results of Experiment 4 show that Italian speaking participants who were familiarized with a stream that contained non-native prosody – i.e. initial pitch for phrases and final lengthening for sentences – preferred prosody over TPs even when the prosodic cues of the familiarization stream were not typical of their native language and inverted with respect to those of Experiments 1 and 2. Participants who were tested for rules at phrase-level (condition 1) preferred novel rule-phrases to part-phrases that actually occurred during familiarization. Similarly participants who were tested for rules on sentence-level (condition 2) preferred novel rule-sentences to part-sentences that occurred in the familiarization stream. Thus even when the specific prosodic cues were reversed, prosodic information was stronger than TPs in signaling constituents and their relations in continuous speech.

Similarly to Experiments 1 and 2, differences emerged also between phrase-level rule learning and sentence-level rule learning. Participants who were tested for phrase-level rules chose significantly more rule-phrases than participants who were tested for sentence-level rules chose rule-sentences. The stronger preference for phrases over sentences in Experiments 1, 2 and 4 could have been

caused by the fact there were double the number of instances of phrases than of sentences in the familiarization stream.

Finally, even though participants used prosodic, rather than statistical cues, for finding constituents and their relations in the familiarization stream of Experiment 4, differences between participants who were familiarized with native (Experiment 1 and 2) and non-native prosody (Experiment 4) did emerge. Participants who were familiarized with non-native prosody chose novel rule-sentences significantly less (Experiment 4 condition 2) than participants who were familiarized with native prosody (Experiment 1 condition 2; Experiment 2 condition 3). The stronger preference for pitch declination as a cue for Intonational Phrases above final lengthening, could have been caused by the fact that the first is unavoidable since it is a consequence of the respiratory system, while the second is not.

General discussion

In four experiments, we investigated whether: (1) listeners perceive prosody as hierarchically organized and assign different cues (e.g. duration and pitch) to specific levels of the prosodic hierarchy, (2) listeners use hierarchically structured prosody both to segment the speech stream and to group the segmented units hierarchically, (3) prosody plays a role in drawing generalizations from continuous speech, and (4) prosody prevails over transitional probabilities.

The results of the four experiments reported above demonstrate that participants used prosodic cues from different levels of the prosodic hierarchy to learn hierarchically organized structural regularities. In Experiment 1 participants were familiarized with a stream that contained prosodic cues to both Phonological Phrases (final lengthening) and Intonational Phrases (pitch declination). In the test phase, participants chose significantly more novel rule-phrases than part-phrases (condition 1), and also significantly more novel rule-sentences than part-sentences (condition 2). These results suggest that listeners can simultaneously keep track of variations in pitch and of syllables' duration, and that they can rely on these cues on-line in order to learn hierarchically organized structural regularities. In order to confirm that participants were indeed relying on both pitch and duration, we ran a second experiment. In Experiment 2, we familiarized participants with variations in either pitch or duration, and we investigated whether this manipulation influenced hierarchical segmentation. In the test phase, participants who were familiarized with final lengthening as a cue to phrasal boundary chose novel rule-phrases significantly more than part-phrases (condition 1). However, they did not choose novel rule-sentences over part-sentences (condition 2). This shows that final lengthening was an effective cue for phrasal but not for sentence boundaries. In contrast, participants who were familiarized with pitch declination chose novel rule-sentences significantly more than part-sentences (condition 3). However, they did not choose novel rule-phrases over part-phrases (condition 4). This shows that pitch declination was an effective cue for sentence but not for phrasal boundaries. The findings of this experiment confirm that participants treat prosodic cues from different levels of the prosodic hierarchy separately. In Experiment 3, no prosodic cues were present and all syllables had the same pitch and duration. The fact that in this case participants relied on TPs suggests that the findings of Experiment 1 and 2 were not due to any biases caused by the structure of the familiarization streams or possible similarities to words in participants' native language. In Experiment 4, participants were familiarized with a stream containing prosodic cues not typical of Italian. The fact that participants still relied on prosody suggests that even non-familiar prosody is a stronger cue than TPs for segmenting speech hierarchically.

In Experiments 1 and 2, significant differences emerged between the results of phrase-level rule testing (condition 1) and those of sentence-level rule testing (condition 2): participants chose significantly more rule-phrases over part-phrases than rule-sentences over part-sentences. This suggests that these differences were either caused by the fact that there were twice as many instances of phrases in the familiarization stream as there were sentences; or by the fact that the phrases were shorter than the sentences and required less memory capacity; or by the fact that final lengthening is a stronger cue to constituent boundaries than is pitch declination. The latter view is generally supported by evidence that final lengthening appears to be a more consistent cue to segmentation than declining pitch contour (de Rooij, 1976; Beach, 1991; Streeter, 1978; for a discussion see Fernald & McRoberts, 1995). However, if final lengthening were a stronger cue to constituent boundaries than initial high pitch, in Experiment 4, we would have expected participants to perform better on sentences, signaled by final lengthening, than on phrases signaled by initial high pitch. Instead, the results of Experiment 4 show that participants still performed better on phrases. Thus the differences between the two conditions in Experiment 1, 2 and 4 must have been caused by either the fact that there were twice as many instances of phrases than of sentences or by the fact that phrases were shorter and thus required less memory capacity. In either case, participants segmented and grouped syllables together online according to specific prosodic cues, and they were able to keep track of multiple prosodic cues from different levels of the prosodic hierarchy.

The results of Experiments 1 and 2 are also suggestive of how participants processed final lengthening and pitch declination, two cues that in many languages signal constituent boundaries at two different levels of the prosodic hierarchy. In natural languages, as well as in our familiarization streams, the final lengthening of the last Phonological Phrase of an Intonational Phrase always coincides with the end of pitch declination. That is, each Intonational Phrase is, in fact, marked by two prosodic cues. Previous studies that have focused solely on speech segmentation have found that duration and pitch declination are either additive in the strength with which they signal boundaries (Streeter, 1978), or are perceived as a single percept (Beach, 1991). However, when we look at participants' performance on sentence-level rule testing in Experiment 1, we see that they did not perform better on sentence-level rule testing (where final lengthening and pitch declined coincided at the sentence final boundary) than on

phrase-level rule testing. Similarly, in Experiment 2, where final lengthening was no longer available as a cue to phrase boundaries in the sentence-level rule testing condition (condition 2), participants did not perform significantly worse than participants in Experiment 1 did on sentence-level rule testing. Thus the strength of a constituent boundary is not perceived as the sum of the two single prosodic cues (in this case, final lengthening and pitch declination). This is particularly interesting because short sentences may contain just one phrase, to which both pitch declination and final lengthening apply.

Instead, the finding that participants who were familiarized with both cues simultaneously (Experiment 1) did not perform significantly better than participants familiarized with one cue only (Experiment 2) suggests that listeners associated each cue with a specific level of the prosodic hierarchy and used the individual cues for finding constituent boundaries at their respective levels only. Participants who were familiarized exclusively with final lengthening could only find phrases, and not sentences (Experiment 2 conditions 1 and 2). Similarly participants who were familiarized exclusively with pitch declination could only find sentences, and not phrases (Experiment 2 conditions 3 and 4). The segregation of the individual prosodic cues may be necessary if we consider that prosody is also used for grouping the segmented units. By using two distinct prosodic cues to signal structural relations at the phrase-level (signaled by final lengthening) and at the sentence-level (signaled by pitch declination), we have shown that participants do not use prosody only for segmenting the speech stream but use it also for finding the structural relations between the segmented units at different levels of the prosodic hierarchy, and thus possibly also the syntactic hierarchy.

Importantly, the finding that participants discovered phrases and sentences also when the prosodic cues were reversed shows that their native language did not determine the hierarchical segmentation principles that participants used. It is possible that participants relied on purely acoustic/perceptual characteristics of sounds to segment and group syllables. Because the auditory system processes sound in a specific manner (e.g. the Iambic-Trochaic Law), we chose to use prosodic cues that are attested in the world languages but were unfamiliar to Italian-speaking participants. Had we decided to use prosodic cues unattested in world languages – e.g. used pitch increase and initial lengthening – it is very possible that participants would have found neither phrases nor sentences. It is therefore likely that prosodic-based segmentation is driven by low-level acoustic biases that may be universal. However, further research is necessary to determine how these biases relate to language and its hierarchical structure.

With respect to rule-learning, our results are in line with the findings of Peña et al. (2002), who demonstrated that statistical computations are powerful enough to segment continuous streams of syllables after a short familiarization (see also Aslin et al., 1998; Saffran, Aslin et al., 1996; Saffran, Newport et al., 1996), but that segmentation cues in the form of pauses are necessary for extracting higher order structural regularities (i.e. the long distance dependency rules) – an ability that emerges within the first year of life (Marchetto & Bonatti, in preparation a, in preparation b). However, while the structure of phrases in our experiments was identical to the structure of words used in Peña et al. (2002), our familiarization streams differed from those in Peña et al. in that the transitional probabilities between phrases were considerably higher (TP 1.0 instead of 0.5). Thus, in our experiments the non-adjacent dependencies (TPs between the first and the last syllable of each phrase and sentence) and the adjacent dependencies (across phrase boundaries) were always 1.0. Because adjacent dependencies are easier to learn than non-adjacent dependencies (cf. Bonatti et al., 2006; Newport & Aslin, 2004 for a discussion), participants familiarized with prosodically flat streams (Experiment 3) preferred part-phrases (that contained the adjacent dependency) to novel rule-phrases (that contained only the non-adjacent dependency).

Our findings not only agree with, but also extend, the results of Peña et al. (2002). Because participants in Peña et al. (2002) draw generalizations only when subliminal 25 ms long pauses were introduced between the basic units – words that contained the long-distance dependencies – the authors argued that prosodic constituent structure (signaled by the subliminal pauses) is a prerequisite for drawing generalizations from continuous speech (Bonatti et al., 2006). However, on the one hand, pauses between words are not systematically present in natural speech. In addition, pauses have been found to be unreliable cues for segmentation (cf. Fernald & McRoberts, 1995). Thus our experiments show that participants could draw generalizations also with more natural cues (final lengthening and pitch declination). This reinforces the idea that prosodic cues may be necessary for inducing structural generalizations from continuous speech. On the other hand, different prosodic cues signal different levels of the prosodic hierarchy even though the specific cues assigned to each of the levels may vary across languages (Nespor & Vogel, 1986; Selkirk, 1984). Our results thus also show that generalizations of the type shown in Peña et al. (2002) can be drawn on multiple levels of the prosodic hierarchy. This is an important feat in language acquisition because non-adjacent regularities are found at all levels of grammar. Among many others, between auxiliaries and inflectional morphemes (e.g. "*is* writ*ing*"); center embedded sentences (e.g. *Tom*, who is very shy, *kisses Mary*); as well as distant relations between sentential constituents (e.g. in *The children use the sand in the garden to play*, the final verb *play* does not refer to its adjacent constituent *garden* but rather to the initial constituent *children*) (c.f. Bonatti et al., 2006; Perruchet et al., 2004).

It is important to note that when we talk about rule learning, it is possible that our participants did not actually learn the long-distance dependency rules either at the phrase or at the sentence level. Participants could simply have remembered the first and the last syllables of the phrases and the sentences. Endress, Scholl, and Mehler (2005) have shown that repetition-based regularities are generalized only at the edges of syllable sequences, suggesting that edges of constituents are powerful cues for tackling the speech stream (c.f. Endress & Mehler, 2009; Endress, Nespor, & Mehler, 2009; Endress et al., 2005). Because in our experiment all the long-distance dependency rules coincided with the prosodic cues signaling phrase

and sentence boundaries, it is possible to successfully complete the task without actually having to compute that the first syllable (A) predicts the last syllable (C) with a probability of 1.0. Thus, because the rules at the phrase-level formed three distinct families (three different A × C rules) and so did the rules at the sentence-level (three different A–C rules), and each rule-family shared distinct initial and final syllables, participants' might have generalized far simpler rules than long-distance dependencies. For example, they could have detected the class of possible A syllables, the class of possible C syllables, and their relative position in the beginning and end of phrases/sentences (without requiring participants to learn the transitional probabilities between specific A–C syllable pairs). However, regardless of the precise mechanism underlying the rule generalization, our results demonstrate that participants are able to use the prosodic cues for extracting hierarchical regularities from the speech stream.

While our results are the first to demonstrate hierarchical rule learning with cues from different levels of the prosodic hierarchy, the idea that multi-level structure may be acquired from the speech stream is not new. Saffran and Wilson (2003) found that 12-month-old infants are able to segment a continuous speech stream using TPs between syllables and consequently also discover the order of the segmented words by using TPs between words. In addition, Kovács and Endress (in preparation) showed that seven-month-old infants can learn hierarchically embedded structures that are based on identity relations of words that followed a syllable repetition ("abb" or "aba", where each letter corresponds to a syllable) that formed sentences based on word repetitions ("AAB" or "ABB", where each letter corresponds to a word). The advantage of prosody, with respect to statistical computations and rule learning is that these latter processes depend on exposure to the speech stream that triggers cognitive processing for structure generalizations (c.f. Peña et al., 2002), whereas prosody can, in theory, be exploited on single trial learning because it relies on perceptual biases (c.f. Bion et al., 2011; Endress, Dehaene-Lambertz, & Mehler, 2007; Endress & Mehler, 2010; Endress et al., 2009).

Our findings complement this body of research with evidence in favor of hierarchical rule learning with cues that correspond to those present in actual speech. On the one hand, prosody provides suprasegmental cues for constituent boundaries at different levels of the prosodic hierarchy. Listeners have been shown to be able to segment speech at Intonational Phrase boundaries (i.e. Shukla et al., 2007; Watson & Gibson, 2004), Phonological Phrase boundaries (Christophe, Nespor, Guasti, & van Ooyen, 2003; Christophe et al., 2004; Millotte et al., 2008) as well as at Prosodic Word boundaries (Millotte et al., 2007). On the other hand, because different levels of the prosodic hierarchy use at least partially different prosodic cues, they additionally signal how the segmented units relate to each other. Because lower levels of the prosodic hierarchy are exhaustively contained in higher ones (Nespor & Vogel, 1986; Selkirk, 1984), it is possible to determine the Phonological Phrases contained in any given Intonational Phrase. Because prosody relies on perceptual, rather than computational mechanisms, it may provide a more direct mapping between the speech signal and the hierarchical structure it instantiates.

Does this mean that statistical computations are irrelevant to language acquisition? As discussed above, participants, who were familiarized with the prosodically flat stream, did use statistical computations for segmentation. Our results thus only suggest that prosody is a stronger cue to segmentation than transitional probabilities. Previous studies have shown that there is an interaction between statistical computations over syllables and detection of prosodic information. Shukla et al. (2007) demonstrated that transition probabilities are computed over syllables automatically, but prosodic cues (specifically, the declining pitch contour) are used as filters to suppress possible word-like sequences that occur across Intonational Phrase boundaries. In fact, similar effects have also been found with other language specific cues, such as the vocalic structure of words (Toro, Pons, Bion, & Sebastian Gallés, 2011). In our experiments, the preference for statistically well-formed phrase-like and sentence-like sequences (Experiment 3) was reverted to a preference for rule-phrases and novel rule-sentences when prosodic cues were present in the familiarization stream (Experiments 1 and 2). Thus final lengthening and pitch declination must be powerful cues to assign a constituent boundary where statistical computations would not assign one (TP 1.0 between phrases). Importantly, our experiments extend the findings of Shukla et al. (2007), who used only Intonational Phrase boundaries, to include final lengthening typical to Phonological phrase boundaries. Because pitch declination might be universally used for signaling Intonational Phrase boundaries and final lengthening has been suggested to be universally used to signal Phonological Phrase boundaries (Pierrehumbert, 1979; Price et al., 1991; Wightman et al., 1992), our findings suggest that prosody filters statistical computations at both the Intonational and the Phonological Phrase levels. Furthermore, the strength of the specific prosodic cues in filtering statistical regularities is evident when considering that the TP-cues used in our studies were extremely salient (1.0 for adjacent and non-adjacent dependencies), while prosodic cues fell within the range used in previous studies (c.f. Bion et al., 2011), and were within the limits of pitch and syllable durations in natural speech (c.f. Shukla et al., 2007). Finally, we do not currently know whether the presence of prosodic cues is blocking the computation of statistical information (like TPs), or the distributional properties are computed but are not taken into consideration in producing the outcome. A comparison to address this question – that we leave for further research – might be to contrast part-sentences and part-phrases against sequences that contain the syllables used in the familiarization streams in a order never appeared before.

The fact that prosody filters transitional probabilities at both the Intonational Phrase and the Phonological Phrase level may mean that transitional probabilities are primarily used for finding constituents below the Phonological Phrase level, i.e. for finding words rather than relations between words. This is supported by the differences between TPs within words and at word boundaries (Saffran, Aslin et al., 1996; Saffran, Newport et al., 1996). However, experimental studies have shown that Intonational Phrase

(Kjelgaard & Speer, 1999; Warren et al., 1995), Phonological Phrase (Christophe et al., 2004) and Prosodic word boundaries constrain lexical access (Millotte et al., 2007; Endress & Hauser, 2010). This may suggest that prosody filters transitional probabilities also at the word level.

While the results strongly suggests that prosodic cues are stronger than statistical cues, it is important to bear in mind that all the experiments reported in this paper relied on very brief familiarizations. Because TPs are calculated online, the majority of studies use considerably longer familiarization streams. Participants in our experiments may thus have failed to build strong enough statistical representations of words that could override the prosodic cues. While the current set of data does not rule out this possibility, the experiments of Peña et al. (2002) showed that increasing the duration of the familiarization sequences did not change participants' preference from prosody to TPs. Instead, longer familiarization streams increased participants' preference for prosodically segmented words than for TP defined words. This suggests that even with longer familiarization sequences participants might have used prosodic cues over statistical ones, a question open for future studies.

Importantly, we are not arguing against TPs. We believe our results constitute solid evidence for people's ability to group sequences of syllables hierarchically based on the well-documented principles of the ITL (e.g., high pitch marks the beginning of a grouping, and long duration marks the end of a grouping). These perceptual biases might be helpful in learning natural languages, as they are strongly correlated with different levels of the prosodic hierarchy. In our brief and deterministic experimental conditions, prosody trumps TPs, but in natural languages they are likely to work in combination. The fact that there is not a one-to-one mapping between prosody and syntax, can be overcome by the interaction of different cues to segment the continuous speech stream (Seidenberg, 1997), including TPs and the knowledge of familiar words.

The second important implication of the present study concerns the issue of how children get from prosodic to syntactic constituents. It is clear that there is no one-to-one mapping between prosody and syntax (see Cutler et al., 1997 for an overview) and that infants follow prosodic constituent boundaries even in the rare cases when these do not coincide with appropriate syntactic boundaries (Gerken, 1994; for a discussion see Gerken, 1996a, 1996b). This suggests that all the hierarchical relations extracted from the speech stream during the first and second year of life are those pertaining to prosody and infants treat prosodic boundaries just as if they were syntactic ones (Gerken, 1996a, 1996b).

There are of course differences between competence and processing of prosodic information. For example, while the stimuli of the experiments reported in the present paper had prosody in a one-to-one correspondence to the prosodic constituents determined by the rules, this is not always the case in real language. For example, while the end of a sentence is always aligned with the end of an Intonational Phrase, an Intonational Phrase boundary can occasionally be found within a sentence. This occurs when a sentence is too long to be uttered in a single breath. Notice, however, that sentences of this length are rare in normal speech, and even rarer in child directed speech (Fodor & Crowther, 2002) the type of speech to which infants pay most attention (e.g. Newport, Gleitman, & Gleitrnan, 1977; Fernald, 1992; Thiessen, Hill, & Saffran, 2005). This indicates that in infant directed speech Intonational Phrase boundaries are usually aligned with sentence boundaries. Thus the language-acquiring child will have to come to terms with the probabilistic nature of the cues in the speech signal (for a discussion see Gerken, 1996a, 1996b). Further research will have to address how language learners and listeners process the variation in prosody and how this interacts with statistical cues. Experiments where prosodic cues probabilistically signal syntactic boundaries may show that processing speech with natural prosody might rely more on TPs than artificial language experiments warrant. The findings reported above do not allow us to rule out this possibility.

How might the results presented above generalize to infants? Previous studies have shown that infants become sensitive to pitch declination by 6-months of age (Nazzi et al., 2000) and to final lengthening by 9-months of age (Jusczyk et al., 1992). Seven-month-old infants have been shown to use pitch cues to group syllables into constituents (Bion et al., 2011), suggesting that major prosodic cues are used for discovering relations between constituents before the end of the first year of life. Experimental evidence also shows that infants are capable of discovering multi-level structures using statistical regularities by 12-months of age (Saffran & Wilson, 2003) and algebraic multi-level rules as early as 7-months of age (Kovács & Endress, in preparation). This suggests that, around the end of the first year of life, infants begin to look for hierarchical relations in continuous speech. Importantly, Marchetto and Bonatti (in preparation b) showed that 12-month-old Italian speaking infants could generalize A–C structures similar to those used by Peña et al. (2002) after being exposed to a segmented stream, but not to a continuous stream. In contrast, 7-month-olds exposed to a segmented stream succeed in learning words, but failed to acquire word-internal structures. This suggests that prosodic boundary cues start triggering the generalization mechanism to learn grammar-like regularities around the time when sensitivity to all major prosodic constituents is well developed and infants have understood that the speech stream is hierarchically structured.

In conclusion, in four experiments, we have investigated whether at least part of the human ability to organize words into phrases and phrases into sentences may be achieved on the basis of the acoustic properties of the speech signal. We have shown that listeners can keep track of prosodic cues from different levels of the prosodic hierarchy, that they perceive prosody as organized hierarchically, and that they are able to use hierarchically structured prosody to acquire hierarchically organized rule-like regularities. These findings extend the role of prosody from providing cues to constituent boundaries to a powerful tool for extracting information concerning the relation among the segmented units in the speech stream. In other words, the information contained in the prosody of speech allows listeners to discover hierarchical constituent structure, one of the core properties of human language.

Appendix A.

HABITUATION PHASE
(The same for all experiments)

Phrases	Sentences
TO x FE se ko lu	TO x FEMU x GI se se ko ko lu lu
MU x GI se ko lu	BA x PETO x FE se se ko ko lu lu
BA x PE se ko lu	MU x GIBA x PE se se ko ko lu lu

TEST PHASE

Phrases (The same for all experiments)		Sentences		
RULE	PART	RULE	PART (Exp. 1 & 3)	PART (Exp. 2)
TO x FE mu gi ba	FEMU x se ko lu	BA x PETO x FE mu mu to gi fe ba	BA x PEMU x GI se se ko ko lu lu	GIBA x PETO x se se ko ko lu lu
MU x GI pe to fe	x FEMU se ko lu	MU x GIBA x PE pe mu to to fe fe	MU x GITO x FE se se ko ko lu lu	FEMU x GIBA x se se ko ko lu lu
BA x PE mu to fe	GIBA x se ko lu	TO x FEMU x GI mu pe gi to ba fe	TO x FEBA x PE se se ko ko lu lu	PETO x FEMU x se se ko ko lu lu
* Same "A" and "C" syllables as habituation phrases but a novel "x" syllable that had not occurred in this position before.	x GIBA se ko lu	* Same "A" and "C" syllables as habituation phrases but a novel "x" syllable that had not occurred in this position before.	* Two phrases that occurred in the habituation phase but did not form a sentence.	x PETO x FEMU se se ko ko lu lu
	PETO x se ko lu			x FEMU x GIBA se se ko ko lu lu
	x PETO se ko lu			x GIBA x PETO se se ko ko lu lu
	* The phrases violate the prosodically determined boundary. However, the sequence of syllables did occur in the habitation phase.			* The sentences violate the prosodically determined boundary. However, the sequence of syllables did occur in the habituation phase.

Acknowledgments

During the period the experimental work was carried out: Marina Nespor was additionally affiliated with the Interdisciplinary Center B. Segre, Accademia dei Lincei (Italy); Erika Marchetto and Ricardo Augusto Hoffmann Bion were affiliated with SISSA/ISAS – International School for Advanced Studies (Trieste, Italy). The research was funded by the European Science Foundation Eurocores OMLL grant, the Italian National Grants (COFIN) 2003–2007, the James S. McDonnell Foundation.

References

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

Bagou, O., Fougeron, C., & Frauenfelder, U. (2002). Contribution of prosody to the segmentation and storage of "words" in the acquisition of a new mini-language. In B. Bel & I. Marlien (Eds.), *Proceedings of the speech prosody 2002 conference* (pp. 59–62). Aix-en-Provence: Laboratoire Parole et Langage.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30*, 644–663.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 15–70.

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

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*, 341–345.

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

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2006). How to hit Scylla without avoiding Charybdis: comment on Perruchet, Tyler, Galland, and Peereman. *Journal of Experimental Psychology: General, 135*, 314–326.

Cambier-Langeveld, G. M., Nespor, M., & Van Heuven, V. (1997). The domain of final lengthening in production and perception in Dutch. *ESCA. Eurospeech proceedings*, 931–935.

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.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological Phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language, 51*, 523–547.

Collier, R., & 't Hart, J. (1975). The role of intonation in speech perception. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 107–121). Heidelberg: Springer-Verlag.

Collier, R., de Pijper, J. R., & Sanderman, A. A. (1993). Perceived prosodic boundaries and their phonetic correlates. *Proceedings of the DARPA Workshop on Speech and Natural Language* (pp. 341–345). Princeton, NJ, March 21–24.

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

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cooper, W. E., & Sorensen, J. M. (1977). Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America, 62*, 682–692.

Cooper, W. E., & Sorensen, J. M. (1981). *Fundamental frequency in sentence production*. New York: Springer-Verlag.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133–142.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*, 141–201.

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

de Rooij, J. J. (1975). Prosody and the perception of syntactic boundaries. *IPO Annual Progress Report, 10*, 36–39.

de Rooij, J. J. (1976). Perception of prosodic boundaries. *IPO Annual Progress Report, 11*, 20–24.

Demuth, K. (1995). Markedness and the development of prosodic structure. In J. Beckman (Ed.), *Proceedings of the north eastern linguistic society of Amherst*. MA: GLSA, University of Massachusetts.

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

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van Der Vreken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *Proc. ICSLP'96* (vol. 3, pp. 1393–1396). Philadelphia.

Endress, A. D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition, 105*, 577–614.

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology, 61*, 177–199.

Endress, A. D., & Mehler, J. (2009). Primitive computations in speech processing. *Quarterly Journal of Experimental Psychology, 62*, 2187–2209.

Endress, A. D., & Mehler, J. (2010). Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: HP&P, 36*, 235–250.

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Science, 13*, 348–353.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General, 134*, 406–419.

Fee, J. E. (1995). Exploring the minimal word in early phonological acquisition. In *Proceedings of the 1992 annual conference of the Canadian linguistic association.*

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). Oxford, England: Oxford University Press.

Fernald, A., & McRoberts, G. W. (1995). Prosodic bootstrapping. A critical analysis of the argument and the evidence. In J. Morgan & K. Demuth (Eds.), *Signal to syntax*. Hillsdale, NJ: Erlbaum.

Fodor, J. D., & Crowther, C. (2002). Understanding stimulus poverty arguments. *Linguistic Review, 19*, 105–145.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 862–877.

Gebhart, A. L., Newport, E. L., & Aslin, R. (2009). *Psychonomic Bulletin & Review, 16*, 486–490.

Gerken, L. A. (1994). Young children's representation of prosodic phonology: Evidence from English-speakers' weak syllable productions. *Journal of Memory and Language, 33*, 19–38.

Gerken, L. A. (1996a). Prosody's role in language acquisition and adult parsing. *Journal of Psycholinguistic Research, 25*, 345–356.

Gerken, L. A. (1996b). Prosodic structure in young children's language production. *Language, 72*, 683–712.

Gerken, L. A., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: Nine-month-old's sensititivy to phonological vs. syntactic phrases. *Cognition, 51*, 237–265.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *PNAS, 105*, 14222–14227.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese–Italian cross-linguistic study. *Cognitive Psychology, 57*, 56–74.

Gervain, J., & Werker, J. F. (2008). Frequency and prosody boostrap word order: A cross-linguistic study with 7-month-old infants. In *The 33rd Boston university conference on language development*, Boston, MA.

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological Phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language, 51*, 548–567.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*, 254–260.

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

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meaning: Infant statistical learning in a natural language. *Cognitive Psychology, 63*, 93–106.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Phonetics and phonology. Rhythm and meter* (Vol. 1, pp. 201–260). San Diego: Academic Press.

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

Hirsh-Pasek, K., Kemler Nelson, D., Jusczyk, P. W., Wright, K., Druss, B., & Kennedy, L. J. (1987). Clauses are perceptual units for young infants. *Cognitive Psychology, 24*, 252–293.

Hirst, D. (1993). Detaching intonational phrases from syntactic structure. *Linguistic Inquiry, 24*, 781–788.

Houston, D., Santelman, L., & Jusczyk, P. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes, 19*, 97–136.

Hyman, L. M. (1977). On the nature of linguistic stress. In L. M. Hyman (Ed.), *Studies in stress and accent* (Southern California Occasional Papers in Linguistics No. 4, pp. 37–82). Los Angeles: University of Southern California.

Inkelas, S., & Zec, D. (1990). *The phonology-syntax connection*. Chicago: The University of Chicago Press.

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

Johnson, E. K., & Jusczyk, P. W. (2003a). Exploring statistical learning by 8-month-olds: The role of complexity and variation. In D. Houston, A. Seidl, G. Hollich, E. Johnson, & A. Jusczyk (Eds.), *Jusczyk lab final report*. <http://hincapie.psych.purdue.edu/Jusczyk>.

Johnson, E. K. & Jusczyk, P. W. (2003b). Exploring possible effects of language-specific knowledge on infants segmentation of an artificial language. In D. Houston, A. Seidl, G. Hollich, E. Johnson, & A. Jusczyk (Eds.), *Jusczyk lab final report*. <http://hincapie.psych.purdue.edu/Jusczyk>.

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

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

Jusczyk, P. W. (1998). Dividing and conquering the linguistic input. In M. C. Gruber, D. Higgins, K. Olson, & T. Wysocki (Eds.), *CLS 34: The panels* (Vol. 2, pp. 293–310). Chicago: University of Chicago.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress pattern of English words. *Child Development, 64*, 675–687.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24*, 252–293.

Jusczyk, P. W., Hohne, E., & Mandel, D. (1995). Picking up regularities in the sound structure of the native language. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 91–119). Timonium, MD: York Press.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The headturn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*, 111–116.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40*, 153–194.

Klatt, D. H. (1974). The duration of [s] in English words. *Journal of Speech and Hearing Research, 17*, 51–63.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59*, 1208–1221.

Kovács, Á. M. & Endress, A. D. (in preparation). Seven-month-olds learn hierarchical "grammars".

Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics, 31*, 279–292.

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

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa, 7*, 107–122.

Lehiste, I. (1974). Interaction between test word duration and the length of utterance. *Ohio State University Working Papers in Linguistics,* 17, 160–169.

Lehiste, I., Olive, J. P., & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America, 60*, 1199–1202.

Marchetto, E., & Bonatti, L. L. (in preparation a). Infants' discovery of words and grammar-like regularities from speech requires distinct processing mechanisms.

Marchetto, E., & Bonatti, L. L. (in preparation b). Finding words and rules in a speech stream at 7 and 12 months.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule-learning in seven-month-old infants. *Science, 283*, 77–80.

Marslen-Wilson, W., & Tyler, L. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1–71.

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14*, 648–652.

Millotte, S., Frauenfelder, U. H., & Christophe, A. (2007). Phrasal prosody constraints lexical access. AmLap – 13th Annual conference on architectures and mechanisms for language processing, Turku, Finland.

Millotte, S., Rene, A., Wales, R., & Christophe, A. (2008). Phonological Phrase boundaries constrain the online syntactic analysis of spoken sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 874–885.

Morgan, J. L., Swingley, D., & Miritai, K. (1993). Infants listen longer to speech with extraneous noises inserted at clause boundaries. *Paper presented at the biennial meeting of the society for research in child development*, New Orleans, LA.

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology, 13*, 477–492.

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six month olds detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy, 1*, 123–147.

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

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

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance. I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127–162.

Newport, E., Gleitman, H., & Gleitrnan, L. (1977). Mother, I'd rather do it myself. In C. Snow & C. Ferguson (Eds.), *Talla'ng to children*. Cambridge: Cambridge University Press.

Nooteboom, S. G., Brokx, J. P. L., & de Rooij, J. J. (1978). Contributions of prosody to speech perception. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language* (pp. 75–107). Chichester: John Wiley & Sons.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America, 54*, 1235–1246.

O'Shaughnessy, D. (1979). Linguistic features in fundamental frequency patterns. *Journal of Phonetics, 7*, 119–145.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development, 80*, 674–685.

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

Peña Bonatti Nespor & Mehler (2002). Signal driven computations in language processing. *Science, 298*, 604–607.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General, 133*, 573–583.

Peters, A. M. (1985). Language segmentation: Operating principles for the perception and analysis of language. In Slobin, D. J. (Ed.), *The crosslinguistic study of language acquisition* (Vol. 2.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of Acoustic Society of America, 66*, 363–369.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of Intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: The MIT Press.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*, 2956–2970.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1925–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*, 273–284.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29*, 169–182.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America, 71*, 996–1007.

Sebastian, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish: The role of stress. *Journal of Memory and Language, 31*, 18–32.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599–1604.
Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: The MIT Press.
Selkirk, E. (1996). The prosodic structure of function words. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–213). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*, 193–247.
Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology, 54*, 1–32.
Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *PNAS, 108*, 6038–6043.
Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49*, 249–267.
Speer, S. R., Warren, P., & Schafer, A. J. (2011). Situationally independent prosodic phrasing. *Laboratory Phonology, 2*, 35–98.
Steedmna, M. (1990). Syntax and intonational structure in a combinatory grammar. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 457–482). Cambridge, MA: MIT Press.
Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America, 64*, 1582–1592.
Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant directed speech facilitates word segmentation. *Infancy, 7*, 53–71.
Thiessen, E. D., & Saffran, J. R. (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.
Toro, J. M., Pons, F., Bion, R. A. H., & Sebastian Gallés, N. (2011). Statistical computations over linguistic stimuli are constrained by suprasegmental rules. *Journal of Memory and Language, 64*, 171–180.
Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America, 61*, 846–858.
Vaissiere, J. (1974). On French prosody. *Quarterly Progress Report, MIT, 114*, 212–223.
Vaissiere, J. (1975). Further note on French prosody. *Quarterly Progress Report, MIT, 115*, 251–262.
Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 53–66). Berlin: Springer Verlag.
Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology and parsing in closure ambiguities. *Language and Cognitive Processes, 10*, 457–486.
Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19*, 713–755.
Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America, 91*, 1707–1717.
Woodrow, H. (1951). Time perception. In S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224–1236). New York: Wiley.
Yang, C. (2004). Universal grammar, statistics or both. *Trends in Cognitive Sciences, 8*, 451–456.
Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., et al. (2010). The development of perceptual grouping biases in infancy: A Japanese–English cross-linguistic study. *Cognition, 115*, 356–361.