Early sensitivity to prosodic phrase boundary cues: Behavioral evidence from German-learning infants

Dissertationsschrift
zur Erlangung des akademischen Grades Doctor philosophiae
im Fach Kognitionswissenschaft

eingereicht bei der
Humanwissenschaftlichen Fakultät
der Universität Potsdam
im Juni 2022

vorgelegt von
Caroline Wellmann

Supervision: Prof. Dr. Barbara Höhle

1st reviewer: Prof. Dr. Barbara Höhle
2nd reviewer: Prof. Dr. Ruben van de Vijver

Day of submission: June 30, 2022
Thesis defense: Potsdam, November 23, 2022

https://doi.org/10.25932/publishup-57393
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-573937

ACKNOWLEDGEMENTS

Ich möchte mich an dieser Stelle herzlich bei denjenigen bedanken, die mich bei der Erstellung der Dissertation unterstützt haben.

An erster Stelle gilt mein Dank Barbara Höhle, die mich als Erstbetreuerin über Jahre kontinuierlich gefördert und begleitet hat. Ihre wertvollen Ideen, Kommentare und Ausführungen haben essentiell zur Entstehung dieser Dissertation beigetragen und meine Sichtweise auf Phänomene des frühen Spracherwerbs maßgeblich geprägt.

Ein Großteil dieser Arbeit ist zwischen 2010 und 2012 in einem Projekt zur Entwicklung prosodischer Fähigkeiten im frühen Spracherwerb innerhalb des DFG-Schwerpunktprogramm *Sprachlautliche Kompetenz: Zwischen Grammatik, Signalverarbeitung und neuronaler Aktivität* entstanden. Ein besonderer Dank geht an Julia Holzgrefe-Lang für die enge Zusammenarbeit in diesem Projekt, für das gemeinsame Konzipieren unserer Untersuchungen, für den konstruktiven Austausch und für ihre zuverlässige Unterstützung in den letzten Monaten des Schreibens. Ich bin froh, dass wir auch weiterhin persönlich verbunden sind. Ein herzlicher Dank gilt Isabell Wartenburger, die als Co-Projektleiterin und Zweitbetreuerin meiner Dissertation für alle fachlichen und praktischen Fragen immer ansprechbar war und konstruktives Feedback und Ideen gegeben hat. Ich danke Caterina Petrone und Hubert Truckenbrodt, die im gemeinsamen Projekt bei Fragen zur Phonetik und Phonologie geholfen haben.

Ich danke der Arbeitsgruppe Spracherwerb für den anregenden Austausch im Kolloquium, darüber hinaus besonders Tom Fritzsche für seine Unterstützung bei der technischen Umsetzung der Experimente, für Ideen zur Auswertung der Daten und für die großartige Organisation des BabyLab. Weiterhin danke ich Natalie Boll-Avetisyan für Feedback zu den Untersuchungen und für die gemeinsame Weiterentwicklung der Thematik in Zusammenarbeit mit Saioa Larrazza, Sandrien van Ommen, Ranka Bijeljac-Babic, Thierry Nazzi und Barbara Höhle.

Ich danke Ruben van de Vijver für den konstruktiven Austausch im Rahmen des Schwerpunkt-programms und seine Hilfe bei der Umsetzung von PRAAT-Skripten. Sandra Hanne sei gedankt, nicht nur für ihre intensive fachliche Unterstützung, sondern auch für ihre motivierenden Worte, die mich bei der Fertigstellung begleitet haben.

Eine große Unterstützung in der Organisation und der Durchführung der Untersuchungen habe ich durch das Team des Potsdamer BabyLab erfahren. Ich danke dafür insbesondere Anne Beyer, Sophia Fischer, Babette Graf, Teresa Leitner, Mareike Orschinsky und Marie Zielina. Des Weiteren geht ein herzlicher Dank an Romy Räling für das Einsprechen der Stimuli. Danke an Sophia Jähnigen and Bernadette Smolibocki für ihre Unterstützung bei der akustischen Manipulation der Stimuli.

Sehr viele Familien haben sich auf den Weg ins BabyLab gemacht, um bereitwillig und interessiert mit ihren Kindern an den Studien teilzunehmen. Ihnen gilt ein besonderes Dankeschön, denn ohne sie wäre diese Arbeit nicht möglich gewesen.

Im Verlauf meines Promotionsprojektes sind meine Kinder geboren und haben diese Zeit auf besondere Weise bereichert. Ich bin dankbar für Zacharias, Nathanael und Rahel, die mich mit ihrer Fröhlichkeit und Leichtigkeit immer wieder ins Hier und Jetzt zurückholen. Ich danke meinem Mann Manuel, für beratende Worte, Korrekturlesen, Diskussionen und die intensive Begleitung in der Schlussphase der Promotion, aber vor allem dafür, dass wir gemeinsam durchs Leben gehen.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADS	Adult-Directed Speech
AE	American English
CPS	Closure Positive Shift
EP	European Portuguese
ERP	Event Related Potentials
H	High Boundary Tone
HPP	Headturn Preference Procedure
IDS	Infant-Directed Speech
IP	Intonation Phrase
IPB	Intonation Phrase Boundary
ISI	Interstimulus Interval
ITL	Iambic-Trochaic Law
L	Low Boundary Tone
ms	milliseconds
N	Name
NP	Noun Phrase
PP	Phonological Phrase
SD	Standard Deviation
SE	Standard Error
s	seconds
VP	Verb Phrase

1 Introduction

Prosody, the rhythm and the melody of speech, is expressed by variations in fundamental frequency, duration, pausing, intensity, and spectral characteristics. In auditory terms, these acoustic features refer to pitch, lengthening, loudness, and voice quality (e.g., Nespor & Vogel, 1986; Vaissière, 1983). The prosodic characteristics of an utterance convey social and communicative aspects such as emotions, attitudes, and turn-ends. Prosody also bears important linguistic functions signaling lexical and tonal stress, sentence types, and syntax as well as information structure.

Prosody is a key source for the language processor for chunking the continuous speech stream into linguistically relevant units. Correspondingly, prosodic information helps to solve ambiguities in spoken language (for comprehensive reviews see Cutler et al., 1997; Wagner & Watson, 2010). For example, the intended meaning of the wording “fruit salad and milk"[1] can only be interpreted if prosody is taken into account: while the presence of an internal prosodic boundary after the first noun indicates the meaning of a list of three simple nouns (‘fruit salad and milk'), the same expression will be referred to a list of a compound noun and a simple noun ('fruit-salad and milk') in the absence of such a boundary. This dissertation focusses on those prosodic features that indicate boundaries and play a significant role in speech segmentation.

Since prosodic characteristics apply to elements larger than individual phonological segments, prosody is also referred to as *suprasegmental phonology*. A theory on the organization of prosody is the *Prosodic Hierarchy* developed by Selkirk (1978, 1984, 1986) and extended by Nespor and Vogel (1986). Following the observation that phonological rules apply within certain domains, this theory assumes that utterances are phonologically phrased and that this phrasing is hierarchically organized. The model of the Prosodic Hierarchy proposed by Nespor and Vogel (1986) assumes the following seven prosodic levels:

Utterance
|
Intonation Phrase
|
Phonological Phrase
|
Clitic Group
|
Prosodic Word
|
Foot
|
Syllable

---

[1] The example is taken from the subtest “Chunking input function” of the intonation test battery PEPS-C (Wells & Peppé, 2003).

At each of these levels a relative prominence between components is assumed with one component being strong in relation to the others. According to the *Strict Layer Hypothesis* (Selkirk, 1984) each prosodic constituent *n* dominates the constituent(s) at the next lower level *n-1*, none of the levels may be skipped. Moreover, a well-formed prosodic structure does not show recursivity, that is, a constituent *n* may not dominate another constituent of the same type *n*.

Even though there is no strict one-to-one mapping, the Prosodic Hierarchy strongly interacts with the syntactic hierarchy, specifically at the levels of the *Intonation Phrase (IP)* and the *Phonological Phrase (PP)*. The edges of IPs and PPs that are acoustically well marked by means of fundamental frequency (F0, henceforth interchangeably used with pitch), duration, and intensity typically coincide with clausal and phrasal syntactic boundaries, respectively. The acoustic correlates of prosodic boundaries may thus deliver essential information to discover the edges of syntactic units within fluent speech, a pivotal task in language acquisition.

Infants easily and quickly master essential steps in language acquisition within the first two years of life. Psycholinguistic research in the last four decades has focused on how infants get into the system of the target language. Equipped with some cognitive mechanisms infants derive linguistic knowledge from the linguistic input. F0, duration, and intensity may initially present mere acoustic information that can be easily noticed by the mammalian auditory system (as reflected in newborns' language discrimination based on prosodic features, e.g., Nazzi, Bertoncini, & Mehler, 1998), but they have different functional roles across languages. The infant must thus learn that these acoustic features are essential prosodic boundary markers that may offer a first access into the underlying grammatical organization. Identifying the relevant acoustic correlates of major prosodic boundaries in the early acquisition of German and how this potentially relates to *perceptual attunement* is the key topic of this dissertation.

The term *perceptual attunement* has been introduced by Aslin & Pisoni (1980) to describe the connection between early speech perception abilities and the linguistic input. Within this connection, language experience modifies existing categories (partially developed at birth) according to the properties of the native language. Specifically, with increasing exposure to the ambient language, infants may show a pattern of *decline*, *maintenance* or *enhancement* in their perceptual sensitivities. A typical pattern is a decline in discriminating non-native phonemic contrasts observed by the second half of the first year of life (e.g., Polka & Werker, 1994; Werker & Tees, 1984; note that decreased sensitivity can be re-instated or at least improved by systematic training, e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). Language experience may also have a facilitative role enhancing sensitivity for consonant as well as vowel discrimination (e.g., Narayan, Werker, & Beddor, 2010, but see Sundara et al., 2018). Moreover, maintenance of discrimination of non-native sound contrasts without exposure to the critical language has also been found (e.g., Best, McRoberts, & Sithole, 1988).

So far, a large body of studies has focused on perceptual attunement in segmental phonology (for a review see Maurer & Werker, 2014). Some studies targeted suprasegmental phonology at the word level tracking perceptional reorganization of lexical stress (Bijeljac-Babic, Serres, Höhle, & Nazzi, 2012; Friederici, Friedrich, & Christophe, 2007; Frota, Butler, Uysal, Severino, & Vigário, 2020; Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Jusczyk, Cutler, & Redanz, 1993; Sansavini, Bertoncini, & Giovanelli, 1997; Skoruppa, Cristià, Peperkamp, & Seidl, 2011; Skoruppa et al., 2013, 2009), lexical tones (Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008; Shi, 2010; Yeung, Chen, & Werker, 2013), and lexical pitch accents (Sato, Sogabe, & Mazuka, 2010). Still, little is known about the developmental path of the perception of prosody at the phrasal level. There is one study on prosodic characteristics distinguishing the intonation contrast in questions versus statements with European-Portuguese-learning infants revealing maintenance (Frota, Butler, & Vigário, 2014). Studies on infants' perception of the acoustic correlates of prosodic boundaries are scarce and it is an open question whether prosodic boundary processing is subject to perceptual attunement.

A first step shedding light on infants' perceptual reliance on individual prosodic boundary cues across the first year of life as well as across languages was made by Seidl and colleagues with a series of three studies (Johnson & Seidl, 2008; Seidl, 2007; Seidl & Cristià, 2008). A clause segmentation task revealed differences in the necessity of the pitch versus pause cue in four- versus six-month-old American-English-learning infants and in American-English- versus Dutch-learning infants. This suggests that prosodic boundary cue perception develops language-specifically, hence undergoes some kind of perceptual attunement. Apparently, infants need to learn which acoustic correlates are important for the detection of boundaries in their native language. Seidl et al.'s studies were fundamental to the empirical studies planned and conducted in the framework of this dissertation. Meanwhile, two more studies investigating infants' discrimination of prosodic boundaries, one on French (van Ommen et al., 2020), the other on European Portuguese (Frota, Butler, Severino, Uysal, & Vigário, 2019) were added to the field, both studies supporting the view of a language-specific development. French infants' boundary perception proved to depend on the presence of the pause cue, whereas for European-Portuguese-learning infants pitch and lengthening were sufficient. Importantly, van Ommen et al. (2020) tested French and German infants with the same French materials revealing differences in infants' sensitivity to boundary cues depending on the language. In turn, van Ommen et al. concluded that prosodic boundary processing is affected by perceptual attunement, reflected as a process of enhancement with infants sharpening their sensitivity to individual cues and to more subtle ones.

This dissertation encloses three independent empirical studies that focus on the role of the acoustic correlates of major prosodic boundaries, specifically changes in pitch, final lengthening, and pause, in the early acquisition of German. When referring to the critical boundaries under investigation I will use the term *Intonation Phrase Boundary (IPB)* interchangeably with *(major) prosodic boundary*. The main research questions tackle the relation of phrasal prosody and developmental speech perception:

- What role do the acoustic correlates of prosodic boundaries play in boundary perception in early language acquisition? Specifically, which of the individual acoustic features are necessary to trigger boundary perception in German-learning infants and is there a change in infants' sensitivity during the first year of life?
- Relating the present findings to the outcome of studies from other languages: Is the perception of phrasal prosody a matter of perceptual attunement within the first year of life?
- Beyond, does infants' initial processing of boundary cues more likely relate to a universal processing mechanism or do infants apply language-specific strategies from the beginning on?

The structure of this dissertation is the following: subsequent to this introduction, Chapter 2 provides the empirical background that underlies the experimental work of this dissertation. First, the *prosodic bootstrapping* hypothesis will be introduced (Chapter 2.1) followed by a review of empirical research that supports its assumptions. The acoustic characteristics marking IPBs in infant- as well as adult-directed speech will be described (Chapter 2.2). Studies from the last four decades on infants' sensitivity (Chapter 2.3) and infants' exploitation of prosodic boundary information (Chapter 2.3) will be summarized. In Chapter 3, a description of prior studies examining a potential weighting of the acoustic correlates of prosodic boundaries builds the starting point for the empirical part of this dissertation. The research questions for each of the three studies of the empirical part are spelled out here. Study I splits up in Chapters 4 and 5. Chapter 4 refers to the peer-reviewed journal article published as *Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012) How each prosodic boundary cue matters: Evidence from German infants. Frontiers in Psychology 3 (580).* Four experiments with German-learning eight-month-old infants systematically examining the impact of individual prosodic boundary cues are reported. Chapter 5 adds a fifth, so far unpublished experiment, to the study described in Chapter 4 that looks into the quality of boundary cues, naturally occurring versus artificially inserted. In Chapter 6, Study II is presented, a peer-reviewed book chapter published as *Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (forthcoming) Developmental changes in prosodic boundary cue perception in German-learning infants. In: Schuboe, F., Hanne, S., Zerbian, S., & Wartenburger, I. (Eds.) Prosodic boundary phenomena. Studies in Laboratory Phonology. Berlin: Language Science Press.* It addresses the question whether boundary cue perception undergoes perceptual attunement during the first year of life. Four experiments will be reported that investigate six-month-old infants' perception of specific cue combinations. Chapter 7 relates to Study III, a pilot study that aims at uncovering whether infants make use of the identified prosodic cues for segmenting the speech stream. Three experiments were conducted searching for a suitable design for testing eight-month-olds' recognition of sequences indicated by prosodic grouping.

Finally, Chapter 8 completes the dissertation by elaborating on the main research questions. The present findings will be discussed in relation to those from other languages. Moreover, in search for the underlying mechanism of prosodic boundary processing the relevance of the findings concerning the *acoustic salience hypothesis* and the *language-specific hypothesis* of prosodic bootstrapping will be debated.

2 Prosody and early language acquisition

Prosodic information is accessible to the fetus before birth. The fetus is able to hear at around the 25th week of gestation. The acoustic signal of speech that reaches the fetus in the womb is low-pass filtered by maternal tissues and amniotic fluid at about 300 to 400 Hz (e.g., Lecanuet & Granier-Deferre, 1993; for overviews of auditory development see Mattock, Amitay, & Moore, 2012; Pujol, Lavigne-Rebillard, & Lenoir, 1998). Acoustic details allowing phoneme identification are missing in this signal, but the contour of fundamental frequency, and thus the overall prosody, is preserved. This allows the fetus to be exposed to the melody and rhythm of her native language at least in the last trimester of gestation. Several studies in the last decades have provided evidence that this prenatal experience is already shaping infants' early perception (for a recent review Gervain, 2018). Newborns have been shown to be highly sensitive to prosodic characteristics. For example, they show a preference for their native language over other rhythmically different languages (Mehler et al., 1988; Moon, Cooper, & Fifer, 1993). Newborns even derive specific prosodic knowledge from the prenatal language exposure as for example revealed by a recent study by Abboub, Nazzi, and Gervain (2016) on prosodic grouping. The *Iambic-Trochaic Law (ITL)* asserts that sounds contrasting in duration are naturally grouped iambically, whereas sound sequences contrasting in pitch and duration are perceived as trochees. For French, a language with final prominence that is mainly cued by duration, a short-long (i.e. iambic) grouping is the typical pattern. Abboub et al. (2016) presented French monolingual newborns with tone sequences that were either consistent with the predictions from the ITL or not. Their brain responses mirrored a distinction between ITL-consistent and non-consistent groupings of the tone sequence when the grouping was cued by duration (short-long vs. long-short), but not when it was cued by pitch (low-high vs. high-low) or intensity (strong weak vs. weak-strong). Hence, already at birth French monolinguals exhibit a preference in prosodic grouping for duration, the most important cue for rhythm in French. Interestingly, a pitch-based (i.e. trochaic) prosodic grouping preference was observed in French newborns growing up in a bilingual environment (French and one other language that makes more use of pitch in signaling prominence than French does). This suggests that newborns have already developed perceptional biases that are conform to what they have heard in the uterus. This is just one example of how prenatal language exposure modulates early abilities in speech perception.

2.1 Prosodic bootstrapping

Pursuing the very first steps that infants take into language development, Gleitman and Wanner (1982) proposed the *prosodic bootstrapping hypothesis*. This hypothesis assumes that the learner exploits prosodic information available in the input to derive knowledge about the lexical and syntactic properties of the ambient language that are not directly available in the input. Specifically, infants are expected to use prosodic cues like stress, rhythm, and intonation to segment the continuous speech stream into linguistically relevant units such as words and clauses (Gleitman & Wanner, 1982; Morgan & Demuth, 1996). The prosodic bootstrapping hypothesis requires three assumptions to hold (Jusczyk & Kemler Nelson, 1996): first, specific cues are available in the speech signal, specifically, the input provides acoustic cues of correlations between prosodic and syntactic boundaries, a close syntax-prosody mapping. Second, infants are highly sensitive to these cues from early on. Third, the young learner indeed makes use of these cues in order to segment the speech signal. In this dissertation, I will focus on those aspects of prosodic bootstrapping that handle phrasal prosody and its relation to syntactic segmentation (for a comprehensive review on infants' sensitivity to and use of prosodic information see Gervain, Christophe, & Mazuka, 2020 and Schmitz, 2008).

Across languages, large and intermediate prosodic units such as the *Intonation Phrase (IP)* and the *Phonological Phrase (PP)* as assumed by Nespor & Vogel (1986) are aligned with the edges of syntactic constituents (Cooper & Paccia-Cooper, 1980). A learner exploiting phrasal prosody can thus access the boundaries of the underlying syntactic units such as clauses through the acoustic correlates of prosodic boundaries available in the speech stream.

The prosodic bootstrapping hypothesis has gained much support from numerous studies in the last four decades. Nevertheless, there are still several issues demanding further research. Höhle (2009) notes that it is still open how infants identify the relevant cues for their native language given crosslinguistic variation. Correlations between prosody and grammar or lexicon exist universally, but the acoustic cues that signal these correlations may vary across languages. Gervain et al. (2020) demand further research to understand how and which of these correlations infants pursue and how they integrate different bootstrapping mechanisms to acquire different languages. Contributing to the field of crosslinguistic variation of prosody in early language acquisition, the studies conducted within this dissertation will examine the role of individual cues that signal prosodic boundaries in the early acquisition of German. These investigations will expand the research on the emergence of language-specificity in the perception of prosodic boundary information.

Before turning to these studies, the subsequent sections will sketch previous research on prosodic boundary cues – concerning their presence, perception, and exploitation in early language acquisition supporting the assumptions of the prosodic bootstrapping hypothesis.

2.2 The acoustics of prosodic phrase boundaries

Across languages, prosodic boundaries are marked by acoustic cues such as final (or preboundary) lengthening, initial strengthening, glottalization, pauses, changes in the F0 contour and in intensity (e.g., Petrone et al., 2017; Vaissière & Michaud, 2006; Wagner & Watson, 2010). Among these, pitch, final lengthening, and pause are best studied and often referred to be the most important for boundary perception.

The pitch cue comprises the F0 contour preceding a boundary expressed by a high (H) or low (L) boundary tone as well as resets in F0 across the boundary. Final lengthening relates to the duration of preboundary segments such as the rhyme, the syllable or just the vowel that may be lengthened before the boundary, that is, at the right edge of a unit, in relation to the same segment at a phrase-internal position. Pauses are silent breaks of the speech stream.

These three main boundary cues often appear together at the edges of IPs that correspond to clauses. In contrast, smaller syntactic constituents that correspond to PPs may be marked by final lengthening and pitch only. Moreover, the phonetic magnitude of each cue may increase from minor to major boundaries (Nespor & Vogel, 1996). Pitch change, preboundary lengthening, and pause may occur as individual single cues or in combinations of three or only two cues. The distribution of the occurrence of prosodic boundary cues and its degree may differ across languages, for example preboundary lengthening of a significant degree is present in English, French, German, Spanish, Italian, Russian, and Swedish, whereas in Finnish, Estonian, and Japanese there is little (if any) final lengthening (Vaissière, 1983).

2.2.1 Prosodic boundary characteristics in adult-directed speech

For *adult-directed speech (ADS)*, several studies have revealed English and German speakers' use of final lengthening, pitch changes, and pauses in order to disambiguate ambiguous algebraic expressions such as "A plus O times E" or lists of names such as "Anna and Bill or Mary" (as for example for German: Féry & Kentner, 2010; Gollrad, 2013; Gollrad et al., 2010; Huttenlauch et al., 2021; Petrone et al., 2017; and for English: Lehiste, Olive, & Streeter, 1976; Scott, 1982; Streeter, 1978; Wagner, 2005), but also in (semi)spontaneous speech the employment of these cues has been reported (for German: Féry, Hörnig, & Pahaut, 2011; Peters, Kohler, & Wesener, 2005).

I will briefly summarize two studies on the production of phrase boundaries by speakers of German: Petrone et al. (2017) as representative for studies using disambiguation tasks in a laboratory context, linking closely to the empirical part of this dissertation, and Peters et al. (2005) reporting boundary cueing in natural speech.

Petrone et al. (2017) examined the employment of F0 range, lengthening, and pause in the phrasing of bracketed lists of names (N). The names "Lena, Lilli, Lola, Manu, Mona, Nina" were concatenated in six different combinations with each name occurring once as N1, N2 and N3. These combinations of names occurred in three different conditions of bracketing: *Late* (N1 N2) N3, e.g., "(Lola oder Mona) und Lena"*; Early* N1 (N2 N3), e.g., "Lola (oder Mona und Lena)" and the control condition *Word* N1 (N2Ur- N3), e.g., "Lola (oder Mona Urlena)" in which a Prosodic Word boundary is expected after

"Mona" compared to a PP boundary at the same position in the *Early* condition, and an IP boundary in the *Late* condition.

The experimenter had the same list of stimuli without brackets. Participants were asked to produce the stimuli in such a way that the experimenter could guess the intended bracketing. Significant differences in the realization of the three cues were found and are here reported with respect to the *Late* and *Early* condition. Final lengthening was measured as the duration of the final vowels in N1 and N2. At the position of the brackets, that is, preceding a major boundary, this vowel was lengthened with a factor of around 1.8 up to 2 compared to the same position in the condition in which there was no bracket, that is, at a PP boundary. F0 cues varied across speakers, the F0 contour employed in most of the cases at the position of the brackets were rising contours on N1 and N2 with final F0 peaks that are high boundary tones. Moreover, 80% of the brackets were cued by pauses. Pause durations at the position of the brackets were on average 119 ms (*Late*) and 174 ms (*Early*). Only few, short pauses of around 50 ms occurred at non-bracket positions. Previous research (Truckenbrodt, 2002; Gollrad, 2013) has shown that pauses frequently occur between two coordinated clauses, a classical IP boundary position, but rarely and if at all only of short duration clause-internally at a PP boundary. For their production data, Petrone et al. concluded that the characteristics at the positions of the brackets, increased duration of the preboundary vowel, higher pitch and the presence of a pause, also indicated the boundary of IPs. These features are congruent with Peters et al.'s (2005) results on prosodic marking of boundaries in spontaneous German ADS.

Peters et al. (2005) analyzed phrase boundary markings in adult-directed utterances taken from the *German Kiel Corpus of Spontaneous Speech*. Differently to the task in Petrone et al.'s study that explicitly forced speakers to employ means of prosody for the purpose of disambiguation, speakers in Peters et al.'s study held a spontaneous dialogue on making an appointment. In the analysis of this conversation, all auditory breaks that occurred within the continuous speech stream were classified as phrase boundaries without any specification regarding its strength or its level within the prosodic hierarchy, but with a classification of turn-internal versus turn-final boundaries with or without hesitations and sentence ruptures. Here, only those breaks that occurred turn-internally and that were not subject to hesitation or syntactic discontinuity are considered (2470 of all 4583 auditory breaks). They may refer to IPs as well as PPs. Peters et al.'s analysis compared the cues marking these boundaries with respect to their frequency of occurrence (Peters et al. 2005: 159, Table 8).

Regarding the overall occurrences, that is, the occurrence of a cue in combination with (an) other cue(s) and as a single cue, the pitch cue was the most frequent one. Comprising the F0 shape preceding the boundary as well as the resets after it, it occurred at 74 % of all boundaries. Final lengthening was present in 66 % of the boundaries. Pause was the least frequent marking occurring only at 38 % of all boundaries. Cues were often – at 62 % of all boundaries – combined. The co-occurrence of pitch and lengthening (24.6 %) and the coalition of all three cues (23.7 %) were the most frequent. Cue combinations of pause and one additional cue were comparatively infrequent: 8.4 % of all boundaries were marked by a combination of pause and lengthening, and 4.9 % by a combination of pause and pitch. Each of the three cues also occurred as a single cue, that is, as the sole marker of a boundary: pitch was the most frequent single cue, marking 20.8 % of all boundaries, while lengthening alone cued 9.4 %, and pause alone only 1.3 %. Taking into account all frequencies, Peters et al.'s (2005) data revealed pause to be the least frequent and the combination of pitch and lengthening to be the most frequent marker.

In sum, in laboratory as well as natural spontaneous ADS, (not only) German speakers' use F0, final lengthening, and pauses to mark boundaries. Pauses are specific indicators for IPs, F0 and lengthening are applied across different levels of the prosodic hierarchy with varying phonetic magnitude.

2.2.2 Prosodic boundary characteristics in infant-directed speech

Prosodic boundary markings have been shown to be present and even more pronounced in *infant-directed speech (IDS)*. Briefly summarized, compared to ADS, IDS – also called motherese – is characterized by an overall slower speech rate, shorter sentences, higher pitch, larger pitch range and variability, stronger preboundary lengthening, and larger vowel space (for a review, see Cristia, 2013). Infants prefer to listen to IDS over ADS (Frank et al., 2020). IDS carries emotional and social contents that attract infants' attention, increases arousal and may in turn enhance cognitive learning. Proposing the *hyperspeech hypothesis,* Fernald (2000) goes one step further and claims that the linguistic structures themselves available in IDS make the learning task easier. Enhancing prosodic boundary cues in IDS would thus help the infant learner to segment the continuous speech stream into smaller units. There are numerous studies on IDS, many of them highlighting prosodic enhancement in IDS in rather global prosodic characteristics, but unfortunately, studies that focus on the use of the three individual boundary cues pitch, lengthening, and pause in IDS at boundary versus non-boundary positions are scarce.

A pioneering study by Fernald et al. (1989) compared different measures of F0, pause, and sentence duration in parents' interactions with their child (IDS) and in conversations with the other parent (ADS). Across six languages, French, Italian, German, Japanese, British and American English, the results were rather consistent. In IDS, F0 values were higher and more variable, utterances were shorter and pauses longer. Language-specificity showed up for example in American English mothers producing the strongest exaggerations in F0.

Recently, prosodic cues at boundary positions were specifically addressed in a study conducted by Ludusan, Cristia, Martin, Mazuka, & Dupoux (2016). Ludusan and colleagues examined differences in markings as well as in the detection of boundaries in IDS compared to ADS. In the first part of their study, they carried out a systematic literature review. They selected eleven papers on different languages (several on American and British and Afro-American English, one on German, one on Mandarin Chinese, and the one by Fernald et al., 1989 on six different languages, see above). From these papers, they retrieved and compared 36 samples of paired IDS-ADS measurements (i.e., for both registers from the same speaker) on pause duration, pitch change and vowel duration. This comparison revealed that pauses were longer in IDS than in ADS, also the preboundary vowel duration was increased in most cases of IDS compared to ADS. For pitch, the pattern was mixed as some studies found larger values for pitch reset in IDS and others found larger values for ADS. In the second part of their study, Ludusan et al. (2016) carried out an analysis of a large corpus on 22 paired samples of IDS and ADS in Japanese (RIKEN corpus). Here, each of the three cues proved to be stronger across two prosodic units than within a unit in both, IDS and ADS. For pause and vowel duration the values were higher in IDS than in ADS. Pitch resets were larger in ADS. Altogether, the authors concluded from these acoustic characteristics that – in line with the hyperspeech hypothesis – IDS might be easier to segment. However, both, the literature review as well as the corpus analysis revealed that the three cues may vary in strength and reliability within and between registers. To investigate whether the cues are nevertheless useful for boundary detection, in the third part of their study, the data of the corpus analysis was subjected to a machine learning experiment. IDS boundaries were always better learned than ADS boundaries with a special role of the pause cue since all cases that showed better performance in IDS than in ADS involved the pause cue. The authors concluded that pause is the primary cue in IDS. Still, the learning algorithm in IDS benefitted from adding one extra cue to any other cue or cue combination (e.g., adding F0 to pause or adding lengthening to the combination of pause and F0). Hence, even though cues and their constellation may vary in the input, the learning mechanism takes them all into account. The authors state that the three cues "are mutually complementary sources of information and are all useful for learning" (Ludusan et al., 2016: 1248).

This section illustrated the presence of acoustic cues at syntactic boundaries in ADS and IDS, supporting the first assumption of the prosodic bootstrapping hypothesis. In the following, an overview of results concerning their influence on infant speech perception will be presented.

2.3 Infants' sensitivity to prosodic boundary information

The previous section showed that the relevant prosodic information to bootstrap syntactic segmentation is indeed available in the speech signal. The next question that arises within the framework of the prosodic bootstrapping hypothesis is: Are the learners sensitive to this information? In the following, studies will be summarized that support the assumption that infants are sensitive to acoustic correlates that may aid syntactic segmentation.

Right after birth, perceiving at least some acoustic correlates of prosodic phrase boundaries seems to be possible. Christophe, Dupoux, Bertoncini, and Mehler (1994) tested French newborns with the high amplitude sucking procedure. The stimuli were bisyllabic items extracted either from the middle of one long word, hence containing no boundary (e.g., *mati* from "… mathé*mati*cien …"), or from two words – the final syllable of one word and the first syllable of the following word with a PP boundary in between (e.g., *mati* from "panora*ma ty*pique"). This PP boundary was acoustically marked by initial consonant lengthening of /*t*/ and preboundary lengthening of the syllable /ma/. Three-day-old newborns discriminated between with- and no-boundary items. This suggests that newborns are sensitive to acoustic correlates of prosodic boundaries such as initial as well as preboundary lengthening. However, since French is a language with phrasal stress, there was a confounding factor regarding the duration of the preboundary syllable. The with-boundary items consisted of an unstressed and a stressed final syllable, whereas no-boundary items had two unstressed syllables. The increased duration of the preboundary syllable observed in the with-boundary items might have been the result of the word-final stress typical for French. Hence, it might be possible that discrimination was at least partially triggered by the differences in stress overlapping with boundary information. To disentangle the influence of stress and boundary markings, Christophe & Mehler (2001) conducted a similar study with French newborns with stimuli in Spanish, a language with flexible word stress. Differently to the French stimuli, the Spanish CVCV items were always unstressed on the first syllable, and stressed on the second syllable. As in the first study, the two syllables (e.g., *lati*) had been extracted from one word (e.g., *gelatína*) or from two words spanning a PP boundary (e.g., Manué*la tí*mida). This boundary was characterized by higher pitch on the preboundary vowel, increased energy of the preboundary vowel and the initial consonant, but no significant preboundary vowel lengthening as in the French stimuli. Still, the distance between the vowel onsets of the two syllables was larger in the boundary condition. Also with these Spanish stimuli, French newborns successfully performed the discrimination task. These two studies suggest that infants have access to acoustic correlates of PP boundaries such as changes in pitch, intensity, and duration from very early on.

Sensitivity to the acoustic correlates of larger prosodic boundaries such as IP boundaries becomes evident only later, in the second half of the first year of life. In a landmark study, Hirsh-Pasek et al. (1987) established the insertion of an artificial pause to stimulus materials as a technique that henceforth would be used in a number of studies tracking infants' sensitivity to the acoustic correlates of syntactic boundaries. Hirsh-Pasek at al.'s materials were excerpts from an interactive conversation and oral storytelling by a mother to her toddler. From these sentences, they created two versions: in the natural (coincident) version a pause of 1 s was inserted at all clause boundaries. In the unnatural (non-coincident) version the authors inserted a 1 s pause between words in the middle of the clause, that is, at a non-boundary location. Moreover, the unnatural version started and ended in the middle of a clause. Examples of the speech passages are cited below:

natural version	"Cinderella lived in a great big house # but it was sort of dark # because she had this mean, mean, mean stepmother. # And, oh, she had two stepsisters # that were so ugly. # they were mean, too."
unnatural version	"… in a great big house, but it was # sort of dark because she had # this mean, mean, mean stepmother. And, oh, she # had two stepsisters that were so # ugly. They were mean, # too. They were …"

Hirsh-Pasek et al. (1987: 273)

Presented with these materials in the *headturn preference procedure*, infants aged six to ten months listened significantly longer to the natural samples.[2] Hirsh-Pasek and colleagues considered different interpretations: on the one hand, the different listening times might reflect a preference for the cohesive structure of clausal units. On the other, they could also mirror infant' disliking of the unnatural version because the artificial pauses within the clause disrupt the prosodic unit and/or because the pauses are not accompanied by other boundary cues typically correlating with a pause. Certainly, the results revealed that infants are sensitive to prosodic cues of clausal boundaries which might in turn enable them to segment the speech stream roughly into clauses (as evidenced in later studies, see Chapter 2.4.1).

The pause-insertion technique was expanded in several studies with respect to age of the tested infants, their language, the relevant syntactic unit (clauses vs. syntactic phrases), the speech register (IDS vs. ADS), and the effect of using low-pass filtered stimuli (for English: Gerken et al., 1994; Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989; for Japanese: Hayashi & Mazuka, 2002; for German: Schmitz, 2008). In all of these studies coincident versions with pauses at boundary locations were contrasted to non-coincident samples where the pauses were inserted at non-boundary locations. Kemler Nelson et al. (1989) replicated Hirsh-Pasek et al.'s results with a group of infants aged seven to nine months who preferred the coincident version of the same materials in an IDS version. However, a group that listened to similar materials in ADS mode did not prefer any of the conditions highlighting the crucial role of motherese in clause segmentation. Jusczyk and colleagues (1992) targeted phonological phrase boundaries and provided evidence that nine- but not six- month-olds prefer to listen to the coincident samples of noun and

---

[2] This listening preference for natural materials showed up in both experiments of the study. In the first experiment all naturally occurring pauses were maintained. This caused a higher frequency of pauses in the unnatural samples because there were artificial pauses within the clause plus natural pauses at the boundary location. Another discrepancy across the natural and unnatural versions concerned the pause durations: pauses occurring at the boundaries were longer in the natural version. To avoid these confounds, in a second experiment all naturally occurring pauses longer than 450 ms were deleted. Even with this modification, infants in the second experiment performed similar to those tested in the first one.

predicate phrases. This preference even held when materials were presented in adult-directed speech, but also with a low-pass filtered version of the stimuli with eliminated phonetic details, but intact prosodic information. This revealed that prosody is sufficient for discrimination, even in the absence of lexical, semantic or syntactic sources.

Hayashi and Mazuka (2002) developed Japanese materials and found that Japanese-learning infants distinguish between clauses with artificially inserted pauses at boundary versus non-boundary locations only from the age of eight months on, but not yet between four and seven months.

For German, Schmitz (2008) conducted several experiments using the pause-insertion technique with six- and eight-month-old infants. Schmitz tested whether infants can discriminate between pauses at boundary and non-boundary positions and if so, whether discrimination is affected by different pause durations.

Schmitz' stimuli were sentences consisting of a matrix clause coordinated with an infinitival clause, together making up two IPs (e.g., "Die Tante verspricht Katrin zu schwimmen und die Insel zu erkunden." 'The aunt promises Katrin to swim and to explore the island'). For the coincident version, the artificial pause was inserted between the two IPs, that is, at the clause boundary position before the conjunction "und" (e.g., "Die Tante verspricht Katrin zu schwimmen # und die Insel zu erkunden."). Creating the non-coincident version, the artificial pause was inserted within the first clause after the infinitive particle "zu" (e.g., "Die Tante verspricht Katrin zu # schwimmen und die Insel zu erkunden"). For each condition, six different sentences were concatenated to one block. The results revealed that six-month-old German-learning infants did not discriminate between the two conditions when the artificially inserted pause was of 1 s duration with an *interstimulus interval (ISI)* between sentences of 1.5 s, which is outside the range of naturally occurring pauses in German speech corpora (cf. Butcher, 1981). However, when the artificial pause was only 750 ms long and the ISI 1 s, infants discriminated the coincident from the non-coincident version. Schmitz concluded that infants recognize the prosodic correlates of clause boundaries, but only when the pause durations reflect conditions of natural speech. She took the results as a hint towards an early sensitivity to the natural timing of pauses. Extending these findings, she addressed the *natural pause hierarchy* in further experiments. The natural pause hierarchy relates to the observation that the duration of a pause correlates with the type of boundary (Butcher, 1981; Goldman-Eisler, 1972), that is, the higher the syntactic level the longer the pause (from no or only minimal pauses at word boundaries to longer pauses at phrase boundaries, again longer pauses at clause boundaries to the longest pauses at sentence boundaries).

Aiming at this phenomenon, Schmitz (2008) created stimuli that were coincident in both conditions, but the durations of the pause and the ISI either corresponded to the natural pause hierarchy or not. For the natural pause condition, a pause of 750 ms was inserted at the natural clausal boundary position and the ISI between sentences was 1 s. For the reversed pause condition, the pause at the clausal boundary was 1 s and the ISI between sentences only 750 ms, hence violating the natural pause hierarchy. Eight-month-old infants preferred to listen to the natural pause condition over the reversed pause condition, whereas six-month-olds did not discriminate between both conditions. Schmitz argued that infants develop their ability to process subtle time differences from six to eight months and hence, are able to demonstrate their sensitivity to the natural pause hierarchy. By eight-months, infants have not only gained knowledge about the natural location of pauses, but also about their timing.

This section presented studies that provide robust evidence that newborns as well as infants are highly sensitive to acoustic cues to prosodic boundary information. By the second half of their first year of life, infants notice the very local violation that arises when a boundary marker such as the pause cue occurs at a position in which it is not accompanied by typical boundary changes in pitch and/ or preboundary lengthening. The sensitivity to the correlation of pitch, lengthening, and pause provides infants with a crucial tool to identify the edges of sentences and clauses. Moreover, infants have been shown to be sensitive to the natural pause hierarchy which may be path-breaking when acquiring the hierarchical structure of syntax.

2.4 Infants' exploitation of prosodic boundary information

So far, we have seen that there is evidence in favor of the first assumption of the prosodic bootstrapping hypothesis: there are acoustic correlates of prosodic boundaries aligned with syntactic boundaries in the speech signal. Also the second assumption has found empirical support: infants are sensitive to prosodic boundary information. The studies described in the following provide evidence supporting the third assumption: infants make use of prosodic boundary characteristics for the segmentation of syntactic units.

2.4.1 Segmentation and recognition of clauses

The question whether infants' sensitivity to prosodic boundaries has an impact on how the learner structures her mental representation of the incoming speech was addressed by studies by Mandel, Jusczyk, and Kemler Nelson (1994), Mandel, Kemler Nelson, and Jusczyk (1996), Nazzi, Kemler Nelson, Jusczyk, and Jusczyk (2000) and Soderstrom, Nelson, and Jusczyk (2005).

Mandel, Jusczyk, and Kemler Nelson (1994) provided the first evidence that prosody indeed affects the learner's perceptual organization of the speech stream. Prosodic structuring was shown to help infants in encoding and remembering speech. In particular, infants at the age of two months detected phonetic changes in word sequences only when the words had been familiarized in a sentence with coherent prosody, but not when they had been presented as lists of isolated words. Mandel et al. concluded that prosodic coherence enhances memory capacities. Such an advantage was also observed in infants' detection of word order changes (Mandel et al., 1996; for a more detailed description of both studies see Chapter 7).

Moving closer towards natural speech conditions, Nazzi et al. (2000) investigated whether six-month-old infants are able to recognize a whole clause within a longer speech passage. They hypothesized that recognition would be better after familiarization to a prosodically *well-formed* clause than after familiarization to a prosodically *ill-formed* sequence of the same words. The following two passages were used as stimuli:

(1) "John doesn't know what *rabbits eat.* ***Leafy vegetables*** **taste so good.** They don't cost much either."

(2) "Many animals prefer some things. **Rabbits eat *leafy vegetables.*** *Taste so good* is rarely encountered."

Nazzi et al. (2000: 127)

From these passages, the sequences "rabbits eat leafy vegetables" ("rabbit" sequence) and "leafy vegetables taste so good" ("leafy" sequence) were extracted, once in a prosodically well-formed version (bold) and once in a prosodically ill-formed one (italics). The latter spanned a clause boundary that was characterized by a pitch rise, increased duration of the final stressed syllable and increased duration of the pause between the two words compared to the same position in the well-formed version (note that the pause durations in the ill-formed versions differed quite largely between the two word sequences, 456 vs. 146 ms). Infants were assigned either to a group that was familiarized with the "rabbits" sequence or to a group that was familiarized with the "leafy" sequence. Both groups heard its sequence in both prosodic conditions – the prosodically well-formed and the ill-formed version. During test, both familiarization groups listened to the same two passages cited above. Across both groups, infants preferred that passage that matched the well-formed familiarization sequence, that is, the group familiarized with the "rabbits" sequences preferred passage 2, whereas the group familiarized with the "leafy" sequences preferred to listen to passage 1. Unfortunately, Nazzi et al. did not report information on potential interactions of the listening times with the kind of familiarization ("leafy" or "rabbit" sequence). It would have been interesting to see whether the different pause durations at the clausal boundaries of the two word sequences affected infants' recognition. However, Nazzi and colleagues interpreted the overall results such that infants recognized the re-occurrence of a prosodically well-formed word sequence within a longer passage of speech.

The preference for the well-formed sequences even held when the familiarization sequences were extracted from another text passage, thus, no longer being an exact acoustic match of the embedded test sequences. Nazzi et al. concluded that the data does not merely reflect enhanced memory for prosodic units as Mandel et al. (1994, 1996) had evidenced. Rather, Nazzi et al.'s study revealed that prosodic packaging affects infants' mental online processing of continuous speech as "the well-formed unit is more easily extractable or unitized within its passage than the ill-formed unit is within its passage" (Nazzi et al., 2000: 140). The authors claimed their study to be the first one that shows infants' use of prosodic cues in the online segmentation of continuous speech into clauses.

Soderstrom, Kemler Nelson, & Jusczyk (2005) built up on Nazzi et al.'s study with a very similar experimental design. Differently, infants were not familiarized with pre-extracted word sequences, but with a whole text passage containing the word sequence (the so-called target sequence). The passages presented during the test phase were again new text passages that contained this word sequence. Hereby, the recognition task was more difficult, but also more similar to the task that infants encounter in natural speech situations.

Familiarization passages (two groups):

(1) "John doesn't know what *rabbits eat.* ***Leafy vegetables*** **taste so good.** They don't cost much either."

(2) "Many animals prefer some things. **Rabbits eat** ***leafy vegetables.*** *Taste so good* is rarely encountered."

Test passages (all infants):

A Children need to have healthy foods. **Leafy vegetables taste so good.** Salad is best with dressing.

B Students like to watch how *rabbits eat. Leafy vegetables* make them chew. They finish the whole bunch.

C Squirrels often feed on acorns. ***Rabbits eat leafy vegetables***. Their favorites are not the same.

D Mothers sometimes do the shopping. They must buy *leafy vegetables. Taste so good* helps their families.

Soderstrom et al. (2005: 92)

Two groups of six-month-old infants were familiarized with one of the familiarizations test passages (1 or 2). Within the familiarization passage, two different target sequences occurred, "rabbits eat leafy vegetables" and "leafy vegetables taste so good". One target sequence occurred in a prosodically well-formed version (bold, e.g., "… **Leafy vegetables taste so good.** …" in Passage 1), the other in a prosodically ill-formed version spanning a clausal boundary (italics, e.g., "… *rabbits eat. Leafy vegetables* …" in Passage 1). During test, all infants heard four different test passages (A-D) each containing one of the target sequences in one prosodic conditions. Results revealed that infants preferred to listen to those two passages that contained the target sequence that was prosodically well-formed in the familiarization passage (e.g., Group 1 prefers Passage A over D). Moreover, infants preferred the test passage in which the prosody of the target sequence at test matched the prosodic structure during familiarization (e.g., Group 1 preferred Passage B over C) suggesting that infants are better able to recognize a word sequence that is reproduced with similar prosodic characteristics (even when it is spanning a boundary). The optimal condition for recognition was when the sequence constituted one cohesive prosodic unit at familiarization and at test. Soderstrom et al. (2005)'s study extended previous studies by demonstrating six-month-olds' use of clause-level prosody across different contexts of fluent speech and even after very few presentations.

Taken together, the studies by Nazzi et al. (2000) and Soderstrom et al. (2005) and Mandel et al. (1994, 1996) provide support for the third assumption of the prosodic bootstrapping hypothesis: infants do make use of prosody in online processing. Specifically, already in the first half year of life infants rely on prosodic boundary information such as pitch changes, preboundary lengthening, and pause to encode, memorize and recognize clausal units within the continuous speech stream.

Going beyond segmentation, Hawthorne and Gerken (2014) addressed the role of prosody in learning syntactic constituents with a very different experimental design. Employing an artificial grammar task, they examined whether toddlers use prosody to chunk groups of words into cohesive units and whether they recognize these units when they move – like syntactic constituents – to a new position in the utterance.

To this end, 19-month-old English-learning toddlers were familiarized with strings of an artificial grammar consisting of six CVC non-words (ABCDEF, e.g., "bup div kag feb has jik"). Each letter referred to a class of two non-words (e.g., A = "bup/ nim", B= "div/ pel", …). Half of the toddlers, the control group, was familiarized with these strings in 1-clause (1-IP) prosody without a boundary that was derived from the sentence "I ate at home with John." and transferred to the non-word string (ABCDEF). The other half of toddlers, the experimental group, was familiarized with the same string in 2-clause prosody that was derived from the sentence "While I ate, John came home", that is, including an IPB, hence expressing 2-IP prosody (ABC, DEF).

Familiarization consisted of ten minutes in a small room, where the child, the experimenter and the parent played quietly while listening to the familiarization strings. Immediately, the familiarization continued for additional two minutes in the HPP testing booth. Subsequently, all participants were tested on strings with 2-clause prosody that were grammatical (DEF, ABC) or ungrammatical (EFA, BCD) re-orderings of the words from the 2-clause familiarization. Grammatical items preserved the groupings that were prosodically cohesive in the 2-clause familiarization, whereas ungrammatical items split the groupings that the 2-clause familiarization had indicated.

The experimental group, but not the control group, successfully discriminated grammatical and ungrammatical re-orderings as expressed by longer listening times to the grammatical test items. The authors concluded that the experimental group had segmented constituent-like chunks from the two-clause familiarization string such as ABC and DEF and were able to recognize these chunks that had moved in the test string to a new position (DEF, ABC). By 19 months, toddlers "treat prosodically-grouped words as cohesive, reorder-able unit that are independent of a particular acoustic manifestation" (Hawthorne & Gerken, 2014: 424).

As Nazzi et al. (2000) and Soderstrom et al. (2005) with infants, Hawthorne and Gerken (2014) verified the use of clausal prosody for segmentation with toddlers. Going beyond, they uncovered that the young learner treats the segmented categories as cohesive units that act together (here, movement), a key feature of syntactic constituents.

Hence, prosody does not only provide the young learner with a tool for segmentation, but also for learning syntactic constituency. Hawthorne and Gerken (2014) point out that this is especially useful during the earliest stages of language acquisition when other sources for deriving constituency like knowledge about semantics, frequently co-occurring morphemes, transitional probabilities, and comparisons across sentences have not yet been acquired.

2.4.2 Disentangling syntactic ambiguities

The studies described above underline the importance of phrasal prosody for segmenting and structuring the speech stream. Furthermore, the learner makes also use of phrasal prosody to constrain the syntactic analysis when encountering ambiguities. Recent studies tested French toddlers (20 and 28 months; de Carvalho, Dautriche, Lin, & Christophe, 2017) and adults and preschoolers (3.5 and 4.5 years; de Carvalho, Dautriche, & Christophe, 2016). De Carvalho and colleagues created ambiguous sentences with noun/verb homophones to be presented in an intermodal preferential looking task with an eye-tracker. In the study with adults and pre-schoolers this was combined with a pointing task. Stimuli in the noun-prosody condition (e.g., "[le bébé souris$_{\text{NOUN}}$] [a bien mangé]" – '[the baby mouse] [ate well]') had a prosodic boundary after the ambiguous target word ("souris/sourit"), whereas sentences in the verb-prosody condition (e.g., ["le bébé] [sourit$_{\text{VERB}}$ à sa maman]" – '[the baby] [smiles to his mother]') had the prosodic boundary before the ambiguous target word. These boundaries were characterized by rising pitch, preboundary lengthening, and phrase-initial consonant strengthening. Note that in contrast to the previously described studies no pauses were present and the boundaries under investigation here relate to PP boundaries.

Participants listened to the beginnings of these sentences, that is, up to and including the ambiguous word and were presented with two pictures, one associated with the noun interpretation of the ambiguous word, the other with the verb interpretation. Adults and preschoolers (de Carvalho et al., 2016) as well as 20- and 28-month-olds (de Carvalho et al., 2017) correctly identified the picture that corresponded to the prosodic structure of the sentence (with the younger children being much slower). The results suggest that from 20 months on children are able to use prosodic information, here the acoustic correlates of the PP boundary, to derive the syntactic structure of an utterance and to determine the syntactic category of an ambiguous word. Hence, during their second year of life infants make use of prosodic boundary information when encountering syntactically ambiguous structures. The authors admit that syntactically ambiguous structures occur rather rarely in IDS, however, they assume that the observed performance regarding these kind of structures reveals that the learner is able to identify syntactic units and moreover to assign syntactic categories such as nouns and verbs. Once syntactic categories are assigned, the learner might guess the meaning of unknown words that they encounter in their natural language environment (as proposed in the syntactic bootstrapping hypothesis by Gleitman, 1990).

3 Research questions

Starting from findings that infants use the acoustic correlates of prosodic boundaries for the segmentation of the continuous speech stream, in three studies Seidl and colleagues investigated the role of the individual boundary cues, pitch change, final lengthening, and pause. They found that American-learning develop from requiring all three cues in combination at the age of four months (Seidl & Cristià, 2008) to requiring only a subset of prosodic cues at six months (Seidl, 2007). Importantly, this subset of cues needs to contain the pitch cue, combined with either final lengthening or pause. A boundary marked by the subset of lengthening and pause did not lead to successful clause recognition (Seidl, 2007). Hence, pitch proved to be a necessary cue for the early segmentation of American English. With the same experimental design, but Dutch stimuli, the pause cue was shown to be necessary for Dutch-learning infants (Johnson & Seidl, 2008; for a detailed description of the studies see Chapter 4.1).

There are two possible explanations for the different outcomes with English and Dutch infants. First, infants may apply different clause segmentation strategies related to a language-specific weighting of prosodic cues. During their first half year of life, American English infants may have learned that pitch is a crucial and reliable boundary cue. In contrast, Dutch infants may have learned to weight pause especially high, reflecting a behavior useful regarding the typical boundary characteristics of the Dutch language. The second explanation is that infants universally apply the same segmentation strategy by attending to acoustically salient features. Thus, the different outcomes might stem from differences in the salience of the pitch and pause cue in the English versus Dutch stimuli. This highlights the importance of a thorough control of the acoustic realizations of boundary cues in future studies. Even though the two explanations could not be disentangled, the work by Seidl and colleagues proposes that language-specificity plays a role in the domain of prosodic boundary cue perception. The impact of individual prosodic boundary cues on infant clause segmentation may differ across languages and infants' sensitivity to these cues may develop according to their native language.

Tracking this assumption, the following three studies provide insights from German prosodic boundary perception. They investigated the role of individual prosodic boundary cues and cue combinations in discriminating between different prosodic groupings and segmenting prosodic units. These are the first studies that used the exact same materials that have also been tested with German adults (Holzgrefe-Lang et al., 2016), allowing for a proper comparison across age. All three studies employed behavioral experiments using the HPP with six- and eight-month-old monolingual German-learning infants tested in the Potsdam BabyLab.

Study I addressed the impact of pitch change, final lengthening, and pause in the perception of a prosodic boundary in German-learning eight-month-old infants when discriminating different prosodic groupings. Five experiments sought to answer the following questions:

- Do eight-month-old infants perceive an IPB within a list of coordinated names when it is signaled by the three main naturally occurring prosodic boundary cues (pitch change, final lengthening, pause)?
- Does infant boundary perception require the occurrence of all three cues together or is it sufficient that only a subset of IPB cues or even a single cue is present? Particularly, are eight-month-old infants able to perceive a prosodic boundary cued by the combination of pitch change and final lengthening, or is pause a necessary boundary cue?
- Are pitch change or final lengthening alone sufficient to trigger boundary perception?
- Does the quality of prosodic boundary cues, either occurring naturally or being inserted by acoustic manipulation, have an impact on infants' listening preferences?

To answer these questions a discrimination task was run that used coordinated lists of names that either contained an internal prosodic boundary or not (e.g., [Moni und Lilli]$_{\text{IP}}$ [und Manu]$_{\text{IP}}$ vs. [Moni und Lilli und Manu]$_{\text{IP}}$). The boundary marking was systematically varied. First, the full set of naturally occurring cues indicated the boundary (Exp. I/1). If indeed, German-learning eight-month-olds proved to perceive this fully marked IPB, this would state the baseline to investigate the relevance of single IPB cues or specific cue combinations in subsequent experiments, in which the boundary was created by inserting specific boundary cues with means of acoustic manipulation. The implementation was first carried out for the combination of pitch change and final lengthening (Exp. I/2), in a further step, for pitch and final lengthening individually (Exp. I/3 and I/4). To control for a potential influence of the factor acoustic manipulation, in a final experiment all three cues were inserted to indicate the boundary (Exp. I/5). Study I provides insight how pitch change, final lengthening, and pause contribute to the perception of German IPBs. The results are compared to findings from German adult listeners (Holzgrefe-Lang et al., 2016) and discussed in the light of previous crosslinguistic research.

Study II examined the developmental course of prosodic boundary cue perception in the acquisition of German. Following Study I, here, infants that are two months younger were tested with similar experiments varying the constellation of prosodic boundary cues. The following questions were addressed in four experiments:

- Do six-month-old infants perceive an IPB occurring within a list of coordinated names when it is signaled by the three main prosodic boundary cues (pitch change, final lengthening, pause)?
- Do six-month-old infants need pitch change, final lengthening, and pause to occur together to perceive a prosodic boundary or is a subset of two cues sufficient? Specifically, is the

combination of pitch and final lengthening sufficient to trigger boundary perception or does the perception of a boundary at six months depend on the pause cue?

- Is the combination of pause and lengthening sufficient to trigger boundary perception?

First, the same baseline experiment as in Study I presenting a fully cued IPB was run (Exp. II/1). Since previous research revealed a high sensitivity to the pause cue and the correlation of boundary cues in German-learning infants (Schmitz, 2008), I hypothesized that this fully cued boundary would be detected by six-month-olds. Next, the boundary was indicated by pitch change and final lengthening. Discrimination with these materials was tested in two experimental versions, after short (Exp. II/2) and after long familiarization (Exp. II/2a). Finally, to evaluate the role of the pause cue in comparison to the coalition of all three cues, the boundary was cued by pause and final lengthening (Exp. II/3). In the light of the findings from Study I, the results are discussed regarding perceptual attunement in prosodic cue perception. Moreover, similarities and differences across the languages studied so far are pointed out. Finally, I reflect on methodological issues in boundary perception in a comparison to a previous electrophysiological study with the same materials.

After Studies I and II have identified the prosodic cues and cue combinations that are essential to perform a discrimination task, **Study III** aimed at examining whether infants make use of the same cue or cue constellations in a task that demands segmentation and recognition of prosodic units. As a pilot study, Study III sought to find an experimental design for a segmentation task that would answer this question:

- Do eight-month-old infants recognize a sequence of two names out of a list of three names when it is prosodically grouped as an IP with its edges cued by pitch change, final lengthening, and pause?

In three slightly varying experimental designs (Exp. III/1, III/2, III/ 3), eight-month-old German-learning infants were familiarized with two different combinations of three coordinated names with different prosodic groupings. Subsequently, infants were tested on the recognition of two out of the three names. Previous studies suggest that IPs are processing units that infants use to memorize and recognize units from a fluent passage of speech (Mandel et al., 1994, 1996; Nazzi et al., 2000). Therefore, I assume that the full set of prosodic boundary cues provided here would help infants in this study to identify the edges of prosodic units and thus to segment the two names out of the list of three names. If one of the experimental designs piloted here proved to be suitable, in a next step – that is not part of Study III – those prosodic boundary cues and cue constellations that Studies I and II have yielded to be relevant in German boundary perception could be systematically examined with this segmentation task.

Study I

4 How each prosodic boundary cue matters: Evidence from German infants[3]

Abstract

Previous studies have revealed that infants aged six to ten months are able to use the acoustic correlates of major prosodic boundaries, that is, pitch change, preboundary lengthening, and pause, for the segmentation of the continuous speech signal. Moreover, investigations with American-English- and Dutch-learning infants suggest that processing prosodic boundary markings involves a weighting of these cues. This weighting seems to develop with increasing exposure to the native language and to underlie crosslinguistic variation. In the following, we report the results of four experiments using the *headturn preference procedure* to explore the perception of prosodic boundary cues in German infants. We presented eight-month-old infants with a sequence of names in two different prosodic groupings, with or without boundary markers. Infants discriminated both sequences when the boundary was marked by all three cues (Exp. I/1) and when it was marked by pitch change and preboundary lengthening in combination (Exp. I/2). The presence of a pitch change (Exp. I/3) or preboundary lengthening (Exp. I/4) as single cues did not lead to a successful discrimination. Our results indicate that pause is not a necessary cue for German infants. Pitch and preboundary lengthening in combination, but not as single cues, are sufficient. Hence, by eight months infants only rely on a convergence of boundary markers. Comparisons with adults' performance on the same stimulus materials suggest that the pattern observed with the eight-month-olds is already consistent with that of adults. We discuss our findings with respect to crosslinguistic variation and the development of a language-specific prosodic cue weighting.

---

[3] This chapter has been published as: Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012) How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology, 3*(580).
The present version was slightly modified with respect to the status of some references and the numbering of the experiments.

4.1 Introduction

The system underlying the prosodic organization of language constitutes a complex linguistic subsystem with strong interfaces to other linguistic domains like the lexicon or the syntax. This paper deals with the correlation between prosodic phrasing and the syntactic structure of utterances which has already been the subject of numerous studies in the area of adult sentence processing as well as of infant language acquisition (e.g., Hirsh-Pasek et al., 1987; Nazzi et al., 2000; Peters et al., 2005; Sanderman & Collier, 1997; Scott, 1982; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003; Streeter, 1978). The question unifying these diverse areas of research is whether prosody provides information that can enter into the processing of the syntactic structure of utterances. In language acquisition research this approach is known as the prosodic bootstrapping account (Gleitman & Wanner, 1982), which assumes that infants can exploit acoustic information from their speech input to find solutions for several tasks they are faced with when accessing the grammatical system of their language. In this paper, we will have a closer look at German infants' sensitivity to the acoustic cues that mark a major prosodic boundary, that is, the intonation phrase boundary (IPB).

There are two properties that render IPBs especially useful within the prosody-syntax mapping. First, a rather clear-cut set of acoustic cues, namely pitch changes, lengthening of preboundary segments, and pauses, is associated with IPBs across different languages (e.g., Féry et al., 2011; Hirst & Di Cristo, 1998; Nespor & Vogel, 1986; Peters et al., 2005; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Vaissière, 1983; Venditti, Jun, & Beckman, 1996; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Secondly, again crosslinguistically, there exists a high coincidence of IPBs with major syntactic boundaries like sentence and clause boundaries (e.g., Cooper & Paccia-Cooper, 1980; Vaissière & Michaud, 2006; Venditti et al., 1996). Hence, sensitivity to the relevant acoustic cues would provide infants with a strong mechanism for chunking incoming speech into syntactically relevant units without requiring lexical or syntactic knowledge.

Indeed, numerous studies within the prosodic bootstrapping account have demonstrated that infants are equipped with a high sensitivity to prosodic information such as stress, rhythm, and intonation (for an overview, see Jusczyk, 1997). This also holds for the perception of acoustic information that is related to the marking of prosodic boundaries. Research in this area started with some landmark studies that tested infants' reactions to the presentation of natural speech in contrast to manipulated speech material in which pauses had been inserted at non-boundary positions (Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989). These studies – using the *headturn preference procedure (HPP)* – showed that American infants as young as seven to ten months prefer to listen to speech material showing a coincidence of the typical acoustic cues occurring at clausal boundaries compared to materials in which the coincidence of pauses with other prosodic cues had been disrupted. The fact that the same preference occurred with low-pass-filtered material strongly suggests that it is the disturbance of the prosodic organization of the utterances that causes the successful discrimination of both kinds of material. Studies with other languages using the same technique of pause insertion have provided evidence that this discrimination ability is not unique to English-learning infants: German as well as Japanese infants have been found to discriminate speech with pauses at clausal boundaries from speech with pauses inserted at non-boundary positions in their language (Hayashi and Mazuka, 2002; Schmitz, 2008).

Also using pause insertion, Jusczyk et al. (1992) investigated infants' sensitivity to boundaries of smaller units, namely clause-internal phrase boundaries. In their material, pauses were inserted either before the main verb, that is, at the boundary between the subject and the verb phrase, or after the main verb, that is, within the verb phrase. English-learning nine-month-olds preferred to listen to the materials in which the pause occurred at the phrasal boundary.

Gerken et al. (1994) compared sentences with lexical subjects (e.g, "The caterpillar ate …") and sentences with pronominal subjects (e.g., "He ate …") in which pauses had been inserted after either the subject or the verb. As lexical subjects form their own phonological phrase, there is a prosodic boundary between the subject and verb in the corresponding sentences while there is typically no prosodic boundary after a pronominal subject. Only in the lexical subject condition did nine-month-old infants prefer to listen to sentences with pauses after the subjects (e.g., "The caterpillar # ate …", "He # ate …") over those with pauses after verbs (e.g, "The caterpillar ate # …", "He ate # …"). These results again suggest that the prosodic organization – and not the syntactic one – is relevant for infants' preference for natural material. Taken together, these studies provide evidence that by nine months infants are sensitive to the acoustic markers at clausal as well as at phrasal boundaries.

More recent work has gone beyond the question of the perception of the acoustic correlates of major boundaries to the question as to whether the occurrence of prosodic boundaries affects the segmentation of continuous speech. Nazzi et al. (2000) were the first to test English-learning six-month-olds' use of prosodic boundary cues to segment continuous speech. At the beginning of the experiment infants were familiarized with a sequence of words, once as a prosodically *well-formed* clause (e.g., "Leafy vegetables taste so good.") and once as a prosodically *ill-formed*, that is, non-clausal, sequence that contained an internal clause boundary (e.g., "… leafy vegetables. Taste so good…"). These word sequences had been extracted from two different continuous passages. After familiarization infants were presented with two passages. One of them contained the familiarized prosodically well-formed sequence, the other the prosodically ill-formed sequence, which was now the end and the beginning of two adjacent sentences. This non-clausal unit contained a prosodic boundary that was marked by a pitch change, preboundary lengthening, and pause. Infants listened significantly longer to the passage containing the clausal sequence than to the passage with the non-clausal sequence. These results suggest that word sequences that constitute a prosodic unit are better recognized than word sequences that span a prosodic boundary. Hence, prosodic boundary cues support the segmentation of clauses within a passage of sentences. These findings were replicated by Soderstrom et al. (2005) with a similar design, but more complex experimental materials. Specifically, it was demonstrated that prosodic boundary cues support English-learning infants' detection of familiar word sequences even across different passages of fluent speech.

Moreover, with a similar experimental design, Soderstrom et al. (2003) provided evidence that six-month-old English-learning infants also use prosodic markers to detect syntactic units that are smaller than the clause, namely phrasal units such as noun and verb phrases. Interestingly, phrase boundaries were characterized by preboundary lengthening and pitch cues while there was no perceivable pause at the crucial position. This suggests that for the detection of phrase boundaries pause is not a necessary cue for six-month-old English-learning infants.

The studies presented so far point to a crucial role of prosodic boundary information in infants' speech segmentation, especially during the first year of life. However, in a critical analysis of the prosodic bootstrapping account, Fernald and McRoberts (1996) doubt the reliability of acoustic correlates of prosodic boundaries as cues to syntactic units. The authors claim that none of the three markers is a reliable cue to syntactic boundaries as each cue also has non-linguistic functions (e.g., pitch changes for the regulation of affect) or linguistic functions other than syntax (e.g., vowel length as phonemic contrast). This would cause ambiguity of the acoustic correlates of boundaries whenever they occur at non-boundary positions. Fernald and McRoberts' argument may be weakened if a comprehensive analysis of a corpus of German adult-directed speech conducted by Peters et al. (2005) is taken into account. They found that IPBs were most frequently marked by pitch changes, followed by preboundary lengthening, while the occurrence of pause is rather rare. In addition, the analysis showed that each cue may occur individually, but that in a great majority of the cases boundaries are marked by a coalition of all three or two of the relevant cues. This convergence may decrease the ambiguity of prosodic boundary cues provided that the infant only considers a combination of cues to be a boundary marker.

In fact a detailed study by Seidl (2007) that tested the perceptual impact of each of the prosodic cues provided evidence that English-learning six-month-old infants rely on a combination of cues in their boundary processing. The investigations were based on the materials and the experimental design used by Nazzi et al. (2000). Seidl successively neutralized each acoustic correlate of the prosodic boundaries in the familiarization sequences. Thereby, the acoustic realization of the cue under investigation no longer differed between the two sequences. The question was whether infants, on the basis of the remaining prosodic cues, would still differentiate the clausal and the non-clausal familiarization sequences and recognize the clausal sequence in the passage during testing.

Infants' detection of the clausal sequence was not disturbed by the neutralization of the pause cue. This indicates that pitch change and preboundary lengthening were sufficient cues for the six-month-old English-learning infants, whereas the pause was not necessary. Furthermore, preboundary lengthening also proved not to be a necessary cue, because infants still recognized the clausal sequence when preboundary lengthening was neutralized. However, when the pitch cue was neutralized the infants no longer detected the clausal sequence in the passage. Hence, pitch change proved to be a necessary boundary cue for American infants' clause segmentation. A further experiment investigated whether pitch change as a single cue would suffice. This kind of acoustic manipulation disturbed infants' detection of the clausal sequence, indicating that a pitch change alone is not sufficient. In conclusion, a combination of pitch change and preboundary lengthening or pitch change and pause was necessary to trigger clause segmentation in six-month-old English-learning infants. Seidl (2007) argued that by six months English-learning infants do not treat prosodic cues equally, but have, at least partially, become attuned to adults' weighting of prosodic cues in their native language (Aasland & Baum, 2003; Scott, 1982; Streeter, 1978).

Seidl and Cristià (2008) expanded these investigations by testing four-month-old English-learning infants with the same materials. In contrast to the six-month-olds, this younger group was successful in clause segmentation only when pitch change, lengthening, and pause in combination signaled the boundary. Neutralization of one of the prosodic cues led to failure in segmentation. Seidl and Cristià (2008) concluded that four-month-old English-learning infants segment clauses by considering all prosodic boundary cues.

In a following study, Johnson and Seidl (2008) explored whether infants' weighting of prosodic boundary cues varies across languages. The experimental design of Seidl (2007) was applied with Dutch material to Dutch six-month-olds. Like the English-learning infants, the Dutch learners segmented the clausal sequence from the text passage. However, when the pause was neutralized in the familiarization sequences Dutch-learning infants failed to segment the clausal sequence from the text passage. Johnson and Seidl (2008) considered two interpretations. One is related to the strength of the prosodic cues. The magnitude of pitch change and preboundary lengthening might not have been salient enough to trigger the clause segmentation. Acoustic analyses of the stimuli had revealed that the saliency of the pitch reset and the pause duration at the clausal boundary differed in the materials used across the two languages. Compared to the English stimuli the pitch reset in the Dutch stimuli was only half the magnitude, whereas the pause was more than twice as long. However, the qualitative difference in the prosodic cues in the Dutch versus English stimuli might reflect language-specific boundary markings as Dutch compared to English generally tends to have a smaller pitch range (Collins & Mees, 1981; Willems, 1982). Therefore, Johnson and Seidl argued for a different interpretation: by six months, with increasing exposure to the native language, Dutch-learning and English-learning infants have developed a language-specific prosodic cue weighting that influences infants' clause segmentation procedures.

Taken together, these findings indicate that infants' sensitivity to acoustic cues as prosodic boundary markers is subject to a developmental change during early infancy – perhaps a change from a more general perceptually driven mechanism that relies on a broad set of acoustic cues to a mechanism that is attuned to the specific properties of the target language.

To further investigate the question of an early weighting of prosodic boundary cues, the present study set out to test infants learning German, a language in which we have – at least for adult-directed spontaneous speech – specific knowledge about the frequency of occurrence of prosodic cues at IPBs (Peters et al., 2005) the prosodic unit under investigation in this study. Moreover, from a study with German listeners, findings on adults' weighting of the relevant acoustic cues are available: in a prosodic judgment task (Holzgrefe-Lang et al., 2016) tested whether the presence of a pitch change and preboundary lengthening in the absence of a pause would suffice to signal a boundary. Listeners were presented with coordinated sequences of three names in different prosodic groupings. Their task was to judge the heard sequence as to whether or not it had an internal boundary. The German adult listeners identified the internal boundary when a pitch change as well as a preboundary lengthening, but no pause, were present in the sequence; however, pitch change alone or lengthening alone was not sufficient. In the present study the same linguistic materials were used to test whether German infants' processing of prosodic boundary cues is similar to that shown for German adults.

Hence, in contrast to previous studies, we did not present complex clauses (Johnson and Seidl, 2008; Nazzi et al., 2000; Seidl, 2007; Seidl and Cristià, 2008), but well-formed sequences that allowed for a precise acoustic characterization of the phonetic instantiation of the crucial prosodic boundaries which we considered to be the basis for a controlled acoustic manipulation of the stimuli. Thus, going beyond the previous studies with English- and Dutch-learning infants, the results of the infants tested in the current study could be related to findings from adults, allowing a direct comparison of German adults' and infants' cue weighting.

Again in contrast to previous studies, we did not test infants' segmentation, but their discrimination ability. We suggest that infants' attunement to specific properties of their native language is not only displayed in segmentation tasks as revealed by the work of Johnson and Seidl (2008), Seidl (2007) and Seidl and Cristià (2008). Instead, perceptual reorganization with respect to cue weighting should also be reflected in discrimination performance as has been shown for tone and phonemic contrasts in previous research (Mattock & Burnham, 2006; Mattock et al., 2008; Polka & Werker, 1994; Werker & Tees, 1984). If prosodic boundary cues are perceptually weighted individually, we assume that the less weighted information will contribute less to both discrimination and segmentation.

Experiment I/1 served as a baseline to ensure that in our experimental design German-learning infants perceive a boundary signaled by all three prosodic cues. In Experiment I/2 we investigated whether the specific combination of a pitch change and preboundary lengthening is sufficient for boundary detection. Hereby, the question whether pause is a necessary cue would be examined. We did not test a combination of two cues that included the pause cue, because we expected that eight-month-olds would discriminate between stimuli with and without a pause easily given that the pause is a rather strong acoustic cue, especially in a mere discrimination task. In fact, in a similar study with younger German infants (Wellmann et al., forthcoming) we found that even six-month-olds are able to use the pause cue. More precisely, a pitch change together with preboundary lengthening was not sufficient for six-month-olds, but the combination of pause and lengthening was. Thus, a pause, but not a pitch change was a necessary cue for six-month-olds. This finding moreover suggests that successful boundary detection depends on the specific cue constellation, rather than on the number of boundary cues provided.

After testing the combination of pitch change and preboundary lengthening, we examined the impact of each of the two as single cues: Experiment I/3 tested pitch change and Experiment I/4 preboundary lengthening.

4.2 Experiment I/1: A baseline study on infants' sensitivity to pitch change, preboundary lengthening, and pause

In Experiment I/1, we sought to ensure that eight-month-old German-learning infants are able to perceive a prosodic boundary that is signaled by the three main prosodic cues pitch change, preboundary lengthening, and pause. This would provide a verification of the experimental design and material as suitable for studying the perception of single prosodic boundary cues. As previous research has revealed that infants are sensitive to prosodic boundary information (e.g., Hirsh-Pasek et al., 1987; Nazzi et al., 2000), infants tested in Experiment I/1 should be able to perceive a prosodic boundary. Experiment I/1 aimed at creating a baseline for the subsequent experiments, in which the constellation of prosodic cues would be systematically varied.

4.2.1 Material and Methods

4.2.1.1 Participants

Twenty-four eight-month-old infants (12 girls) were tested. The mean age was 8 months, 16 days (range: 8 months, 3 days to 8 months, 30 days). All infants who participated in this and the following experiments were from monolingual German-speaking families, born full-term, and normal-hearing. Eleven additional infants were tested but their data were not included in the analysis for the following reasons: failure to complete the experiment (2), crying or fussiness (3), mean listening times of less than 3 seconds per condition (3), technical problems (2), and experimenter error (1).

4.2.1.2 Stimuli

The stimuli consisted of a sequence of three German names that were coordinated by *und* ('and'). The advantage of using coordinated structures instead of clauses lies in the better control of phonological and thus prosodic parameters. Thus, we used the following three names, which only contained sonorant sounds: *Moni, Lilli, Manu*. This allowed for a reliable measurement of the fundamental frequency – the acoustic correlate of the pitch contour.

Several recordings of the same sequences of names were made in an anechoic chamber equipped with an Audio-Technica AT4033A studio microphone, using a C-Media Wave soundcard at a sampling rate of 22050 Hz with 16 bit resolution. A young female German native speaker from the Brandenburg area was instructed to read the sequence in two different prosodic groupings, as indicated by different bracketing as in (1).

(1) a. (Moni und Lilli und Manu)
b. (Moni und Lilli) (und Manu)

Each name is a syntactic XP and is correspondingly set off by a phonological phrase boundary from the other names (Gussenhoven, 1992; Truckenbrodt, 1999, 2007a). Both sequences contain the same string and are disambiguated either by grouping all three names together as shown in (1a) or by grouping the first two names together and the final one apart as shown in (1b). This disambiguation employs the next higher level of the prosodic hierarchy, that is, the intonation phrase (IP). Thus, sequences of type (1a) are produced as a single IP, that is, without an internal boundary. In contrast, sequences of type (1b) are produced with an IPB after the second name, and consequently consist of two IPs. For each type of prosodic phrasing, the speaker produced six different acoustic realizations (tokens). The intended prosodic grouping was confirmed by two independent listeners who were naïve to the given bracketing.

The presence of the characteristics of an IPB in the sequences of names were confirmed by a detailed acoustic analysis of the recordings using PRAAT software (Boersma & Weenink, 2010). Measurements were carried out at the critical boundary position, namely on and after the second name. The analysis concentrated on the three acoustic correlates of prosodic boundary cues – fundamental frequency (F0), the duration of the final vowel, and the pause. Examples of the oscillogram and the fundamental frequency aligned with the segments for sequences without an IPB are shown in Figure 4.1(A), and for sequences with an IPB in Figure 4.1(B). Details of the acoustic analysis are presented in Table 4.1.

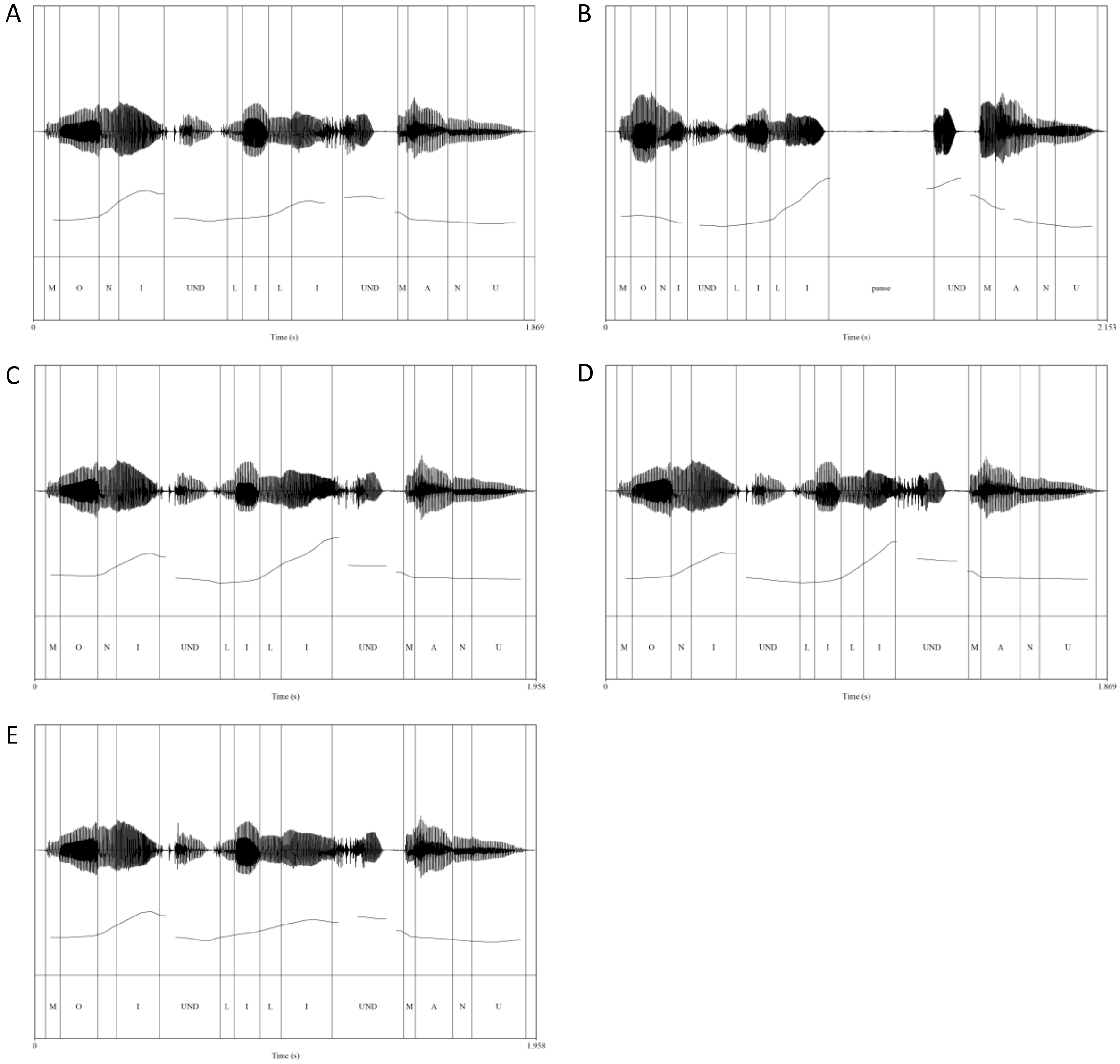

Figure 4.1 Oscillograms and pitch contours aligned to the text. (A) Sequence without an internal IPB used in Exp. I/1, (B) Sequence with an internal IPB used in Exp. I/1, (C) Sequence with pitch and preboundary lengthening used in Exp. I/2, (D) Sequence with pitch used in Exp. I/3, (E) Sequence with preboundary lengthening used in Exp. I/4.

The target word for the analysis was decomposed into four intervals corresponding to the phonetic segments, that is, the single consonantal and vocalic parts of the signal. F0 was measured at the midpoint of the first segment and at the position of the maximum F0 on the final vowel. The difference between these values was used to calculate the pitch change preceding the boundary. In sequences with an IPB, a pitch rise occurred, starting at the second syllable of the word and leading to a high boundary tone at the final vowel. This pitch change was 2.5 times greater in sequences with an IPB compared to sequences without an IPB (see Tab. 4.1 and Fig. 4.1(A) vs. Fig. 4.1(B)). A mean pitch reset of 25 Hz from the high boundary tone to the midpoint of the following conjunction *und* ('and') was measured in sequences with an IPB, whereas the pitch change was only 3 Hz at the same location in sequences without an IPB. Thus, the pitch reset was greater in sequences with a boundary, but compared to Seidl's (2007) stimuli the overall extent of the reset was rather small, as the conjunction *und* was also uttered on a high pitch level (see the pitch contour in Fig. 4.1(B)). First and foremost, in our stimuli the pitch cue in sequences with an IPB was provided by the pitch rise on the target name. Preboundary lengthening was calculated by measuring the length of the final vowel in both prosodic types. Transitions between the final vowel and the onset of the conjunction *und* were not included. The vowel duration was about 1.8 times longer before a boundary compared to the same vowel in the sequence without an internal IPB. The duration of the pause after the target name had a mean of 506 ms in sequences with an internal IPB. In contrast, no pause was present at this position in sequences without an internal IPB.

To summarize, sequences with an internal IPB clearly revealed the acoustic correlates of the three main prosodic boundary cues similar to IPBs in German spontaneous speech (Peters et al., 2005). A pitch rise occurred on the target name followed by a pitch reset after a pause. Preboundary lengthening was observed at the final vowel of the target name.

Table 4.1 Mean values and range of the acoustic correlates of prosodic boundary cues on and after the second name in the experimental stimuli.

acoustic correlate	**without an internal IPB** [Moni und Lilli und Manu]	**with an internal IPB** [Moni und Lilli] # [und Manu]
pitch rise in Hz	88 (77-110)	220 (197-240)
pitch rise in semitones	6.7 (5.8-8.2)	14.0 (12.8-14.6)
maximum pitch in Hz	277 (264-293)	397 (371-422)
final vowel duration of name 2 in ms	99 (91-110)	175 (162-186)
pause duration after name 2 in ms	0	506 (452-556)

Following the acoustic analyses the different recordings (tokens) were used to create sound files for presentation as trials during the experiment. For each prosodic type, the six tokens were randomly concatenated with a silent interval of 1 s inserted between them. In this way, six sound files per prosodic grouping were created such that each file consisted of a different order of tokens. The average duration of tokens without an IPB was 1.76 s (range: 1.71-1.87 s), while it was 2.16 s (range: 2.13-2.2 s) for tokens with an IPB.

To match the sound files of the two prosodic types with respect to length the number of tokens within each file was varied. Files of the grouping with an IPB contained six tokens and had an average duration of 18.97 s. However, files of the condition without an IPB contained seven tokens (i.e., one random token was repeated), leading to an average duration of 19.32 s (range: 19.16-19.43 s).

4.2.1.3 Procedure

The HPP including a familiarization phase (Hirsh-Pasek et al., 1987; Jusczyk and Aslin, 1995) was used in this and all subsequent experiments. During the experimental session, the infant was seated on the lap of a caregiver in the center of a test booth. The caregiver listened to music over headphones to prevent influences on the infant's behavior. Furthermore, she was instructed not to interfere with the infant’s behavior during the experiment. The experimenter sat in an adjacent room, where she observed the infant’s behavior on a mute video monitor and controlled the presentation of the visual and the acoustic signals by a button box.

Three lamps were fixed inside the booth: a green one on the center wall, and red ones on each of the side walls. Directly above the green lamp on the center wall was an opening for the lens of a video camera. Behind each of the red lights a JBL Control One loudspeaker was mounted. Each experimental trial started with the blinking of the green center lamp. When the infant oriented to the green lamp, it was turned off and one of the red lamps on a side wall started to blink. When the infant turned her head towards the red lamp, the speech stimulus was started, delivered via a Sony TA-F261R audio amplifier to the loudspeaker at the same side. The trial ended when the infant turned her head away for more than 2 s, or when the end of the speech file was reached. If the infant turned away for less than 2 s, the presentation of the speech file continued but the time spent looking away was not included in the total listening time. The whole session was digitally videotaped. The experimenter’s coding was recorded and served for the calculation of the duration of the infant’s head turns during the experimental trials (for comparable experimental setups, see Höhle et al., 2009; Höhle et al., 2006).

Half of the infants were familiarized to the sequences without an IPB (Group 1), while the other half were familiarized to the sequences including an IPB (Group 2). The familiarization was set such that at least 20 tokens in each familiarization condition were presented, that is, when familiarized to sequences without an IPB the familiarization lasted until the infant had accumulated 55 s of listening time. For the familiarization with an IPB the criterion was 63 s of accumulated listening time. This requirement was chosen to match the familiarization duration used in Nazzi et al. (2000).

Two different kinds of familiarization were chosen to control for a possible effect of the prosodic structure of the sequences presented. One might hypothesize that a familiarization to sequences without an internal IPB might be more effective. This is supported by Nazzi et al.’s (2000) findings that infants recognize word sequences that constitute a prosodic unit better than sequences that are a non-unit like our sequences with an IPB. Therefore, we planned to compare the data of both familiarization groups.

The familiarization was followed by a test phase that comprised twelve test trials. In six trials, the sound files without an internal IPB were presented, in the other six trials the sound files of the sequences with an IPB. Thus, half of the test trials contained exactly the same sound files that the infants had previously heard during familiarization, whereas the other half consisted of sound files with the type of prosodic grouping that had not been presented during familiarization. The test trials were grouped in three blocks of four trials each (two with and two without an internal IPB in a random order). Additionally, within each block the side of presentation of the sequences of the two prosodic types was counterbalanced so that the prosodic condition and the side of presentation were not associated. The duration of each experimental session depended on the infant's behavior and varied between four and six minutes.

4.2.2 Results and Discussion

Mean listening times to the test trials with and without an IPB were calculated for each infant. Because all listening times were shorter than 18.97 s (the maximum trial length in the condition with an IPB), an adjustment of the listening times to the longer duration of the trials without an IPB was not necessary.

On average, infants listened for 6.32 s (*SD* = 2.39) to the familiarized prosodic grouping, and for 7.13 s (*SD* = 2.12) to the novel prosodic grouping (see Fig. 4.2). This difference was significant, *t*(23) = 2.30, *p* = .031, two-tailed. Eighteen out of 24 infants had longer listening times to the novel test items. A repeated-measures ANOVA with the within-subject factor familiarity (familiarized versus new prosodic pattern) and the between-subject factor prosodic type (familiarization with versus without an internal IPB) showed a main effect of familiarity, *F*(1,22) = 5.36, *p* = .030, and a main effect of prosodic type, *F*(1,22) = 4.44, *p* = .047, but no significant interaction between prosodic type and familiarity, *F*(1,22) = 1.237, *p* = .278.

A further analysis of the data separated by prosodic type heard during familiarization was conducted. This analysis revealed a significant preference for new test items in the group familiarized with the sequences without an IPB, *t*(11) = 2.40, *p* = .035. The mean listening time to the new prosodic pattern was 6.48 s (*SD* = 1.23) and to the familiarized prosodic pattern 5.29 s (*SD* = 1.46). No such preference was present in the group familiarized with sequences including an IPB, *t*(11) = .860, *p* = .408. Infants in this group listened to the novel test trials on average for 7.77 s (*SD* = 2.64) and to the familiar test trials for 7.36 s (*SD* = 2.73). Mean listening times separated by familiarization group are depicted in Figure 4.3.

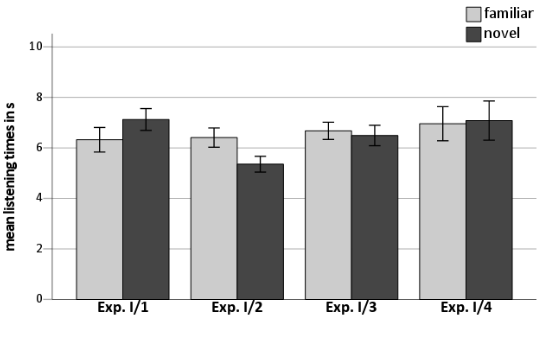

Figure 4.2 Mean listening times for Experiment I/1-4. Error bars indicate +/- 1 SE.

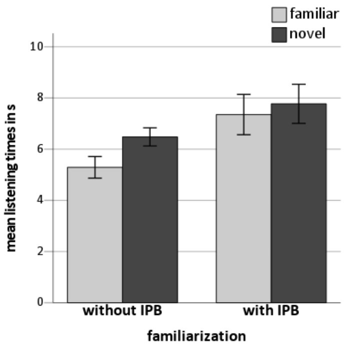

Figure 4.3 Mean listening times for Experiment I/1, separated by familiarization group. Error bars indicate +/- 1 SE.

Experiment I/1 served as a baseline study to ensure that the stimuli – sequences of names that have two different prosodic groupings – and our experimental design are suitable for studying the perception of prosodic boundary cues in German-learning infants.

After being familiarized with one of the two prosodic phrasings, eight-month-old infants showed an overall preference for the novel prosodic grouping. Thus, German-learning infants are able to discriminate the two prosodic groupings. Even though we found no significant interaction between prosodic type and familiarity, a separate analysis of the two familiarization groups revealed that the difference in listening times was significant only when the familiarization strings did not have an internal IPB. Thus, discrimination of sequences with versus without an IPB was affected by the prosodic type heard during familiarization. How can we explain this effect? During familiarization the infants' task is to build up a representation of the auditory stimulus, to which they will compare the test stimuli. Presumably, infants can more easily build up representations of sequences without an IPB because these are easier to process and memorize, as Nazzi et al.'s (2000) study demonstrated. Secondly, both familiarization conditions differ in the number of IPs: Stimuli played to Group 1 do not contain any prosodic boundary cue and, hence, only consist of a single IP, that is, one prosodic unit. In contrast, stimuli presented to Group 2 were sequences including an IPB, which splits the sequences into two separate IPs, that is, two prosodic units.

A study by Mandel et al. (1994) suggests that infants at the age of two months already perceive prosodic units as an organizational unit in the speech stream. Infants detected phonetic changes in word sequences when the words were prosodically grouped into a major linguistic unit, but not when the words were presented as isolated words in a list or as a fragment of two adjoining clauses. Mandel et al. argued that the organization of words in a prosodic unit helps infants to process and memorize the speech signal. For our experiment this implies that the representation of the familiarization sequence is built up more easily when the sequence consists of a single prosodic unit, like our sequences without an internal IPB. These are – compared to the sequences with an IPB – easier to process during the familiarization phase and thus can be better remembered during the test phase.

The difference found in the two familiarization groups motivated a modification of the experimental design implemented in the subsequent experiments. As a full design with two separate familiarization conditions was not relevant to our research question, we decided to only use strings without an internal IPB as familiarization stimuli. In doing so, we chose the condition that yielded the most robust results.

In sum, Experiment I/1 showed that eight-month-old German-learning infants are sensitive to the presence of an IPB in short coordinated sequences of names when the IPB is marked by the acoustic correlates of the main prosodic boundary cues pitch change, preboundary lengthening, and pause. Hence, not only clauses – like those that were used in previous studies (e.g., Hirsh-Pasek et al., 1987;

Nazzi et al., 2000; Schmitz, 2008; Seidl, 2007) – are suitable for investigating infants' sensitivity to prosodic boundaries. Rather, coordinate structures, which can be carefully controlled for phonological parameters, may serve as stimuli to characterize the impact of each prosodic boundary cue in a discrimination task.

The subsequent experiments contain only one kind of familiarization, namely the familiarization to sequences without an IPB. In these experiments, the number of prosodic boundary cues in the stimuli is reduced stepwise. This is done to determine whether infants' discrimination ability remains or is disturbed when different constellations of prosodic boundary cues are given.

4.3 Experiment I/2: Sensitivity to pitch change and preboundary lengthening

In Experiment I/2, we investigated infants' sensitivity to two of three prosodic boundary cues, namely pitch change in combination with preboundary lengthening. Specifically, we asked whether the pitch change and the lengthening of the preboundary vowel suffice as boundary cues or whether the pause is a necessary prosodic boundary cue. If pause is a necessary cue for the discrimination of two prosodic groupings, infants were not expected to show significantly different listening times to novel versus familiar test items. In contrast, if pitch and preboundary lengthening are sufficient cues, we expected a significant listening preference.

4.3.1 Material and Methods

4.3.1.1 Participants

Sixteen eight-month-old infants (8 girls) were tested. The mean age was 8 months, 11 days (range: 8 months, 1 day to 9 months, 8 days). Ten additional infants were tested but their data were not included in the analysis for the following reasons: crying or fussiness (6), mean listening times of less than 3 seconds per condition (2), and noise (2).

4.3.1.2 Stimuli

In Experiment I/2, again sequences with and without an internal prosodic boundary cues were presented. The stimuli without any boundary cues were the same as the stimuli used in Experiment I/1. The sequences containing a pitch change and preboundary lengthening were construed from the sequences without an IPB by acoustic manipulation – according to the values that had been measured in the sequences with an IPB recorded for Experiment I/1. Hereby, we created two types of stimuli that only differed in fundamental frequency and duration at the critical boundary position, that is, on the second name. Apart from that, the sequences of both prosodic types were acoustically identical.

The manipulation was carried out with the PRAAT software. For duration, the final vowel of the target name was lengthened to 180%. This factor was chosen because in Experiment I/1 the crucial vowel was on average 1.8 times longer in sequences with an IPB than in sequences without an IPB.

For the manipulation of the pitch contour, first the sequences without an IPB were stylized (2 semitones), that is, the number of pitch points was reduced. The reference values of the fundamental frequency were measured on the target name in the sequences with an internal IPB from Experiment I/1 – at the midpoints of the four segments [l], [ɪ], [l], and [i] and at the position of the maximum pitch present on the preboundary vowel. Then, pitch points with the mean values at these time points were inserted at the same positions into the sequences without an IPB. We obtained new stimuli for the prosodic type with pitch and preboundary lengthening. They contained a natural sounding pitch rise of 212 Hz (13.65 semitones) and a preboundary lengthening with a factor of 1.8. The pitch contour and wave form of a sequence with manipulated pitch and lengthening are depicted in Figure 4.1(C).

To avoid comparing natural with acoustically manipulated stimuli we carried out a slight acoustic manipulation of the sequences without an IPB as well: a stylization of the pitch contour (2 semitones). After acoustic manipulation, all sequences were resynthesized using the PSOLA function in PRAAT.

Six differently ordered speech files with the same set of tokens in each prosodic condition were created from the acoustically manipulated sequences. The speech files of the condition without an IPB contained seven tokens (i.e., one random token was repeated) and had an average duration of 18.33 s (range: 18.23-18.43 s). The files of the condition with added pitch and lengthening cues also contained seven tokens (again one random token was repeated) and had an average duration of 18.81 s (range: 18.79-19.01 s).

4.3.1.3 Procedure

The procedure was the same as in Experiment I/1 with a modification concerning the familiarization phase. Infants in Experiment I/2 were only familiarized to sequences without an IPB, but not to sequences with boundary cues. The familiarization lasted until at least 20 sequences had been presented leading to a minimum duration of 52 s.

4.3.2 Results and Discussion

Infants oriented on average for 6.41 s (*SD* = 1.53) to the familiarized prosodic grouping, and for 5.36 s (*SD* = 1.25) to the new prosodic grouping (see Fig. 4.2). This difference was significant, $t(15) = -3.59$, $p = .003$, two-tailed. Thirteen of 16 infants had longer listening times to the familiar test items.

Experiment I/2 tested whether German-learning infants still perceive an IPB when only a subset of prosodic cues, pitch change, and preboundary lengthening, is present. A significant familiarity effect was displayed indicating that the infants were able to discriminate the stimuli of the two prosodic patterns in Experiment I/2. Interestingly, the direction of preference reversed from Experiment I/1 to Experiment I/2. While infants in Experiment I/1 preferred to listen to the novel prosodic pattern, in Experiment I/2 the familiar pattern was preferred. According to the model by Hunter and Ames (1988), this shift in preference can be explained by higher task demands in Experiment I/2. Hunter and Ames claimed that the direction of preference is affected by three factors: age, duration of familiarization, and task difficulty. As we held the first two factors constant, we assume that the shift in preference from Experiment I/1 to Experiment I/2 is caused by increased task difficulty: if only two instead of three prosodic cues mark the difference between the stimuli, it becomes harder to distinguish both conditions as less information is available. In turn, the task of discriminating the two prosodic patterns is more difficult and leads infants to a preference for the familiar sequences. Hence, for German eight-month-olds pitch change and preboundary lengthening in combination are sufficient. Pause is not a necessary boundary cue, however, processing different prosodic groupings without the information provided by the pause cue seems to be more demanding.

4.4 Experiment I/3: Sensitivity to pitch change

Experiment I/2 showed that German infants are able to discriminate the two prosodic groupings when a boundary is signaled by a pitch change and preboundary lengthening in combination. In Experiment I/3 we asked whether only one cue, the pitch change, is sufficient for German eight-month-olds to perceive a boundary.

4.4.1 Material and Methods

4.4.1.1 Participants

Seventeen infants (7 girls) were tested. The mean age was 8 months, 13 days (range: 8 months, 4 days to 8 months, 29 days). Six additional infants were tested but their data were not included in the analysis for the following reasons: crying or fussiness (4), and mean listening times of less than 3 seconds per condition (2).

4.4.1.2 Stimuli

In Experiment I/3, sequences without an IPB and sequences with an inserted pitch rise were contrasted. For the condition without an IPB the same sequences as in Experiment I/2 were used. For the condition with added pitch cue a manipulation of the pitch contour was carried out similar to that in Experiment I/2: a pitch rise was inserted on the second name of the six sequences without an IPB. In contrast to the stimuli in Experiment I/2 no duration manipulation was conducted. Thus, the pitch change with the high boundary tone was the only signal of an IPB (see Fig. 4.1(D)). From these pitch-manipulated sequences six differently ordered speech files were created with seven tokens per file (i.e., one of the six exemplars was randomly repeated). The speech files of the condition without an IPB were the same as in Experiment I/2. The average duration of the speech files was the same in both prosodic conditions as there was no duration manipulation (*M* = 18.33 s; range: 18.23-18.43 s).

4.4.1.3 Procedure

The procedure was the same as in Experiment I/2. Infants were familiarized to sequences without an IPB until at least 20 sequences had been presented. This led to a minimum duration of 52 s.

4.4.2 Results and Discussion

Infants listened on average for 6.68 s (*SD* = 1.41) to the familiarized prosodic grouping and for 6.49 s (*SD* = 1.67) to the novel prosodic grouping (see Fig. 4.2). This difference was not significant, *t*(16) = .522, *p* = .609. Ten of 17 infants had longer listening times to the familiar test items.

In Experiment I/3 only a pitch rise indicated a different prosodic grouping. Neither a pause nor lengthening of the preboundary vowel was present. The infants did not differentiate between sequences with added pitch cue and sequences without an IPB. Hence, the presence of a pitch change alone is not sufficient for German infants to perceive a prosodic boundary.

Apart from the specific cue constellation presented, Experiment I/3 generally differs from Experiment I/2 with regard to the number of IPB cues provided in the stimuli, that is, whereas in Experiment I/2 two boundary cues were available, in Experiment I/3 we only inserted one cue. Hereby, the boundary is generally less marked in Experiment I/3. The failure to discriminate the two conditions could hence be due to the mere number of cues being relevant for boundary detection, instead of the specific kind of cue or cue constellation (but see General discussion, Chapter 4.6).

4.5 Experiment I/4: Sensitivity to preboundary lengthening

German eight-month-olds are able to perceive an IPB when a pitch change and preboundary lengthening occur together (Exp. I/2) but not when only a pitch change is present (Exp. I/3). Experiment I/4 tested whether preboundary lengthening as a single boundary cue is sufficient.

4.5.1 Material and Methods

4.5.1.1 Participants

Sixteen infants (8 girls) were tested. The mean age was 8 months, 10 days (range: 7 months, 30 days to 8 months, 29 days). Six additional infants were tested but their data were not included in the analysis for the following reasons: failure to complete the experiment (1), crying or fussiness (2), and mean listening times of less than 3 seconds per condition (3).

4.5.1.2 Stimuli

In Experiment I/4, sequences without an IPB and sequences with inserted preboundary lengthening were contrasted. For the condition without an IPB the same sequences as in Experiment I/2 were used. For the condition with inserted preboundary lengthening a manipulation of the duration of the final vowel was carried out similar to that in Experiment I/2: in six exemplars of the sequences without an IPB the final vowel was lengthened to 180 % (see Fig. 4.1(E) for an example). The sequences were concatenated in a random order to speech files.

The speech files of the condition without an IPB were the same as in Experiment I/2. They contained six different tokens and had an average duration of 18.33 s (range: 18.23-18.43 s). The speech files of the condition with preboundary lengthening also contained six tokens and lasted for 18.89 s on average (range: 18.79-19.01 s).

4.5.1.3 Procedure

The procedure was the same as in Experiment I/2. Infants were only familiarized to sequences without an IPB until at least 20 sequences had been presented. This led to a minimum duration of 52 s.

4.5.2 Results and Discussion

The listening time in one individual trial of the condition with the lengthening cue exceeded the duration of the longest speech file in the condition without an IPB. Therefore, the listening time in this trial was reduced to the maximum trial length of sequences without an IPB, which was 18.43 s.

The mean listening time to the familiarized prosodic grouping was 6.96 s (*SD* = 2.7) and to the new pattern 7.08 s (*SD* = 3.1) (see Fig. 4.2). This difference was not significant, $t(15) = -.221$, $p = .828$. Nine of 16 infants had longer listening times to the familiar test trials.

Experiment I/4 suggests that preboundary lengthening as a single cue is not sufficient to trigger the perception of a prosodic boundary in German eight-month-old infants. However, in combination with a pitch cue, as tested in Experiment I/2, it becomes an effective boundary marker. As for Experiment I/3, we also have to consider that the insufficiency of preboundary lengthening alone compared to its effectiveness in combination with a pitch change could also be explained by the number of cues (but see General discussion, Chapter 4.6).

4.6 General discussion

The aim of the present study was to specify the relevance of pitch change and preboundary lengthening as combined and as single prosodic cues in German-learning infants' perception of major prosodic boundaries. Experiment I/1 showed that eight-month-olds are able to discriminate different prosodic groupings – specifically, familiar sequences without a prosodic boundary from unfamiliar sequences with a prosodic boundary – when the boundary is clearly marked by all three boundary markers.

In further experiments stimuli were acoustically manipulated with respect to pitch and preboundary lengthening. We focused on investigating infants' processing of boundaries in the absence of the pause cue. Pauses are perceptually highly salient and we assumed that in a discrimination task like ours, infants would easily detect the presence of a pause. Especially in short coordinated structures as used in this study pauses are easy to notice as they constitute approximately a fourth of the overall duration of the sequence. Furthermore, we know from other studies (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Schmitz, 2008; Wellmann et al., forthcoming) that infants by the age of six to ten months are highly sensitive to pauses.

When we manipulated the stimuli such that only a pitch change and preboundary lengthening indicated the presence of an IPB (Exp. I/2), infants still detected the boundary. We concluded that pause is not necessary, but it seems to ease infants' processing. This was indicated by a shift in preference from a novelty effect in Experiment I/1 to a familiarity effect in Experiment I/2. We argued that higher task demands in Experiment I/2 are responsible for the preference for familiar stimuli (see Hunter and Ames, 1988). In Experiments I/3 and I/4 the impact of the single prosodic cues pitch change and preboundary lengthening were tested. Sequences with pitch as a single cue (Exp. I/3) were not differentiated from sequences without any boundary cue. Nor was preboundary lengthening alone (Exp. I/4) sufficient to trigger the perception of a boundary. This might indicate that infants do not take single cues into account, as cue combinations are very frequent whereas the occurrence of single cues is rather rare (Peters et al., 2005). However, the weighting of prosodic boundary cues might depend on the strength of the specific cue, that is, its phonetic magnitude. When implementing the cues in Experiments I/2-4 we used the acoustic values measured in natural sequences that contained all three cues. It is conceivable that the specific strength of each cue in production depends on the constellation of cues, that is, when a cue occurs alone or in a subset its magnitude might be larger than when it occurs together with all main cues. Thus, it remains possible that a larger pitch rise in Experiment I/3 or longer preboundary lengthening in Experiment I/4 might have been sufficient to trigger boundary perception by a single cue. We also considered the reduced number of boundary cues as an explanation for the insufficiency of the single cues compared to their occurrence in combination. However, in a study with six-month-olds (Wellmann et al., forthcoming) we found that pause, but not a pitch change, was sufficient although the number of cues was kept constant. Therefore, we argue that the specific cue constellation, and not the number of cues, is decisive for the detection of a boundary.

Another restriction when interpreting the data concerns the fact that the stimuli presented during the test phase differed across experiments in the presence or absence of boundary cues, but potentially also with respect to their naturalness. Thus, infants' different performance patterns could be due to infants' disliking of one kind of stimuli in one but not the other experiment. Pitch change or preboundary lengthening might be effective as single cues when produced naturally, but infants could find stimuli with a single inserted cue odd, thus, would not pay attention and consequently fail to discriminate test stimuli. Hereby, infants' weighting and their liking of stimuli might be confounded. However, when editing the stimuli with inserted cues, we took special care to create stimuli that are perceptually distinguishable, but comparably natural sounding in all experiments. Hence, we rather argue that the different performance patterns suggest that perception depends on the specific cue constellation: Pitch change and preboundary lengthening in combination are sufficient to trigger boundary perception in German eight-month-old infants and hence, pause is not a necessary cue. Whether pitch change or preboundary lengthening is a necessary cue cannot be answered from these experiments. Still, both of them are not sufficient as single boundary cues: when they occur individually, stimuli are not differentiated from sequences without prosodic boundary marking – at least if the single cues are presented with the same acoustic parameters as when they occur combined.

In summary, two parallels of these findings to previous research are obvious: first, they resemble findings on the processing of these cues in German adults (Holzgrefe-Lang at al., 2016), and secondly, they show a strong overlap with the findings by Seidl (2007) for English-learning infants. Both parallels will be discussed separately in the following section.

To our knowledge, the present study is the first that has used the same material with infants that had previously been used with adults in a prosodic judgment task (Holzgrefe-Lang et al., 2016). In this study, adults were asked to interpret the aurally presented sequences as having no internal boundary *[a and b and c]*, or as having an internal boundary after the second name *[a and b] [and c]*. The effects that the specific prosodic cues had on these decisions mirror the pattern we found with the German-learning infants: Sequences that provided pitch change or preboundary lengthening as single cues either were judged as having no boundary or listeners performed at chance level. However, when a combination of pitch change and preboundary lengthening occurred in the sequence, they were clearly identified as consisting of two prosodic units. Moreover, infants' behavior in our study is in line with the distribution of prosodic boundary cues found in spontaneous speech of German adults (Peters et al., 2005): First, the majority of IPBs are marked by a coalition of cues. Secondly, compared to pitch change and preboundary lengthening, pause is a rather rare marker of IPBs. This suggests that pause is not reliable and listeners should be able to cope without it.

It is rather surprising that the experiments with the adults and the infants show exactly the same pattern of results with respect to cue effectiveness even though the tasks that had to be performed by the participants were clearly different: while the adults had to exploit the acoustic information to assign a prosodic phrasing to the utterances, the children only had to discriminate between the different prosodic contours. If we consider these findings in the light of Johnson and Seidl's (2008) assumption that a language-specific weighting of prosodic boundary cues takes place, our results suggest that the German eight-month-olds have already attuned to the German system as they show a parallel pattern of responding to the cues to that of adults. Furthermore, our results indicate that cue weighting leads to a perceptual reorganization that has an effect on the ability to discriminate verbal materials containing the relevant phonetic information.

Additional empirical support for this conclusion is required and may come from crosslinguistic studies that compare children learning languages that exhibit relevant differences in the acoustic instantiation of prosodic boundary cues. In addition, one may compare the current findings to the performance of younger infants. This would allow a developmental trajectory to be followed from a language-general perceptual system that is not yet fully adapted to the properties of the phonological system of the ambient language to a language-specific perceptual system that is attuned to these properties.

Crosslinguistic research in the area of the processing of prosodic boundaries is still sparse. Additionally, a crosslinguistic comparison may be impeded because of differences in the experimental material of our and previous studies: we used coordinated noun phrases, whereas previous studies on English and Dutch (Johnson and Seidl, 2008; Seidl, 2007) presented clauses. Even though both kinds of material have a different syntactic structure, the prosodic structure is similar. Clause boundaries in Seidl's (2007) and Johnson and Seidl's (2008) studies coincide with IPBs. In our sequences of names each name forms a phonological phrase. To convey the intended internal grouping, that is, separating the first two names from the third, our speaker needed to group the first two names into a larger prosodic unit by producing a larger prosodic boundary after the second name. In line with current models of prosodic phrasing (Gussenhoven, 1992; Truckenbrodt, 1999, 2007a) we argue that therefore the first two names of the internally grouped sequences constitute an intonation phrase. This account is supported by the acoustic analysis we carried out on the respective IPB cues. Hence, even though the stimuli differ across studies, the prosodic level under investigation is comparable allowing us to compare ours and previous findings crosslinguistically. German infants' behavior compared to American six-month-olds' (Seidl, 2007) shows no indications of crosslinguistic variation. Like the German infants in our study, the six-month-old American infants did not provide any evidence of detecting a boundary when it was solely cued by pitch change or preboundary lengthening, but only if a combination of these cues occurred in the stimuli. However, given the high overlap in the prosodic systems of English and German, the lacking crosslinguistic variation could simply reflect the fact that the two languages do not differ crucially in the area under investigation.

However, a comparison of the results of the experiments with German- and English-learning infants on the one hand and Dutch-learning infants on the other gives some indications of crosslinguistic variation. While the six-month-old Dutch infants tested by Johnson and Seidl (2008) needed a pause to detect the prosodic boundary, the German and American infants were able to perceive a boundary with pitch change and preboundary lengthening only. This might indicate a true crosslinguistic variation between German and Dutch and English and Dutch.

Regarding the difference observed between the German and Dutch infants’ reliance on the prosodic cues, we have to take into account that it may arise from a purely developmental change. The Dutch infants were two months younger than the German ones. It is thus possible that older Dutch babies will be able to detect prosodic boundaries that are not marked by pause. In addition, it is feasible that German six-month-olds will not detect a prosodic boundary when no pause is present. This would suggest a developmental change in prosodic cue perception from six to eight months in Dutch and German infants. Future studies comparing German and Dutch infants of the same age will have to disentangle whether the observed difference is due to crosslinguistic variation or is caused by developmental aspects.

Regarding the difference between English- and Dutch-learning infants’ sensitivity to prosodic boundary markers, Johnson and Seidl (2008) took this as an indication of the emergence of a language-specific cue weighting, as the results reflected differences in the way that the prosodic boundaries were marked in the Dutch material and the English material, with a longer pause but smaller pitch reset in Dutch as compared to English. Additional evidence for this view comes from the study by Seidl and Cristià (2008), which revealed that younger, four-month-old English-learning infants only rely on a combination of all three cues. The authors argued that younger infants’ perception reflects holistic mechanisms that do not depend on language-specific factors. Later in development, infants follow an analytical segmentation strategy that implies language-specific processing (Seidl, 2007). This indicates a developmental shift from four to six months of age. Based on this reasoning, a further study with German-learning infants younger than the age tested in our study would be necessary to provide more evidence for this kind of developmental change.

Furthermore, it would be highly interesting to look at languages in which the way prosodic boundaries are marked is more different than in the closely related languages English, German, and Dutch. The advantage of the linguistic material used in this study is that it can easily be adapted to other languages. One relevant language to look at would be French. Two features might lead to a greater saliency of preboundary lengthening. First, French does not have lexical stress and thus has no pitch accents. In languages without pitch accents syllable duration is much less varied within phrases. Secondly, French is a syllable-timed language. The inventory of syllable types is smaller in syllable-timed than in stress-timed languages. Smaller syllable inventories comprise simpler syllables, whereas languages with more syllable types tend to have heavier syllables (Ramus et al., 2000). Consequently, syllable duration is less varied in syllable-timed than in stress-timed languages. Both aspects, no lexical stress and a smaller syllable inventory, lead to the assumption that whenever syllables are lengthened, namely phrase-finally, this provides a clear acoustic contrast to phrase-internal syllable durations. Empirical evidence for a greater phonetic extent of preboundary lengthening comes from a production study with German and French adults by Féry et al. (2011). They found that the difference in duration between phrase-internal and phrase-final words was significantly higher in French speakers than in German speakers, who used preboundary lengthening to a smaller degree. Thus, preboundary lengthening might be a more important cue for the perception of prosodic boundaries in French adults and infants compared to the speakers and learners of the languages looked at so far. Again, this question is left open for further research.

Also, tone languages that deploy lexical tones on each syllable should be studied (e.g., Chinese). Where pitch is used to encode lexical distinctions, its role in encoding boundaries is reduced (Fernald & McRoberts, 1996). Therefore, one can hypothesize that infants acquiring such a tone language focus more on other boundary cues, like pause and preboundary lengthening. Pitch would then be perceptually weighted less in this kind of tone language than in non-tone languages.

The results of our study contribute in an important way to our understanding of how prosodic information may support children's early phrasing of incoming linguistic material and hence provide further evidence for the prosodic bootstrapping account. Fernald and McRoberts (1996) outlined the unreliability of prosodic cues due to their multiple functions. Our results as well as Seidl's (2007) data show that infants only consider a combination of at least two cues as a marker for a prosodic boundary – and even younger infants rely on the convergence of all cues that serve as prosodic boundary markers (Seidl and Cristià, 2008). With these constraints infants have a powerful mechanism to make specific use of these correlations of cues as boundary markers and to ignore the same acoustic information when it is not accompanied by correlating cues.

5 Add-on Study I Experiment I/5: Assessing the quality of prosodic boundary cues – natural versus inserted cues

The study presented in the previous chapter revealed that eight-month-old German-learning infants are able to discriminate sequences with an IPB from sequences without an IPB in an HPP experiment. This is possible when the three main boundary cues pitch change, final lengthening, and pause signal the IPB (Exp. I/1), but also when only a subset of cues, pitch change and final lengthening, is present (Exp. I/2). In both experiments, discrimination ability was evidenced by significant differences in infants' listening times to novel versus familiar trials. However, the direction of infants' listening preference differed between the two experiments: while infants showed a preference for novel material in the experiment with all three boundary cues, a familiarity effect occurred in the experiment that provided only two boundary cues. Hunter & Ames (1988) consider different factors accounting for familiarity effects: age, stimulus complexity, exposure time, processing speed, and task difficulty. In the case of the familiarity effect in the experiment that used a subset of cues, Wellmann, Holzgrefe, Truckenbrodt, Wartenburger, & Höhle (2012) argued for higher task demands since the differences between the stimuli were more subtle compared to those with the fully marked IPB.

Still, interpreting the opposing directions of preference one should consider the different materials across both experiments, naturally produced stimuli on the one hand and acoustically manipulated stimuli on the other. The experiment that used sequences with inserted cues revealed a familiarity effect: specifically, all infants were familiarized with sequences without IPB cues and tested on their discrimination of the same sequences without cues and sequences with inserted cues. Infants preferred to listen to the familiar test trials, which means, that they listened significantly shorter to sequences with inserted pitch and lengthening. In principle, this could reflect a rejection of the stimuli, that is, infants might simply dislike the manipulated stimuli due to unnaturalness. As we took special care implementing the cues and given that we also carried out a slight pitch manipulation to sequences without an IPB I assume that this explanation is rather unlikely. However, to rule out that a potential disliking of manipulated stimuli caused the shift in listening preference, in this chapter I present an experiment that modified the stimuli that were used in Wellmann et al.'s (2012) baseline experiment by using sequences with inserted cues instead of naturally occurring cues.

5.1 Materials and Methods

To create the condition *Inserted Cues*, the full set of the three main boundary cues was inserted to the original sequences without an IPB. These were presented in contrast to sequences of the condition *Without an IPB* (but stylized pitch, see below) to test infants' discrimination. I expected a novelty preference similar to the outcome of Wellmann et al.'s (2012) experiment with the naturally occurring three main cues, if the quality of the boundary cues (natural vs. inserted) did not matter. Specifically, the group of infants familiarized with sequences without an IPB were expected to show longer listening times to sequences with inserted cues. Conversely, the group of infants familiarized with sequences with inserted cues should listen longer to sequences without IPB cues. Such a finding would support the interpretation that the familiarity effect found in the experiment with inserted pitch and lengthening is not an artefact of using stimuli that were acoustically manipulated and legitimate the interpretation that different directions of preference stem from differences in task difficulty. However, if the concern is true that infants dislike the manipulated stimuli per se, infants should generally listen shorter to sequences with the three inserted cues irrespective of the kind of familiarization.

5.1.1 Participants

Twenty-five eight-month-old infants (13 girls) were tested. The mean age was 8 months, 14 days (range: 8 months, 2 days to 8 months, 29 days). Twelve additional infants were tested but their data were not included in the analysis for the following reasons: failure to complete the experiment (1), crying or fussiness (4), mean listening times of less than 3 seconds per condition (5), technical problems (1), and infant's first name being identical to one of the stimulus names *(Lilli*, 1). All infants were from monolingual German-speaking families, born full-term, and normal-hearing. Informed written consent was obtained from all parents.

5.1.2 Stimuli

Similar to the stimuli used in Chapter 4, the speech materials consisted of lists of coordinated names in two different prosodic groupings: sequences without an internal IPB, *[Moni und Lilli und Manu]*$_{IP}$, and sequences with an internal IPB, *[Moni und Lilli]*$_{IP}$ *[und Manu]*$_{IP}$. The latter were derived from the original sequences without an IPB from Experiment I/1 to which the three main boundary cues (a pitch rise, final lengthening, and a pause) were added on and after the second name by acoustic manipulation with the Software PRAAT (Boersma & Weenink, 2010). The insertion of pitch and lengthening cues was employed in the same manner as carried out for the stimuli used in the study by Wellmann et al. (2012). First, the pitch contour was stylized and then pitch points corresponding to the values measured in natural sequences with an IPB were added. This resulted in a mean pitch rise of 218 Hz (13.9 semitones, measured as the difference between the maximum and the minimum pitch on the second name) with a high boundary tone H% of 394.8 Hz on the second name. Second, the final vowel was lengthened to 180%. Third, a pause was inserted at the prospective boundary position in the following way: any co-articulation between the second name and the subsequent conjunction, that is, the section of formant transition from the final vowel [i] to the vowel [u], was cut out at zero

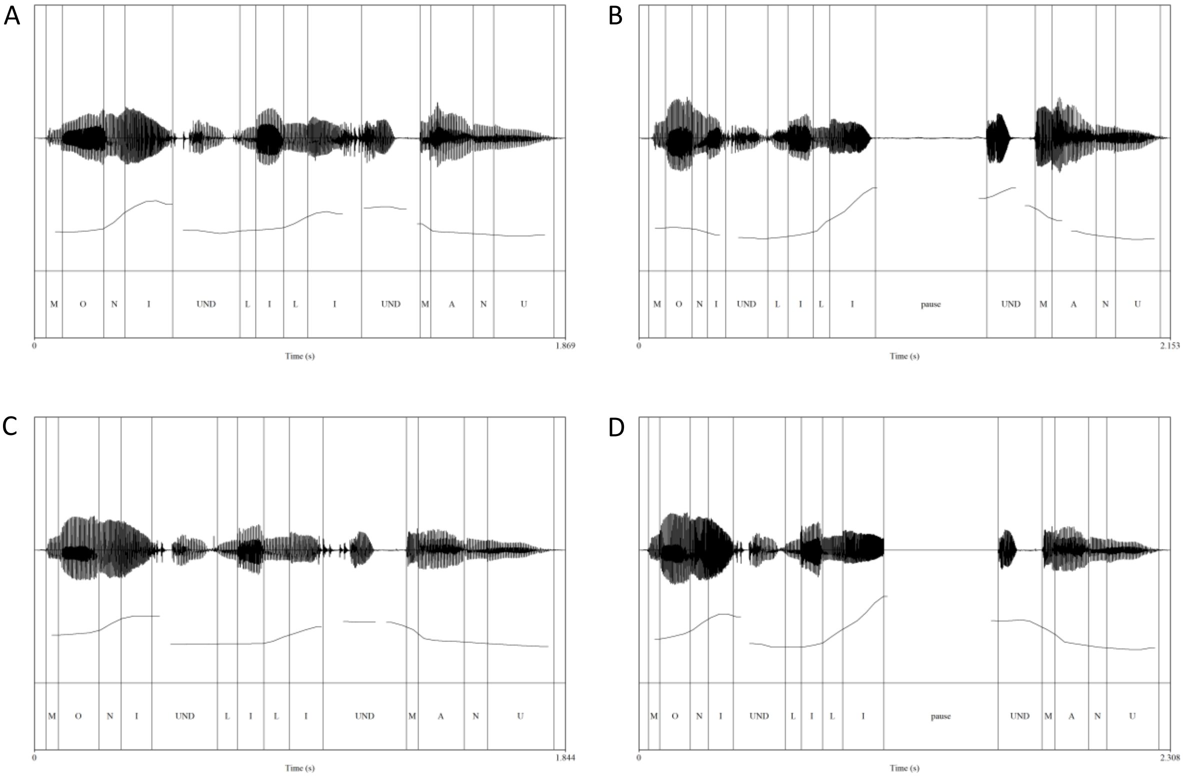

Figure 5.1 Oscillogram and pitch contour aligned to the text. (A) Original sequence without an IPB used in Exp. I/1, (B) Sequence with an IPB used in Exp. I/1, (C) Sequence without an IPB and stylized pitch used in Exp. I/5, (D) Sequence with inserted pitch, lengthening, and pause used in Exp. I/5.

crossings. Then, a silent interval of 500 ms – corresponding to the mean duration of pauses measured in natural sequences with an IPB from Experiment I/1 – was created and inserted at the offset of the final vowel.

The very local implementation of boundary cues led to some differences in the pitch contour between the original sequences with an IPB and sequences with inserted cues (see Fig. 5.1(B) vs. Fig. 5.1(D)). First, the original recordings without an IPB that were the basis for cue manipulation had an accentual peak on the first name that we did not manipulate. However, this accentual peak was always lower than the peak of the H% in the second name (*M* $F0_{MAX}$ = 317 Hz vs. 388 Hz). Second, there were differences in the pitch movement on the postboundary conjunction: the postboundary pitch reset was measured as the difference between the maximum pitch on the final vowel of the second name and the maximum pitch on the vowel of the conjunction. In sequences with inserted cues, this pitch reset was expressed by a pitch fall of 5.4 semitones on average. This was twice as much as the pitch reset in natural sequences with an IPB (Wellmann et al. 2012, Exp. I/1). The difference can be explained by the pitch stylization that was carried out to the global pitch contour in manipulated sequences and led to fewer pitch points in the signal and hence, to a sharper pitch contour compared to the natural sequences with an IPB. Taken together, on the second name the acoustics of the manipulated sequences were similar to the ones of the original sequences with an IPB in Experiment I/1. However, on the first name and on the conjunction the pitch contour differed and resembled that of the natural sequences without an IPB. Hence, the acoustic difference between sequences without an IPB and sequences with inserted cues was slightly less pronounced than the one between the original sequences without an IPB and sequences with an IPB.

For sequences without an IPB, the pitch contour was stylized. No other manipulation was carried out. On the second name, sequences without an IPB, but stylized pitch, had a mean pitch rise of 83 Hz (6.2 semitones) with a high boundary tone H% of 275 Hz on average. There was neither a pause after the second name, nor a relevant pitch reset to the conjunction.

All sequences were resynthesized using the PSOLA function in PRAAT and scaled to 70dB. Examples of sequences with inserted cues and sequences without an IPB, but stylized pitch in contrast to the original sequences used in Experiment I/1 are depicted in Figure 5.1.

For each prosodic condition, the acoustically manipulated sequences were randomly concatenated with a silent interval of 1 s to six differently ordered speech files. The speech files of the condition Without IPB contained seven sequences (i.e., one sequence was randomly repeated) and had an average duration of 19.56 s (range: 19.43-19.67 s). The files of the condition Inserted Cues contained six sequences and had a duration of 19.84 s.

5.1.3 Procedure

During familiarization, one group of infants was presented with sequences without an IPB and another group of infants with sequences with inserted cues. For both groups, familiarization lasted until at least 20 sequences were presented. Given that sequences without an IPB were shorter than sequences with inserted cues, familiarization time slightly differed. That is, when infants were familiarized with sequences without an IPB, familiarization lasted until 56 s of listening time had been accumulated. When familiarized with sequences with inserted cues, infants had to accumulate 66 s of listening time. Subsequently, infants went through a test phase of twelve test trials. In six trials, the sound files without an IPB were presented, in the other six trials the sound files of the sequences with inserted IPB cues. Thus, half of the test trials contained exactly the same sound files that the infants had previously heard during familiarization, whereas the other half consisted of sound files with the type of prosodic grouping that had not been presented during familiarization. The test trials were grouped in three blocks of four trials each (two with and two without an internal IPB in a random order). Additionally, within each block the side of presentation of the sequences of the two prosodic types was counterbalanced so that the prosodic condition and the side of presentation were not associated. The duration of each experimental session depended on the infant's behavior and varied between four and six minutes.

5.2 Results and Discussion

In the overall analysis, infants showed a mean listening time of 6.87 s (*SD* = 2.33 s) to the familiar prosodic grouping, and a listening time of 6.99 s (*SD* = 2.38 s) to the novel prosodic grouping (see Fig. 5.2). A repeated-measures ANOVA with the within-subject factor Familiarity (familiar versus novel prosodic pattern) and the between-subject factor Familiarization Group (without IPB vs. inserted cues) neither showed a significant main effect of Familiarity, *F*(1,23) = .066, *p* = .799, nor of Familiarization Group, *F*(1,23) = .538, *p* = .471, but a significant interaction between the factors Familiarity and Familiarization Group, *F*(1,23) = 4.439, *p* = .046.

Decomposing the significant interaction of Familiarity and Familiarization Group revealed that the group of infants that was familiarized with the sequences without IPB listened on average for 6.18 s (*SD* = 2.27 s) to the familiar prosodic pattern and for 7.09 s (*SD* = 2.55 s) to the novel one. This difference was marginally significant, *t*(12) = 2.048, *p* = .063, two-tailed. The data indicated a trend for a preference for the novel condition Inserted Cues.

The group of infants that was familiarized with sequences with inserted cues, listened for 7.62 s (*SD* = 2.24 s) to the familiar test items and for 6.91 s (*SD* = 2.30 s) to the novel test items. This difference was not significant, *t*(11) = -1.11, *p* = .291, two-tailed. Mean listening times to familiar and novel stimuli by Familiarization Group are depicted in Figure 5.2.

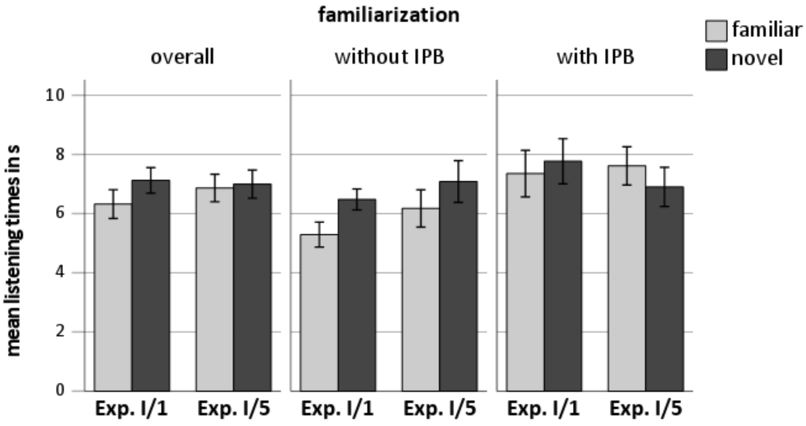

Figure 5.2 Mean listening times to familiar and novel test trials overall for both familiarization groups (left panel) and split by familiarization group without IPB (middle panel) and with IPB (right panel) for Experiment I/1 and Experiment I/5. Error bars indicate 1 SE.

In a next analysis, the data was combined with those from Wellmann et al. (2012) that had used natural stimuli. In an ANOVA with the within-subject factor Familiarity and the between-subject factors Familiarization Group and Experiment (Exp. I/1 vs. Exp. I/5) revealed marginally significant main effects of Familiarity *F*(1,45) = 3.02, *p* = .089 and of Familiarization Group, *F*(1,45) = 3.87, *p* = .055, along with no main effect of Experiment, *F*(1,45) = .143, *p* = .707. The interaction of Familiarity, Familiarization

Group and Experiment was not significant, *F*(1,45) = .668, *p* = .418, nor the interaction of Familiarity and Experiment, *F*(1,45) = 1.841, *p* = .182. However, a significant interaction of Familiarity and Familiarization Group, *F*(1,45) = 5.307, *p* = .026, showed up. This interaction was due to the fact that in the combined analysis the groups of infants familiarized with sequences without an IPB listened significantly longer to novel test items, *t*(24) = 3.203, *p* = .004, two-tailed. In contrast, no significant listening preference occurred in the groups of infants that were familiarized with the sequences with an IPB/ with inserted cues, *t*(23) = -.358, *p* = .723, two-tailed.

Looking more thoroughly at the data of the present experiment, the listening times to the two test conditions Without IPB and Inserted Cues were compared irrespective of their familiarity. This means that across both familiarization groups familiar and novel test trials of Without IPB trials were combined and familiar and novel test trials of Inserted Cues were combined. This revealed that infants tended to listen longer to the condition Inserted Cues (*M* = 7.34 s, *SD* = 2.37 s) compared to the condition Without IPB (*M* = 6.53 s, *SD* = 2.27 s). A repeated-measures ANOVA with the Within-subject factor Test Condition (Without IPB vs. Inserted Cues) and the between-subject factor Familiarization Group revealed a significant main effect of Test Condition, *F*(1,23) = 4.439, *p* = .046, no effect of Familiarization Group, *F*(1,23) = .538, *p* = .471, and no interaction of Condition and Familiarization Group, *F*(1,23) = .066, *p* = .799.

The present experiment aimed at modifying Wellmann et al.'s (2012) findings using stimuli with inserted pitch, lengthening, and pause instead of natural occurring cues. Hereby, it should be ruled out that infants generally dislike and reject stimuli that underwent acoustic manipulation. From the results several conclusions can be drawn.

First, infants showed successful discrimination of sequences without cues and sequences with inserted boundary cues, but only after familiarization with sequences without an IPB. This was indicated by a marginally significant novelty effect. This finding is very similar to the one for the original experiment that used stimuli with naturally cued boundaries. The similarities were confirmed by the combined analysis of both data: performance does not differ across both experiments, but is affected by Familiarization Group: in both experiments, the group that was familiarized with sequences without an IPB significantly preferred to listen to the novel test condition (though only marginally in Experiment I/5[4]), whereas the other group had no significant listening preference. In sum, the results

---

[4] The slightly weaker effect in Experiment I/5 compared to Experiment I/1 may result from acoustic differences between stimuli. The boundary characteristics in sequences with inserted cues were obtained from a local manipulation on the second name of sequences without an IPB. Both test conditions (Without IPB and Inserted Cues) differed only at the critical boundary position. In contrast, the original natural stimuli from Experiment I/1 differed more globally from sequences without an IPB, that is also on the first name and on the postboundary conjunction (see 5.1.2). Thus, the mere acoustic difference between the test conditions was smaller here compared to the original experiment. This may have increased the difficulty to discriminate both conditions.

from the experiment with acoustically manipulated stimuli mirrored those from Experiment I/1 with natural materials.

A more thorough look at the data revealed a difference between both experiments. In the present experiment, infants tended to listen longer to the condition Inserted Cues irrespective of the kind of familiarization, that is, it did not matter whether the test condition was familiar or novel. Hence, infants seem to have a preference for sequences with inserted prosodic boundary cues over sequences without any boundary cues during the testing phase. Since such a general preference was not found for the stimuli with inserted pitch and lengthening cues (Wellmann et al., 2012), this effect can be mainly attributed to the inserted pause cue. Overall, the data does not support the view that infants generally dislike and reject manipulated materials. Even though the manipulation carried out here was even stronger, infants favored the manipulated stimuli. Hence, infants’ shorter listening times to stimuli with inserted pitch and lengthening as found in Experiment I/2 (Wellmann et al., 2012) are not due to a general disliking of manipulated material. Rather, this seems to reflect a true familiarity effect caused by higher task demands when the stimuli only differ in the employment of pitch and lengthening.

A preference for with-boundary material that is independent of the kind of familiarization was not observed in the original experiment with natural materials. Why do infants display a preference for stimuli with inserted cues, but not for stimuli with naturally occurring cues? This may again be caused by acoustic differences. Due to co-articulation the insertion of a silent interval is a rather delicate procedure. Any parts of co-articulation were sought to be deleted before the pause was inserted (since co-articulation does not occur before and after a pause in natural speech). However, this deletion led to cutting points with remaining energy in the oscillogram. Consequently, the acoustics of the ending of the preboundary vowel and the onset of the postboundary phoneme did not reflect fully natural sounding. This may presumably have increased the acoustic salience of the boundary and hence, may have strongly attracted infants’ attention as reflected in longer listening times to manipulated materials irrespective of the type of familiarization. To conclude, acoustic manipulations may have effects on perception – but in the way they were carried out in this series of studies infants do not generally refuse them. Rather, especially, when the acoustically manipulated materials do not involve the pause cue whose insertion is more delicate than the one of lengthening or pitch, they are suitable for testing the impact of each individual boundary cues or combinations.

Study II

6 Developmental changes in prosodic boundary cue perception in German-learning infants[5]

Abstract

Previous investigations suggest that the main prosodic cues characterizing intonation phrase boundaries (IPBs), namely pitch change, final lengthening, and pause, have different weightings in perception. Such a weighting of IPB cues may be subject to crosslinguistic variation and seems to develop during the first year of life (Johnson & Seidl, 2008; Seidl, 2008; Seidl & Cristià, 2008). For German, eight-month-old infants were found to detect an IPB signaled by pitch change combined with final lengthening in a behavioral discrimination task, even in the absence of a pause (Wellmann, Holzgrefe, Truckenbrodt, Wartenburger, & Höhle, 2012).

Assessing the developmental course of prosodic boundary detection, the present study tested six-month-old German-learning infants with the same discrimination task in the *headturn preference paradigm* as used previously with the eight-month-olds (Wellmann et al., 2012). Stimuli were presented in two different prosodic groupings, as a sequence either without or with an internal boundary after the second name, [Moni und Lilli und Manu]$_{IP}$ vs. [Moni und Lilli]$_{IP}$ [und Manu]$_{IP}$. Infants were familiarized to sequences without an IPB and then tested on their discrimination of both prosodic patterns. The boundaries were either cued naturally, that is, by pitch change, final lengthening, and pause (Exp. II/1), or cued by the subset of pitch and lengthening (Exp. II/2 and II/2a) or the subset of pause and lengthening (Exp. II/3). Analyzing infants' listening times across all four experiments revealed a significant interaction of familiarity and experiment. Post-hoc comparisons revealed that six-month-old infants detected the boundary when it was cued by all cues (Exp.II/1) and by pause and lengthening (Exp. II/3). However, when the IPB was only marked by a combination of pitch and lengthening, six-month-old infants failed (Exp. II/2), even when familiarization duration was doubled (Exp. II/2a). This points to a crucial role of the pause cue at six months. Taken together with the results from Wellmann et al. (2012) and Holzgrefe-Lang et al. (2016) our data suggest a development to an adult-like boundary perception that no longer requires the pause cue between six and eight months. We argue that this behavioral change reflects a shift of attention to boundary markings that are functionally relevant in the ambient language.

---

[5] This chapter has been accepted as: Wellmann, C., Holzgrefe-Lang, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (forthcoming) Developmental changes in prosodic boundary cue perception in German-learning infants. In: F. Schuboe, S., Hanne, S., Zerbian, & I. Wartenburger (Eds.) *Prosodic boundary phenomena*. Studies in Laboratory Phonology. Berlin: Language Science Press.
The present version was slightly modified with respect to the numbering of the experiments.

6.1 Introduction

During their first year of life infants pass through a phase of perceptual reorganization, in which their speech perception is sharpened for acoustic properties that are functional in the language they are exposed to, but attenuates for acoustic properties that are not (for a review, see Maurer & Werker, 2014). Perceptual reorganization was initially shown for vowels and consonants, with numerous findings suggesting an increasing ability to discriminate native sound contrasts but decreasing performance with non-native sounds (for vowels: e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994; for consonants: e.g., Best & McRoberts, 2003; Kuhl et al., 2006; Werker & Tees, 1984). More recently, perceptual reorganization has also been shown for suprasegmental prosodic aspects of language like lexical tone (Götz, Yeung, Krasotkina, Schwarzer, & Höhle, 2018; Mattock & Burnham, 2006; Mattock et al., 2008; Yeung et al., 2013) and lexical stress (Bijeljac-Babic, Serres, Höhle, & Nazzi, 2012; Höhle et al., 2009; Jusczyk, Cutler, & Redanz, 1993; Skoruppa et al., 2009). The present paper deals with a potential developmental change in a further area of prosody, namely in the perception of acoustic cues that mark major prosodic boundaries, specifically boundaries at the edges of *Intonation Phrases* (IP, (Nespor & Vogel, 1986; Selkirk, 1986). Three main cues are considered to mark these boundaries across languages: a change in fundamental frequency (F0) in terms of boundary tones or a pitch reset, a lengthening of preboundary segments, and the insertion of a pause (e.g., Hirst & Di Cristo, 1998; Nespor & Vogel, 1986; Price et al., 1991; Vaissière, 1983).

The edges of IPBs usually coincide with syntactic clause boundaries (Selkirk, 2005; for German: Truckenbrodt, 2005), and infants as well as adults benefit from this close syntax-prosody mapping. Over the past thirty years, multiple studies have provided consistent evidence that infants are highly sensitive to prosodic boundary information and use it to segment the continuous speech signal into linguistically relevant units (Gout, Christophe, & Morgan, 2004; Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989; Nazzi et al., 2000; Schmitz, 2008; Shukla, White, & Aslin, 2011). Assuming that each prosodic boundary cue contributes individually to boundary perception, previous research has focused on the specific roles of pitch change, final lengthening, and pause in adult as well as infant listeners (e.g., Aasland & Baum, 2003; Johnson & Seidl, 2008; Lehiste et al., 1976; Petrone et al., 2017; Sanderman & Collier, 1997; Scott, 1982; Seidl, 2007; Streeter, 1978; Zhang, 2012).

Studies with American-English-learning infants found evidence of a developmental change in perception of these boundary cues (Seidl, 2007; Seidl & Cristià, 2008). In a series of experiments with varying cue manipulations, Seidl (2007) tested six-month-old infants with an experimental design based on Nazzi et al. (2000). In a familiarization phase, infants were presented with two sequences of the same words extracted from two different naturally recorded text passages of infant-directed speech. One familiarization sequence was a complete clause (*clause* stimuli, e.g., “# Leafy vegetables taste so good #”). The other one contained an internal clause boundary and thus a major prosodic boundary (*non-clause* stimuli, e.g., “leafy vegetables # taste so good”). During the test phase, the original text passages from which the familiarization sequences had been extracted were presented. In the first experiment, the internal boundary of the non-clausal familiarization sequence was fully marked by all three boundary cues, that is, a pitch reset at the juncture, final lengthening, and a pause. Results indicated that infants were better able to recognize the clausal than the non-clausal sequence in the continuous speech presented during the test phase. In subsequent experiments, the familiarization sequences were varied by acoustic manipulation of one or two of the three essential prosodic cues to find out if clause segmentation would still be possible. When either pause or final lengthening was neutralized, infants still preferred the passage with the familiar clausal sequence. Consequently, neither of these two cues was necessary to evoke the detection of a prosodic boundary. However, when the pitch cue was neutralized, infants no longer showed a preference for the clausal sequence. Yet pitch alone was not a sufficient marker: when it was present as a single cue with neutralized pause and final lengthening, clause segmentation was not successful. Hence, Seidl (2007) concluded that American-English-learning six-month-old infants need the pitch cue in combination either with pause or with final lengthening for clause segmentation.

Seidl and Cristià (2008) extended these investigations to four-month-old American-English-learning infants. Tested with the same experimental material and design, the younger group was only successful in clause segmentation when all three cues in combination signaled the boundary. The authors concluded that four-month-old infants’ clause segmentation is based on a holistic processing that relies on a coalition of all boundary cues while six-month-old American-English-learning infants are already able to process the cues independently of each other.

Crosslinguistic evidence provides first hints that boundary perception by six-month-olds is modulated by specific prosodic properties of the language that the infants are learning. Johnson and Seidl (2008) used the same experimental procedure as Seidl (2007) to test Dutch-learning six-month-olds with Dutch materials. In contrast to their American English age mates, the Dutch infants only showed evidence of clause segmentation when the prosodic boundaries were marked by the combination of pitch, lengthening, and a pause, but not when the pause was removed from the materials. The authors argue that pauses may be a strong marker of prosodic boundaries in Dutch and thus important for infants’ boundary detection. In fact, a comparison of the speech materials that were naturally recorded for the American English and the Dutch experiments showed that in the Dutch passages the pauses at the clause boundaries were twice as long as in the American English ones while the American English passages showed a much larger pitch reset after the juncture than the Dutch materials.

Broadening the crosslinguistic perspective, Wellmann et al. (2012) investigated German-learning infants on their perception of boundary cues in bracketed lists of names as *[A and B and C]* in contrast to *[A and B] [and C]* indicating either one group of three people or a groups of two and a single person.

According to Petrone et al. (2017), German speakers typically employ the IP to signal such a grouping (see also Huttenlauch, Hansen, de Beer, Hanne & Wartenburger, this volume). Indeed, analyzing the respective cues at the natural internal boundary in Wellmann et al.'s materials revealed typical characteristics of an IPB: pitch changes, specifically an upstep on the second peak and a partial pitch reset, a lengthening of the preboundary vowel, and the employment of a pause (for similar findings concerning prosodic boundary cues in German, see also Féry & Kentner, 2010; Truckenbrodt, 2007a, 2016). Hence, the sequences either formed a single IP, [Moni und Lilli und Manu]$_{\text{IP}}$, or were made up of two IPs with an internal IPB after the second name, [Moni und Lilli]$_{\text{IP}}$ [und Manu]$_{\text{IP}}$. The internal IPB was the focus of our investigations on prosodic cue perception.

Unlike previous studies (Johnson & Seidl, 2008; Nazzi et al., 2000; Seidl, 2007), Wellmann et al. (2012) and the present study tested the detection of the prosodic boundary not by a clause segmentation task, but by a discrimination task. In this discrimination task, two groups of eight-month-old infants were familiarized with a sequence of one prosodic type, either with or without an internal IPB. In the subsequent test phase, all infants were presented with sequences of both prosodic types to test whether they discriminated between both of them. Given that previous research on infants' attunement to features of segmental phonology also used discrimination tasks (Werker & Tees, 1984; Polka & Werker, 1994; Mattock & Burnham, 2006; Mattock et al., 2008), we assumed that such a discrimination task should be suited to reveal differences in prosodic boundary information as well. We considered the methodology as an important contribution since the same materials were suitable for use in a behavioral study with adults, as well as for use in ERP studies with adults and infants (Holzgrefe-Lang, Wellmann, Höhle, & Wartenburger, 2018; Holzgrefe-Lang et al., 2016). Moreover, the material can in principle be used in other languages as well (for French: van Ommen et al., 2020).

Wellmann et al.'s (2012) results revealed that eight-month-old infants preferred to listen to sequences of the new prosodic grouping (i.e., sequences with an IPB) after familiarization with sequences without an IPB. This indicated successful discrimination of the prosodic patterns. In subsequent experiments the prosodic boundary information was systematically varied by adding a cue or a subset of cues to the original sequence without an internal IPB after the second name. When the IPB was signaled by a pitch rise and final lengthening in combination but without a pause, eight-month-olds still successfully discriminated the sequences with boundary cues from the sequences without an internal IPB. However, when the IPB was signaled solely by a pitch rise or by final lengthening, infants did not discriminate the two prosodic conditions. These findings suggest that pitch change and lengthening in combination, but not as single cues, are sufficient for IPB detection in eight-month-old German-learning infants while the presence of a pause is not necessary.

Interestingly, the discrimination pattern of the German eight-month-olds mirrored a pattern that Holzgrefe-Lang et al. (2016) observed in a prosodic judgment task using the same stimuli with German speaking adults. In this task, participants judged via button-press whether the stimuli contained an internal boundary or not. The results revealed that stimuli containing a pitch change and final lengthening in combination but no pause were judged as sequences with an IPB. In contrast, stimuli that contained only a pitch change were predominantly judged as sequences without an internal boundary, and sequences with only lengthening were judged at chance level, indicating no categorization.

Although the two tasks – the discrimination task with infants and the prosodic judgment task with adults – may place some different requirements on the participants and each group's data was analyzed on its own, the similarity in the results across the two studies indicates that German-learning eight-month-olds' sensitivity to prosodic boundary cues already resembles that of German adults. The question arises whether the discrimination pattern found in the German eight-month-olds is in fact the result of a perceptual attunement from a solely acoustically driven perception based on the presence of the salient pause cue to a more sophisticated linguistically affected perception relying on pitch change and lengthening.

Therefore, six-month-old German-learning infants were studied with the same experimental paradigm and the same stimuli as used by Wellmann et al. (2012), that is, sequences without an internal IPB and sequences with an internal boundary cued by the full set or a subset of pitch, lengthening, and pause cues had to be discriminated.

In Experiment II/1, detection of a prosodic boundary that is fully marked by the combination of all naturally occurring cues was investigated. We assumed that infants are able to detect this boundary, as previous research presenting infants with artificial pauses at boundary and non-boundary locations has revealed that German infants are highly sensitive to the correlation of prosodic boundary information already in their first half year of life (Schmitz, 2008).

Experiment II/2 examined whether pitch change and final lengthening in combination are sufficient boundary markers or whether pause is a necessary cue for this age group. Here we aimed to clarify whether prosodic boundary detection at six months already reflects an attunement to linguistically relevant markings or whether it is rather influenced by the perceptual salience of cues.

In Experiment II/2a we used the same stimuli as in Experiment II/2, but a prolonged familiarization. We hypothesized that infants develop a sensitivity to boundaries that are not cued by pause between six and eight months. In Experiment II/2a, we asked whether this sensitivity would show up already at six months under optimized experimental conditions, that is, after a prolonged exposure. We hypothesized that the double amount of presentations of the familiarization sequence might lead to a more stable mental representation and would thus release (working) memory capacity to thoroughly explore the new stimulus with its differences.

In Experiment II/3, we investigated whether the combination of pause and final lengthening provides sufficient information for boundary detection or whether – as has been shown for younger American-English-learning infants – only a combination of all cues would evoke boundary detection.

In the following, we will successively introduce each experiment with its participants, stimuli, procedure, and its descriptive results. Subsequently, a statistical analysis across all four experiments will be reported, followed by a general discussion.

6.2 Experiment II/1: The influence of pitch, final lengthening, and pause on six-month-olds' boundary detection

Experiment II/1 tested whether German-learning six-month-old infants are able to perceive an IPB that is signaled by the three main prosodic cues pitch change, final lengthening, and pause.

6.2.1 Participants

A group of twenty-four six-month-old infants (12 girls) was tested. Their mean age was 6 months, 11 days (range: 6 months, 2 days to 6 months, 27 days). Nine additional infants were tested but not included in the data for the following reasons: failure to complete the experiment (1), crying or fussiness (4), mean listening times of less than 3 seconds per condition (1), technical problems (2), and noise in the surroundings due to construction work (1).

All infants who participated in this and the following experiments were from monolingual German-speaking families, born full-term, and with normal hearing. They were recruited from birth lists obtained through the Potsdam city hall archives. All parents signed informed consent. None of the infants tested in the present study participated in more than one experiment.

6.2.2 Stimuli

All stimuli used in Experiments II/1 and II/2 were identical to those that were presented to eight-month-olds in the study by Wellmann et al. (2012): the stimuli consisted of a sequence of three German names containing only sonorant sounds ("Moni, Lilli, Manu"), which allowed a reliable measure of F0 and were suitable for acoustic manipulation. The names were coordinated by und ('and'). A young female adult, a German native speaker from the Brandenburg area, was instructed to read the sequence in two different prosodic groupings indicated by different bracketing:

a. (Moni und Lilli und Manu)$_{IP}$ – without internal IPB

b. (Moni und Lilli)$_{IP}$ (und Manu)$_{IP}$ – with internal IPB

Both sequences contained the same string of names and differed only in grouping either all three names together as shown in (a) or grouping the first two names together and the final one apart as shown in (b). Sequences of type (a) were produced as a single IP, without an internal boundary. In contrast, sequences of type (b) consisted of two IPs, with an internal IPB after the second name. The speaker repeated each sequence six times, resulting in six recordings per prosodic type. The intended grouping was confirmed by two independent listeners who were naïve with respect to the given bracketing. Recordings were made in an anechoic chamber equipped with an Audio-Technica AT4033A studio microphone, using a C-Media Wave soundcard at a sampling rate of 22050 Hz with 16-bit resolution. Examples of both kinds of prosodic phrasing are depicted in Figures 6.1(A) and 6.1(B).

The acoustic analysis of the recordings revealed clear acoustic differences between the two prosodic phrasings on and after the second name (see Tab. 6.1).

Preboundary pitch movement: Sequences without an internal IPB that form one single IP were characterized by F0 lowering, with an accentual pitch rise on the first name, followed by a smaller pitch rise on the second name, that is, a downstep pattern (Truckenbrodt, 2007b). Sequences with an internal IPB exhibited a flat tonal contour (plateau) on the first name, followed by a large pitch rise on the second name, starting at the second syllable, and leading to an upstepped peak, a high boundary tone, at the final vowel. This pitch rise on the second name (measured as the difference between the maximum and minimum pitch on the second name in semitones) was 2.5 times greater and led to a

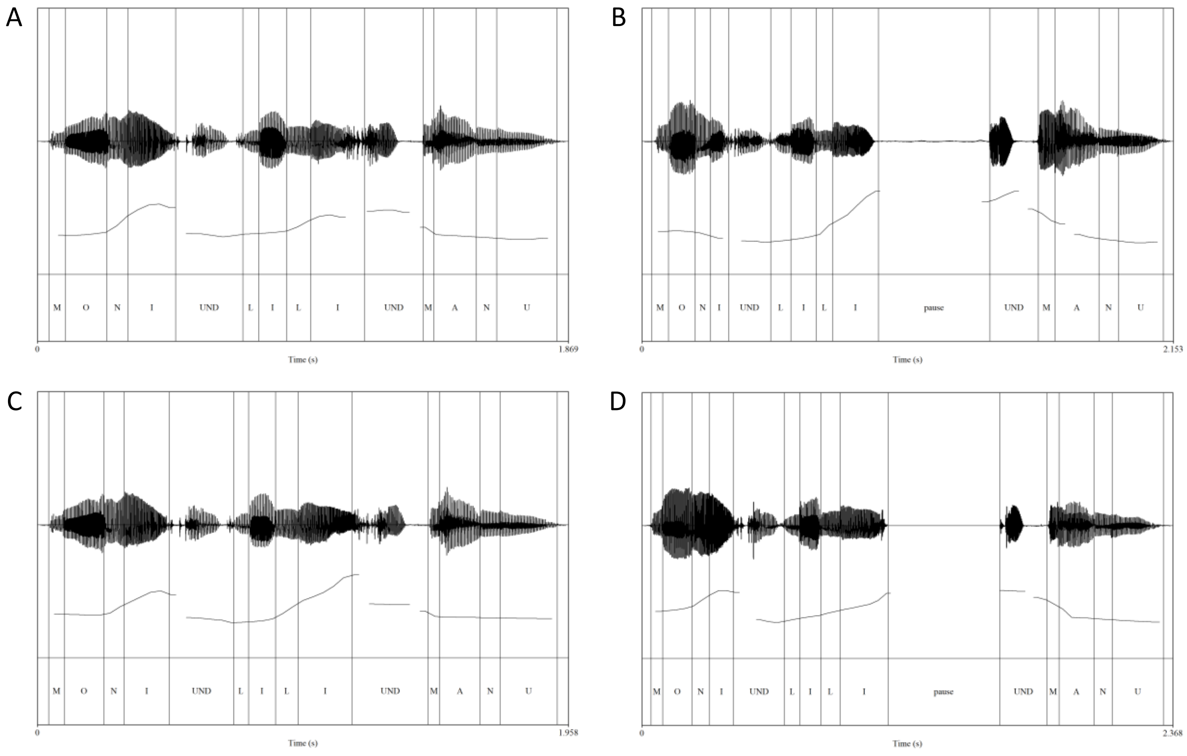

Figure 6.1 Oscillograms and pitch contours aligned to the text. Vertical lines mark the segmental boundaries. (A) Sequence without an IPB used in Exp. II/1, (B) Sequence with a fully marked internal IPB used in Exp. II/1, (C) Sequence with pitch change and final lengthening used in Exp. II/2 and II/2a, (D) Sequence with pause and final lengthening used in Exp. II/3.

Table 6.1 Mean values and range of the acoustic correlates of the prosodic boundary in the six experimental sequences *Moni und Lilli und Manu* without and with an internal IPB, respectively.

boundary cue	**acoustic correlate**	**without internal IPB** [Moni und Lilli und Manu]	**with internal IPB** [Moni und Lilli] [und Manu]
boundary tone H%	max F0 in Hz on final vowel in name 2	277 (264 to 293)	397 (371 to 422)
pitch rise	difference between the maximum and the minimum F0 in name 2 in …		
	… Hz	88 (77 to 110)	220 (197 to 240)
	... semitones	6.7 (5.8 to 8.2)	14.0 (12.8 to 14.6)
partial pitch reset	difference between the maximum F0 on the final vowel in name 2 and the maximum F0 on the vowel of the following conjunction in …		
	… Hz	5x rise: -12 (-23 to -5); 1x fall: 18	6x fall: 55 (36 to 96)
	… semitones	5x rise: -0.7 (-1.3 to -0.3); 1x fall: 1.1	6x fall: 2.5 (1.7 to 4.4)
final lengthening	duration of the final vowel in name 2 in ms	99 (91 to 110)	175 (162 to 186)
pause	duration of the pause after name 2 in ms	0	506 (452 to 556)

higher maximum pitch than the small rise occurring at the same location in sequences without an IPB. Hence, the F0 contour of sequences with an internal IPB clearly indicated the following prosodic boundary. This tonal contour resembled the most common realization of internal IPBs in similar German sequences of names investigated in Petrone et al. (2017, see LATE condition).

Postboundary pitch movement: Postboundary pitch reset was measured as the difference between the maximum pitch on the final vowel of the second name and the maximum pitch on the vowel of the conjunction. In sequences without an internal IPB there was no relevant pitch difference: in five of the six recordings the height of the downstepped second peak was slightly higher at the conjunction (on average by 0.7 semitones), whereas in one recording it was slightly lower (by 1.1 semitones). In sequences with an internal IPB a partial pitch reset, one step below the preboundary upstep, occurred (see Truckenbrodt, 2007b for a similar partial reset). This was expressed by a pitch fall of 2.5 semitones on average.

Final lengthening: To explore final lengthening, the duration of the second name's final vowel [i] was compared in sequences with and without an internal IPB. Its duration was 1.8 times longer in the grouping with an IPB, indicating a strong lengthening cue (cf. Kohler, 1983).

Pause duration: Finally, the duration of the silent interval after the second name was measured. A pause with an average duration of 506 ms occurred in sequences with an internal IPB, whereas no pause was present in sequences without an internal IPB (see Tab. 6.1).

Taken together, in sequences with an internal IPB all acoustic correlates of the three main prosodic boundary cues were observed: a change in F0 (mainly, a preboundary pitch rise), a lengthening of the preboundary vowel, and the occurrence of a pause.

The recorded sequences were used to create sound files for presentation during the experiment. All recordings were scaled to a mean intensity of 70 dB. For each prosodic type, the six recordings were randomly concatenated with a silent interval of 1 s inserted between them. In this way, six sound files per prosodic grouping were created such that each file consisted of a different order.

Due to the missing durational cues (final lengthening and pause), sequences without an internal IPB were shorter than those with an internal IPB. The average duration of sequences without an internal IPB was 1.76 s (range: 1.67–1.87 s), while it was 2.16 s (range: 2.13–2.2 s) for sequences with an internal IPB. To match the sound files of the two prosodic types with respect to overall length, the number of sequences within each file was varied. As a result, sound files of condition with an internal IPB contained six sequences and had an average duration of 18.97 s, and sound files of the condition without an internal IPB contained seven sequences (i.e., one random recording was repeated), leading to an average duration of 19.32 s (range: 19.16–19.43 s). The difference in the number of sequences was crucial in order to present sound files of similar lengths during the experimental trials.

6.2.3 Procedure

In all experiments presented here, infants were tested using the *headturn preference procedure (HPP)* including a familiarization phase. During the experiment, the infant was seated on the lap of a caregiver in the center of a test booth. Inside this booth three lamps were fixed: a green one on the center wall, and a red one on each of the side walls. Directly above the green lamp was an opening for the lens of a video camera. Behind each of the red lights a JBL Control One loudspeaker was mounted. Each experimental trial started with the blinking of the green center lamp. When the infant oriented to the green lamp, it was turned off and one of the red lamps on a side wall started to blink. When the infant turned her head towards the red lamp, the speech stimulus was started, delivered via a Sony TA-F261R audio amplifier to the loudspeaker on the same side. The trial ended when the infant turned her head away for more than 2 s, or when the end of the speech file was reached. If the infant turned away for less than 2 s, the presentation of the speech file continued but the time spent looking away was not included in the total listening time. The whole session was digitally videotaped. The experimenter's coding was recorded and served for the calculation of the duration of the infant's head turns during the experimental trials. The caregiver listened to music over headphones to prevent influences on the infant's behavior. Furthermore, she was instructed not to interfere with the infant's behavior during the experiment. The experimenter sat in an adjacent room, where she observed the infant's behavior on a mute video monitor and controlled the presentation of the visual and the acoustic signals by a button box. The experimenter was blind with respect to the type of acoustic stimuli presented during familiarization and testing.

An experimental session consisted of a familiarization phase immediately followed by a test phase. For Experiment II/1, we familiarized half of the infants with sequences without an internal IPB, while the other half listened to sequences with an internal IPB. For both groups, familiarization lasted until at least 20 sequences were presented. Given that sequences without an IPB were shorter than sequences with an internal IPB, familiarization timing differed slightly. That is, when infants were familiarized with sequences without an internal IPB, familiarization lasted until 55 s of listening time had been accumulated. When familiarized with sequences with an internal IPB, infants had to accumulate 63 s of listening time.

After familiarization, infants immediately passed through the test phase that comprised twelve trials. Half of the test trials contained the identical sound files previously presented during familiarization (familiar test trials). The other six trials contained the sound files of the other prosodic grouping (novel test trials). The test trials were grouped in three blocks of four trials each. Two out of these four trials contained sequences with an internal IPB, the others contained sequences without an internal IPB. Within each block, test trials were randomly ordered with the side of presentation being counterbalanced for each prosodic type. Based on the infant’s head turns the listening time to each test trial was measured. The duration of each experimental session varied between four and six minutes, depending on the infant’s behavior.

6.2.4 Descriptive results

We analyzed the data for each familiarization group on a descriptive level. Within this sample we observed clear numerical differences in the listening times. The group of infants that was familiarized with sequences with an internal IPB listened for 7.69 s (*SD* = 3.41 s) to novel test trials and for 7.52 s (*SD* = 2.90 s) to familiar test trials, that is, the mean listening time between novel and familiar test trials only differed by M_{Diff} = .17 s. Six out of twelve infants listened longer to the novel test trials.

The group of infants that was familiarized with sequences without an internal IPB listened on average for 9.65 s (*SD* = 2.96 s) to novel test trials and for 8.06 s (*SD* = 3.02 s) to the familiar ones (see Fig. 6.2). Hence, the familiarization with sequences without an internal IPB yielded a novelty preference with a mean listening time difference of M_{Diff} = 1.58 s. Nine out of twelve infants listened longer to the novel test trials.

Overall, a clearer numerical difference between the two prosodic patterns showed up after familiarization with sequences without internal IPB, whereas the familiarization with sequences with an internal IPB seemed to be much less or even not effective as also evidenced in other studies (see Van Ommen et al., 2020 and Wellmann et al., 2012 for results and a discussion of this asymmetric behavior). Given the constraints we usually encounter in infant research (small sample sizes, high drop-out rates related to infant behavior) we therefore decided to run only the familiarization with sequences without an internal IPB in the subsequent experiments of the present study.[6]

6.3 Experiment II/2: The influence of pitch and final lengthening on six-month-olds' boundary detection

Experiment II/2 examined whether a subset of prosodic cues would suffice to trigger the perception of a boundary in six-month-old infants. Specifically, the impact of the combination of a rising pitch contour and final lengthening was under focus, questioning the necessity of the pause cue.

6.3.1 Participants

Sixteen infants (8 girls) were tested. The mean age was 6 months, 14 days (range: 5 months, 28 days to 6 months, 29 days). Four additional infants were tested but not included in the data analysis for the following reasons: failure to complete the experiment (1), crying or fussiness (1), and mean listening times of less than 3 seconds per condition (2).

6.3.2 Stimuli

Experiment II/2 involved stimuli of two prosodic types: one condition comprised the same six sequences without an internal IPB as in Experiment II/1. The other one consisted of six sequences with only pitch rise and lengthening cues indicating the boundary. For this prosodic type, sequences without an internal IPB were locally acoustically manipulated with respect to F0 on the second name and the duration of its final vowel. A specific pitch reset cue, that is, a manipulation of F0 at the position of the postboundary conjunction, was not implemented, since in the stimuli of Experiment II/1 the postboundary peak was utterance-final and generally low. Stimuli manipulations were carried out the same way as in Wellmann et al. (2012).

The stimuli without a boundary were selected as the basis for the acoustic manipulations to avoid a potential influence of additional cues that may contribute to IPB marking and perception. Thus, the crucial boundary information, here a rising pitch contour and final lengthening, was added to the sequences without an internal IPB. Hence, experimental effects can clearly be attributed to the acoustic properties under investigation. By using these local cue manipulations the stimuli with IPB

---

[6] With the same Experiment II/1, we also tested 24 four-month-old infants (n=12 in each familiarization group) in a slightly modified HPP setup (to adapt for the limited head movements at that age the position of the side lamps was moved to the edges of the front wall). Four-month-old infants that were familiarized to sequences without an IPB had mean listening times of 10.5 s (*SD* = 3.59 s) to novel test trials and 10.25 s (*SD* = 3.26 s) to the familiar ones, *t*(11) = .483, *p* = .639, two-tailed. The group familiarized to sequences with an IPB listened on average 10.01 s (*SD* = 4.12 s) to the novel test condition and 10.67 s (*SD* = 4.90 s) to the familiar one, *t*(11) = -1.1, *p* = .295, two-tailed. Given this null result we did not continue in testing four-month-olds.

cues differed from sequences that were used in the condition without an internal IPB only within a predefined critical region and by controlled acoustic properties. However, this local manipulation led to the concession that sequences with inserted pitch rise and lengthening differed from natural sequences with an internal IPB with respect to the pitch contour of the first name. The original recordings without an internal IPB had an accentual peak on the first name that is, a pitch cue to a phonological phrase boundary (see Fig. 6.1(A)). The accentual peak on the first name was always lower than the peak of the H% in the second name (M $F0_{MAX}$ = 317 Hz vs. 388 Hz). However, this kind of cue was not present in the naturally produced sequences with an internal IPB, but was preserved in sequences with inserted cues since sequences without an internal IPB were the base for sequences with inserted cues and cue manipulations were restricted to the second name. Hence, the inserted pitch rise was preceded by another smaller pitch cue. If the pitch rise on the second name is parsed globally, that is, in relation to the previous pitch contour, the pitch rise cue in the manipulated sequences is less pronounced and potentially less salient than the pitch rise cue in natural sequences with an internal IPB.

The acoustic manipulation was carried out with PRAAT software (Boersma & Weenink, 2010). As the phonetic magnitude of prosodic cues differs across languages, and also within a language depending on the syntactic structure (e.g., for pausing in German, see Butcher, 1981), there is no unique value for each prosodic cue. Hence, we decided to implement the same phonetic magnitude for each cue that was present in the corresponding naturally produced stimulus with an internal IPB, proceeding in the same way as other studies that have employed cue manipulations (e.g., Seidl, 2007).

Manipulation steps were the following: to implement the pitch rise, first, the pitch contour of the sequences without an internal IPB was stylized (two semitones). This transformation decreases pitch perturbations by reducing the number of pitch points. Second, the pitch points on the second name were set to the reference values. The reference values of F0 were measured on the second name in the six original sequences with an internal IPB (used in Experiment II/1), namely at the midpoints of the four segments [l], [ɪ], [l], [i] and at the position of the maximum pitch present on the final vowel. For the manipulation of the pitch contour, pitch points with the mean values at these time points (176 Hz, 183 Hz, 224 Hz, 305 Hz, and 397 Hz) were inserted into the stylized sequences without an internal IPB at the same positions. After PSOLA resynthesis in PRAAT, the six new stimuli contained a natural sounding pitch rise of 212 Hz (13.65 semitones) leading to an H% with a mean value of 388 Hz. To implement final lengthening, the final vowel [i] of the second name was lengthened to 180 %. This factor was chosen because in the natural stimuli, the crucial vowel was on average 1.8 times longer in sequences with an internal IPB than in sequences without an internal IPB (see Tab. 6.1). A sequence with manipulated pitch and lengthening is depicted in Figure 6.1(C).

To avoid comparing natural with acoustically manipulated material, we carried out a slight acoustic manipulation in sequences without an internal IPB as well, that is, the stylization of the pitch contour (two semitones). After pitch stylization, sequences were resynthesized using the PSOLA function. Sequences without an internal IPB lasted on average 1.76 s (range: 1.67–1.87 s), while sequences with inserted pitch and lengthening had a mean duration of 1.84 s (range: 1.74–1.96 s). From these sequences six differently ordered sound files per prosodic type were created to be used as experimental trials. The interstimulus interval between the sequences and within a sound file was 1 s. All sound files contained seven sequences (one random recording was repeated). The files containing sequences without an internal IPB had an average duration of 18.33 s (range: 18.23–18.43 s) and the files containing sequences with inserted pitch and lengthening cues lasted on average 18.81 s (range: 18.79–19.01 s).

6.3.3 Procedure

All infants were familiarized with sequences without an internal IPB. The familiarization lasted until at least 20 sequences had been presented, resulting in a minimum of 52 s of accumulated listening time. The familiarization was immediately followed by a test phase with twelve trials. As in Experiment II/1, half of the test trials contained the same sound files that the infants had heard during familiarization. The other half contained the files of the sequences with pitch and lengthening cues, which had not been presented during familiarization. All twelve test trials were grouped in three blocks of four trials each (two of each prosodic type in a random order). The infant's listening time to each test trial was measured.

6.3.4 Descriptive results

Infants tested in Experiment II/2 showed a mean listening time of 7.09 s (*SD* = 2.1 s) to the novel test trials and a mean listening time of 7.20 s (*SD* = 2.43 s) to the familiar test trials (see Fig. 6.2). Eight out of 16 infants had longer listening times to the familiar test trials.

6.4 Experiment II/2a: The influence of pitch and final lengthening on six-month-olds' boundary detection after prolonged familiarization

To verify that the non-discrimination in Experiment II/2 was due to the composition of the stimuli, we modified the experimental design by doubling the familiarization time. Considering a longer familiarization to enable successful discrimination stems from findings of studies that tested French-learning infants' discrimination of rhythmic patterns (Bijeljac-Babic et al., 2012; Höhle et al., 2009; Skoruppa et al., 2009). For French, a language without contrastive stress at the word level, the perception of prosodic cues indicating lexical stress has been shown to be hard for infant learners and adult listeners (Bhatara, Boll-Avetisyan, Unger, Nazzi, & Höhle, 2013; Höhle et al., 2009). In the study by Bijeljac-Babic et al. (2012), monolingual ten-month-old French-learning infants exhibited a null result in discriminating an iambic and a trochaic version of a pseudo-word after a one-minute familiarization. However, when familiarization duration was increased to two minutes, they were successful at discriminating the stress patterns as indicated by a novelty effect. Bijeljac et al. concluded that the null result after the short familiarization could not be interpreted as a general inability to distinguish both stress patterns, but was due to the short familiarization. Regarding the novelty effect after long familiarization, they drew on the model by Hunter & Ames (1988) that would predict novelty preferences in relatively easy discrimination conditions.

In our case of discriminating lists of names with and without IPB cues, it is important to consider that boundary perception without the pause cue is successful in older infants. This sensitivity seems to arise between six and eight months. If six-month-old German-learning infants are already at the beginning of this development, in the light of the Hunter and Ames model, discrimination might show up with reduced task difficulty. Following Bijeljac et al. (2012) we hypothesized that a more robust mental representation of the stimuli presented during familiarization may improve the ability to detect differences between the familiar and the novel stimuli. In the following experiment, we therefore doubled the amount of presentations of the familiarization stimulus in order to help infants building up a more robust mental representation of the sequences without an internal IPB. Through this modification, infants might be able to accomplish a still difficult task for their age such as the detection of a boundary signaled only by pitch and lengthening cues.

6.4.1 Participants

Twenty-three six-month-old infants (12 girls) were tested. The mean age was 6 months, 15 days (range: 6 months, 0 days to 6 months, 26 days). All infants were from monolingual German-speaking families, born full-term, and normal-hearing. Thirty-one additional infants were tested but their data were not included in the analysis for the following reasons: failure to complete the experiment (6), crying or fussiness (15), mean listening times of less than 3 seconds per condition (4), technical problems (2), experimenter error (2), parental interference (1), and outlying listening times due to steady fixation (1).

Drop-out rate was especially high, primarily due to infants' fussiness and failure to finish the experiment (accounting for 68 % of all drop-outs). The longer lasting familiarization which increased the total duration of the experiment to about 6 to 10 minutes (in contrast to six minutes with the original familiarization duration) may have reduced infants' attention.

6.4.2 Stimuli

Stimuli were exactly the same as in Experiment II/2.

6.4.3 Procedure

Infants were familiarized with sequences without an internal IPB. The familiarization duration was set to 104 s. After familiarization, infants listened to exactly the same twelve test trials as used in Experiment II/2, half of them being sequences with inserted pitch and lengthening cues, the other half sequences without an internal IPB.

6.4.4 Descriptive results

Infants tested in Experiment II/2a showed a mean listening time of 6.86 s (*SD* = 2.01 s) to the novel test trials, and a listening time of 7.52 s (*SD* = 2.25 s) to the familiar test trials (see Fig. 6.2). Fifteen out of 23 infants listened longer to familiar test trials.

6.5 Experiment II/3: The influence of pause and final lengthening on six-month-olds' boundary detection

In the following Experiment II/3 the boundary was cued by a pause in combination with final lengthening, but without any pitch cue. The aim of this experiment was to investigate whether six-month-olds would respond to a boundary that is cued by a subset of the naturally occurring cues including a pause. The combination of pause and lengthening was chosen because a pause rarely occurs as the only cue in German (only at 1.3 % of all boundaries in the analysis by Peters, Kohler, & Wesener, 2005), and would thus sound unnatural as the only inserted cue[7], and because the combination of pause with final lengthening occurs more frequently (8.4 %) in spoken German than the combination of pause and pitch (4.9 %, values by Peters et al., 2005). Moreover, this combination is interesting to look at crosslinguistically, as six-month-old American-English-learning infants failed to perceive a boundary signaled only by the combination of pause and lengthening cues (Seidl, 2007).

6.5.1 Participants

Sixteen infants (8 girls) were tested. The mean age was 6 months, 10 days (range: 5 months, 14 days to 6 months, 28 days). Eleven additional infants were tested but not included in the data for the

---

[7] We are grateful to on reviewer who raised the question whether discrimination would be possible with pause as the only boundary cue. We hypothesize that a similar experiment with a boundary cued by pause only would lead to successful discrimination as well. This hypothesis is based on infants' successful discrimination between clauses with artificially inserted pauses at non-boundary locations and clauses with pauses at natural boundary positions with co-occurring boundary cues (Schmitz, 2008).

following reasons: failure to complete the experiment (1), crying or fussiness (6), mean listening times of less than 3 seconds per condition (2), and technical problems (2).

6.5.2 Stimuli

In Experiment II/3, we contrasted sequences without an internal IPB and sequences that contained a pause and final lengthening. To create the latter, five recordings of sequences without an internal IPB were acoustically manipulated on and after the second name. We did not use exactly the same set of recordings without an internal IPB as in Experiments II/1 and II/2 because some of the sequences contained co-articulation between the final vowel of the second name and the initial vowel of the conjunction such that the insertion of a pause would have created an unnaturally sounding stimulus. Hence, for Experiment II/3, we chose five sequences with no or only minimal co-articulation: three sequences that had been used in the previous experiments and two more sequences recorded with the same speaker. First, any co-articulation between the second name and the subsequent conjunction, that is, the section of formant transition from the final vowel [i] to the vowel [u], was cut out at zero crossings. Second, a silent interval of 500 ms – corresponding to the mean duration of pauses measured in natural sequences with an internal IPB from Experiment II/1 – was inserted at the offset of the final vowel. Then, the final vowel was lengthened to 180 %, according to the average lengthening factor found in the acoustic analysis of sequences with an internal IPB in Experiment II/1. A sequence with inserted pause and lengthening cues is depicted in Figure 6.1(D). For both stimulus conditions, the pitch contours were stylized (two semitones) and sequences were resynthesized using the PSOLA function in PRAAT.

Sequences without an internal IPB lasted on average 1.82 s (range: 1.71–1.89 s), while sequences with inserted pause and lengthening had a mean duration of 2.36 s (range: 2.27–2.42 s). The sound files for the condition without an internal IPB contained seven sequences (two of the five recordings were randomly chosen and repeated at the end of a sound file) and had an average duration of 18.75 s (range: 18.59–18.86 s). To achieve a similar mean duration, the sound files for the condition with inserted pause and lengthening cues contained only six sequences, resulting in an average duration of 19.2 s (range: 19.09–19.24 s). The interstimulus interval between the sequences in each type of sound file was 1 s.

6.5.3 Procedure

The procedure was identical to that of Experiment II/1 and II/2. All infants were familiarized with sequences without an internal IPB until at least 20 sequences had been presented. This led to a minimum of 54 s of accumulated listening time. The familiarization was immediately followed by a test phase of twelve test trials, half of them containing familiar sequences without an internal IPB, the other half, containing new sequences with an internal IPB cued by pause and final lengthening.

6.5.4 Descriptive results

Infants tested in Experiment II/3 showed a mean listening time of 8.65 s (*SD* = 3.97 s) to the novel test trials and a mean listening time of 7.1 s (*SD* = 3.52 s) to the familiar test trials (see Fig. 6.2). Eleven out of 16 infants had longer listening times to the novel test trials.

6.6 Joint statistical analysis of the experiments

We statistically analyzed the data of all four experiments in a repeated-measures ANOVA with Familiarity as within-subject factor (mean listening times to novel vs. familiar test trials) and Experiment as between-subject factor (Exp. II/1, II/2, II/2a, II/3)[8]. This revealed a significant main effect of Familiarity, $F(1,63) = 5.480$, $p = .022$, but not of Experiment, $F(3,63) = 1.334$, $p = .271$. However, there was a significant interaction of Familiarity and Experiment, $F(3,63) = 5.628$, $p = .002$.

To dissolve the significant interaction and to determine which experiments differed from each other we carried out pairwise comparisons. Therefore, we compared the results of each experiment with those from Experiment II/2 as a control experiment – the one that yielded the smallest listening time differences between novel and familiar test trials. We ran a post-hoc t-test on the difference scores (mean listening time to novel test trials minus mean listening time to familiar test trials) in Experiment II/2 versus II/1, Experiment II/2 versus II/2a, and Experiment II/2 versus II/3. Difference scores are depicted in Figure 6.3.

To adjust for multiple comparisons, the alpha-level of post-hoc t-tests was corrected according to Holm (1979). With three levels of comparisons this resulted in $\alpha = .05$ for the largest *p*-value, $\alpha = .025$ for the mid *p*-value, and $\alpha = .017$ for the smallest *p*-value.

Pairwise comparison of Experiment II/2 versus II/1

The post-hoc test for Experiment II/2 versus II/1 failed to reach significance, $M_{\text{Diff}} = 1.694$ s, $t(26) = - 2.073$, $p = .048$, $\alpha = .025$. However, on the descriptive level, we see a much larger listening time difference in Experiment II/1 compared to Experiment II/2. Infants in Experiment II/1 had a mean listening time difference of $M = 1.584$ s with a preference for novel test trials, whereas infants in Experiment II/2 had a mean listening time difference of $M = - 0.110$ s. Considering the small sample size that presumably prevents statistically significances, this may indicate that six-month-old infants tend to discriminate the two types of prosodic patterns when the IPB is indicated by pitch, lengthening,

---

[8] Note that from Experiment 1 only the data from the group familiarized without IPB was considered (n=12).

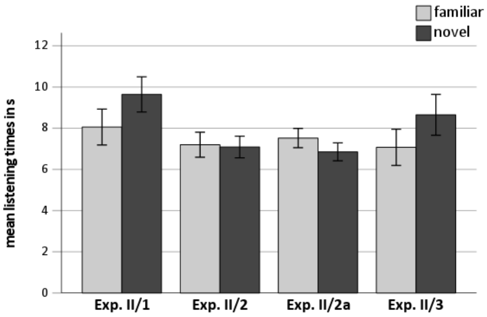

Figure 6.2 Mean listening times to familiar and novel test trials after familiarization with sequences without an internal IPB in Experiment II/1 (all cues), II/2 (pitch and lengthening, short familiarization), II/2a (pitch and lengthening, long familiarization) and II/3 (pause and lengthening). Error bars indicate ± 1 SE.

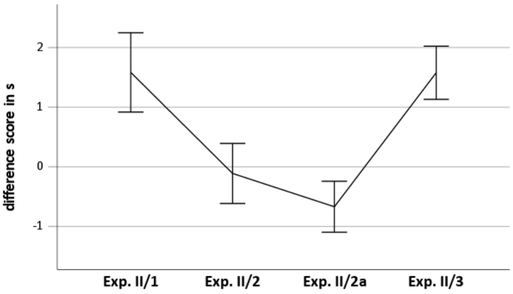

Figure 6.3 Mean listening time differences to novel minus familiar test trials. Error bars indicate ± 1 SE.

and pause, but not when it is cued by pitch and lengthening only (also see the comparison to Experiment II/3 below).

Pairwise comparison of Experiment II/2 versus II/2a

The post-hoc test for Experiment II/2 versus II/2a was not significant, M_{Diff} = .558 s, $t(37)$ = .843, p = .404, α = .05. Infants in Experiment II/2a had a mean listening time difference of M_{Diff} = - .668 s with a slight preference for familiar trials. Infants in Experiment II/2 had a mean listening time difference of M_{Diff} = - .110 s. This comparison indicates that infants' behavior does not differ between Experiments II/2 and II/2a. Hence, the data do not support the hypothesis that a longer familiarization phase leads to better discrimination in six-month-olds suggesting that – unlike eight-month-olds – they still need the pause cue to detect the boundary (see Exp. II/1). However, it is possible that doubling the familiarization time may have reduced infants' general attention during the test phase and may have obscured their discrimination of the test stimuli. This is also indicated by the high drop-out rate, which suggests the modified version of the experiment was especially hard.

Pairwise comparison of Experiment II/2 versus II/3

The post-hoc test for Experiment II/2 versus II/3 almost reached significance, M_{Diff} = 1.688 s, $t(30)$ = - 2.512, p = .018, α = .017. Infants in Experiment II/3 had a mean listening time difference of M_{Diff} = 1.578 s with a preference for novel test trials. Infants in Experiment II/2 had a mean listening time difference of M_{Diff} = - .110 s. We interpret this as a tendency towards a better discrimination of stimuli, in which pause and lengthening indicate the boundary, instead of pitch and lengthening.

Moreover, the results obtained from Experiment II/3 support the interpretation that infants in Experiment II/1 detected the boundary that was marked by pause, lengthening, and pitch. Note that the number of participants was higher in Experiment II/3 (n=16) compared to Experiment II/1 (n=12).

This underlines the issue of low statistical power in Experiment II/1. Across Experiment II/1 and II/3, the mean listening time scores were very similar with both revealing a numerically strong novelty effect. Overall, we interpret the six-month-olds' data as an indicator for successful perception of boundaries that are marked by the full set of cues or by the subset of pause and lengthening. In contrast, the combination of pitch and lengthening seems to be a non-sufficient marking. This points to a crucial role of the pause cue in early prosodic boundary processing in German.

6.7 General discussion

Perceptual reorganization in early speech perception has been reported extensively and in numerous languages for aspects of segmental phonology. However, research on the development of prosody – specifically phrasal prosody – is still sparse. The present study – in combination with the findings from Wellmann et al. (2012) and Holzgrefe-Lang et al. (2016, 2018) – contributes to uncovering a developmental change in the processing of prosodic boundary information in German.

The experiments presented in this paper have addressed the role of pitch change, final lengthening, and pause in boundary detection by German-learning infants. Although the statistical power of our experiments is small due to the low number of infants and the limited amount of trials, our data yield three major results that need to be discussed with caution. First, German six-month-olds (but not four-month-olds) are able to detect a major prosodic boundary signaled by all the cues. Second, pitch change combined with final lengthening did not appear as a sufficient marking for six-month-olds, neither after a prolonged familiarization. Third, six-month-olds do not generally need a combination of all the three cues, but a combination of pause and final lengthening is sufficient to detect the boundary.

We will focus our discussion on the questions that were raised in the introduction: First, what do these results tell us about developmental changes in infant prosodic cue perception? Second, we will embed the results into the previous research on American-English- and Dutch-learning infants, focusing on crosslinguistic similarities and differences in prosodic cue weighting. Beyond, we will compare the present behavioral outcomes to electrophysiological findings (Holzgrefe-Lang et al., 2018) and discuss the cognitive demands underlying a potential asymmetry in prosodic cue perception.

6.7.1 Developmental changes in German boundary perception

The results suggest that German-learning infants have developed a sensitivity to fully marked IPBs by six months and even to a subset of boundary cues containing pause and final lengthening. While six-month-olds heavily rely on the pause cue to detect a boundary, two months later infants have enhanced their sensitivity to boundaries by perceiving a subtle difference indicated by pitch and lengthening (Wellmann et al., 2012)[9]. Perceptual attunement in the acquisition of German phrasal prosody mainly concerns the necessity of the pause cue. Two observations from German, 1) a rather inconsistent occurrence of the pause cue in German adult-directed speech (ADS) as well as 2) no or only short pauses at minor syntactic boundaries, underline the usefulness of a German learner's ability to detect a boundary without pause, and hence support an enhancement in the sensitivity to boundary cues.

At first glance, the occurrence of a pause seems to be highly reliable with respect to its function as a linguistic boundary cue in infant-directed speech (IDS): whenever pauses occur they are likely to indicate a sentence boundary (Fernald & Simon, 1984; Fisher & Tokura, 1996). However, a pause does not seem to be the predominant boundary cue in German ADS and occurs only rarely as a single cue: Peters et al.'s (2005) analysis of phrase boundary markings in the German Kiel Corpus of spontaneous ADS[10] showed that pauses occurred only at 38 % of all boundaries, while pitch changes did so at 74 % and lengthening at 66 %. An essential finding was that cue combination at boundaries was a frequent pattern, occurring at 61.6 % of all boundaries. Among these, the co-occurrence of pitch and lengthening (24.6 %) and the coalition of all three cues (23.7 %) were the most frequent. Cue combinations including only pause and one additional cue were comparatively infrequent: only 8.4 % of all boundaries were marked by a combination of pause and lengthening, and 4.9 % by a combination of pause and pitch. Each prosodic cue also occurred as a single cue: pitch alone marked 20.8 % of all boundaries, while lengthening cued 9.4 %, and pause only 1.3 %. In brief, Peters et al. (2005) revealed pause to be the least frequent and the combination of pitch and lengthening to be the most frequent marker. This implies that, at least in German ADS, a large proportion of phrase boundaries are not signaled by a pause, which would cause a segmentation problem for learners who overly rely on the occurrence of a pause. Unfortunately, corresponding data on cue frequency in German IDS are lacking, but Fernald and Simon (1984) report longer pause duration and a higher correspondence between pause and sentence boundaries in German IDS compared to ADS. Also, a systematic review by Ludusan et al. (2016) across several languages suggests that pause duration is increased in IDS. So, it may be the case that a high reliance on pause as a boundary cue is appropriate when exposed to IDS, but not

---

[9] Four-month-old infants displayed a null result in the same Experiment II/1 which allows for two interpretations: either, four-month-olds are not yet able to detect fully marked IPBs, or they are able to, but can't show their ability with this kind of method. Even though there are few studies showing that the HPP methods in principle works with four-month-olds (e.g., Bosch & Sebastián-Gallés, 2001; Herold, Höhle, Walch, Weber, & Obladen, 2008; Seidl & Cristià, 2008), the familiarization design might not be optimal for this young age group as it depends on a rather high working memory load (process the auditory information and store them to detect the change in the test phase). A study by Höhle et al. (2009) revealed a null result for four-month-olds in a familiarization technique, whereas Herold et al. (2008) evidenced discrimination in the same age group, with the same materials, but a change in the experimental setup to a discrimination technique without a familiarization phase. Hence, a more simple preference paradigm might have worked better with our materials in the four-month-olds; however, the data would not be directly comparable to the data of the older infants.

[10] In this analysis, all auditory breaks that occurred turn-internally within the continuous speech stream were classified as phrase boundaries.

when exposed to ADS, making a change in the reliance on this cue necessary to become a proficient processor of ADS.

The second argument that relying solely on the pause cue is not an optimal strategy comes from the fact that pauses signal major prosodic and syntactic boundaries, whereas minor prosodic and syntactic boundaries like phrase boundaries are less often marked by a pause (Strangert, 1991; Terken & Collier, 1992) or they are marked by pauses of shorter durations (e.g., Butcher, 1981; Goldman-Eisler, 1972). Thus, if prosodic cues are essential for infants for detecting not only major clause boundaries in the signal but also boundaries of smaller units within these larger domains, children must become more sensitive to boundary markings that do not involve a pause. To sum up, the developmental change evidenced here seems to be in line with the requirements of the ambient language German.

6.7.2 Crosslinguistic comparison

Turning to the next point of the discussion – the crosslinguistic dimension of boundary cue perception – our results reveal similarities as well as dissimilarities in Dutch, German, and American English infants. First, the finding that six-month-old German infants are sensitive to naturally occurring, fully marked IPBs is in line with results from previous studies with Dutch- and American-English-learning infants (Johnson & Seidl, 2008; Nazzi et al., 2000; Seidl, 2007; Soderstrom et al., 2005) and thus expands the crosslinguistic evidence that young infants are sensitive to natural prosodic phrasing. For American English, even four-month-olds have been shown to use fully marked boundaries for the segmentation of complex clauses.

We consider our materials – rather short and phonologically highly-controlled sequences with successively inserted cues – an important extension to the crosslinguistic field of infant prosodic boundary perception. A recent study by van Ommen et al. (2020) created similar stimuli in French to be presented to French- and German-learning infants. The sequences were three coordinated French names either with a major prosodic boundary *[Loulou et Manu][et Nina]* or without a boundary *[Loulou et Manou et Nina]*. Hence, materials were identical in their structure to the concatenation of three German names used in the present study. Also the procedure was the same, a discrimination task in the HPP paradigm with a familiarization phase followed by a test phase. The results showed that French six- and eight-month-olds perceived the boundary when it was signaled by all the three cues, but none of the two age groups was successful when the boundary was cued by pitch and lengthening only. Interestingly, in contrast, German infants presented with the same French materials perceived the pitch-lengthening cued boundary at eight months, but not at six months. This result reinforces the data from the present study and from Wellmann et al. (2012) and supports the interpretation of a developmental change in German boundary perception related to the pause cue. Moreover, no such developmental change can be observed in French infants' performance pointing to language-specificity of the respective development.

Crosslinguistic similarities

A similarity found across the American English, Dutch, German, and French studies is that pause seems to be a necessary boundary cue for the youngest groups of tested infants (Dutch- and German-learning six- and American-English-learning four-month-olds, for French even in both six- and eight-month-olds) – independently of whether the specific task requires a segmentation or a discrimination of stimuli. This points to a language-general way of processing prosodic boundaries in the first months of life that is strongly related to the acoustically salient pause cue.

A strong reliance on the pause is useful since among the three main boundary cues, pause is the most "universal" one. In the languages in which infant boundary perception has been studied, pauses have many pragmatic and paralinguistic functions, however, when it comes to linguistic structure, it serves only one function, that is, the marking of syntactic boundaries. This may render pause a crosslinguistically highly reliable cue. Note that preboundary lengthening and pitch may bear more than one linguistic function. Duration as the acoustic correlate of lengthening is also used to express lexical and/or phrasal stress as well as phonemic contrasts (vowel duration). Regarding pitch, the majority of the world languages is tonal, this means that pitch is used to express different lexical items. In pitch-accent languages like Japanese or stress languages such as English pitch can also be used to distinguish word meanings. Moreover, pitch bears several functions at the sentence level for example such as the distinction between declaratives and questions. Moreover, pause is a perceptually rather salient feature of an acoustic signal and it provides categorical information that can be processed locally because the presence or absence of silence can be detected immediately in the signal. This may be different for the other two boundary cues, pitch changes and lengthening, which constitute relational information and require the parsing of longer strings to recognize any changes in pitch and duration at the location of the boundary in relation to the whole speech string.

Therefore, the available results by infants learning American English, Dutch, German, and French revealing that the occurrence of a pause is initially required for boundary detection may reflect a rather universal processing that initially relies on the pause as an acoustically salient categorical cue that can be easily processed independent of the contextual information in every language environment. [11]

---

[11] We are grateful to one reviewer who suggested to link our findings to those of individuals with acquired language impairments in which the special relevance of the pause cue is also evident (Aasland & Baum, 2003). When tested on resolving syntactic ambiguities in coordinate structures a group of individuals with aphasia after left-hemispheric brain damage – in contrast to a control group – was not able to consistently identify the phrase boundary cued by lengthening and pause (with neutral pitch). However, when the pause duration was increased beyond normal ranges accuracies improved. Thus, pause also seems to play a crucial role in impaired comprehension and may enable boundary detection even in the absence of the pitch cue.

Crosslinguistic differences

The major difference in the development between Dutch, American English, German, and French concerns if, and if so, when, infants respond to boundaries that are not marked by a pause. For Dutch, we only know that six-month-olds need the pause for boundary detection since older infants were not tested (Johnson & Seidl, 2008). French infants still need the pause cue at eight months (van Ommen et al., 2020). In American-English-learning infants, the necessity of a pause as a boundary cue disappears already between the ages of four and six months (Seidl, 2007; Seidl & Cristià, 2008). This is when in German-learning infants the perception of fully cued boundaries first emerges. Only between the ages of six and eight months a developmental change occurs that makes the pause cue no longer necessary.

Seidl and Cristià (2008) interpret the behavior of the four-month-old American-English-learning infants as a so-called holistic processing in which all cues are equally attended to. They argue that this reflects a general processing mechanism rather than a linguistically based strategy. By six months, American-English-learning infants do assign more weight to pitch. Seidl and Cristià explain this development through the increased language exposure allowing to observe the distribution of boundary cues in their native language. Infants may have learned by this age, that pauses are unreliable boundary cues, whereas pitch is a more reliable cue to syntactic boundaries in American English.

Comparing the developmental trajectory between American-English- and German-learning infants the data reveal that the development is different at six months. American-English-learning six-month-olds need the pitch cue (in any combination with another cue), while this seems not to be the case for the German six-month-olds. This may suggest that pitch information is more salient and thus more important for American English learners than for German learners. In fact, crosslinguistic comparisons of pitch in IDS have found that the mean, minimum, and maximum F0 as well as the F0 variability is significantly higher in American English IDS than in German IDS (Fernald & Simon, 1984; Fernald et al., 1989). Hence, American-English-learning infants may be more prone to attend to pitch variation in their input than German-learning infants.

A comparison of German- and Dutch-learning six-month-olds (Johnson & Seidl, 2008) reveals similarities. Like their German age-mates, Dutch-learning infants did not respond to the prosodic boundary marked by pitch change and final lengthening, indicating a crucial role of the pause in Dutch as well. Interestingly, Dutch – like German – IDS was found to show a lower mean and a lower range in F0 difference compared to American English IDS (Fernald et al., 1989; Weijer, 1997), suggesting again that the properties of the specific speech input relate to crosslinguistic differences in how infants process prosodic information and that with less pitch variation pauses may become a more crucial cue for the marking and the perception of prosodic boundaries.

At eight months, German infants' sensitivity has developed to perceiving a pitch-lengthening cued boundary to such an extent that it can even be applied in a non-native language (van Ommen et al., 2020). The result that French eight-month-olds' boundary detection still depends on all three cues points to a delay in comparison to the German-learning infants. This is supported by van Ommen et al.'s experiment with French and German adults who did not differ in their discrimination of sequences with pitch and lengthening only. Apparently, French listeners catch up at one point. Van Ommen et al. (2020) argue that the language-specific differences at eight months might stem from a higher prosodic variability in German providing a larger basis to attend to prosodic details. French does not use prosodic characteristics to mark lexical stress. It uses prosody for phrasal stress, however, phrasal stress coincides with phrasal boundaries by default. Hereby, French is highly regular in the employment of prosodic cues. German, on the contrary, employs a larger variety of tonal and duration patterns at the phrasal as well as at the lexical level, and these are not strictly aligned to boundaries. This might explain why the German-learning infants show an earlier sensitivity to the specific cue combinations than their French peers.

The comparisons between the German, American English, Dutch, and French studies (that, notably, varied in the experimental paradigms: discrimination vs. segmentation) only give first, still vague indications of crosslinguistic differences in prosodic boundary cue weighting that support the assumption that perceptual attunement occurs in this domain. Future research using more comparable materials and methods across languages is necessary to provide a reliable picture of potential crosslinguistic effects of boundary perception and their development. In addition, it is not clear whether the few studies so far have used acoustic instantiations of the different cues that are typical for the specific language and typical for the infants' input. Corresponding prosodic analyses of ADS and IDS in the respective languages are therefore needed to broaden our understanding of the early prosodic development. A further limitation in the interpretation of the results concerns the strength of single cues, which might differ between different cue constellations. There is evidence that marking of prosodic boundaries is subject to cue trading relations, that is, an interaction between the strength of the cues that mark the boundary with one cue being stronger when another cue is weaker (e.g., Beach, 1991). The cue insertion applied to the German stimuli was based on the acoustic parameters of pitch, lengthening, and pause that had been measured in natural sequences with a fully marked IPB. Although these values were already quite high (a pitch rise of 212 Hz/13.65 semitones, a final lengthening factor of 1.8, and a pause of 500 ms duration; cf. Peters, 2005[12]), they might be even higher in natural sequences with a boundary that is only marked by the subset of pitch and lengthening; in other words, when pause is missing, the other cues may be enhanced. Thus, we cannot

---

[12] In a perception study with adult listeners, Peters (2005) implemented lengthening factors of 1.2, 1.4, and 1.6 and pauses with a duration continuum between 50 ms and 890 ms—based on the values found in German ADS (Peters et al., 2005).

exclude that six-month-olds might also be able to detect a boundary without pause if we had implemented stronger pitch and/or lengthening cues. Given that continuous stimulus manipulations can hardly be investigated in behavioral tasks in infants, the present experimental design did not consider cue trading relations. Still, we can conclude that developmental changes in behavior occur, since the eight-month-olds were able to detect the boundary using identical materials.

6.7.3 Behavioral versus neurophysiological methods

In the final section, we compare the present behavioral finding to those of a previous electrophysiological study. Using the very same stimuli, Holzgrefe-Lang et al. (2018) investigated boundary perception in eight- and also six-month-old German-learning infants by means of event-related potentials (ERP). In adults, the processing of IPBs with and without a pause evokes a specific ERP component, the so-called closure positive shift (CPS; e.g., Holzgrefe-Lang et al., 2016; Steinhauer, Alter, & Friederici, 1999), which is assumed to reflect the perception of a prosodic boundary. Holzgrefe-Lang et al. (2018) investigated whether an infant CPS can be elicited in response to different cue constellations. Specifically, they compared six- and eight-month-old infants' brain response to stimuli containing either no boundary cue, a combination of pitch change and lengthening, or only a pitch cue. The ERPs in response to the latter condition did not differ from the condition without any boundary cues, but the combined occurrence of pitch change and final lengthening elicited a positivity that resembled the adult CPS in both age groups (Holzgrefe-Lang et al., 2016). Hence, the electrophysiological data suggests that six- and eight-month-old German infants do not differ in IPB perception, whereas the current HPP data provides no evidence that six-month-olds detect pitch and lengthening cued boundaries, but suggests a developmental change between the ages of six and eight months. Thus, prosodic boundary perception without the pause cue is evidenced earlier at the electrophysiological level (but see Männel & Friederici, 2009 and Männel, Schipke, & Friederici, 2013 for data that indicate that stimuli with neutralized pause cues would only elicit a CPS in children older than three years). In line with this asymmetry, there is ample evidence from other studies (Friederici et al., 2007; Höhle et al., 2009; Schipke, 2012) that a specific brain response may precede the corresponding behavioral response in the course of development. For instance, the recognition of the ambient language's dominant stress pattern has been shown for four-month-old German learners using ERPs (Friederici et al., 2007), whereas a behavioral preference is evident only at six months (Höhle et al., 2009).

We assume that the diverging results across the different methods are due to different cognitive demands during testing. ERPs are measured on-line during infants' passive listening, and hence do not depend on task demands or an overt response performance involving additional processing requirements (see Männel & Friederici, 2008). In the case of the ERP study by Holzgrefe-Lang et al.

(2018), the brain response indicating the perception of a prosodic boundary marked by pitch and lengthening occurs right after the presentation of the phrase-final syllable. Hence, the ERPs represent immediate processing responses evoked by the presence of specific boundary cues in the stimuli. In the HPP, the infants' behavioral response is measured by the amount of listening time indicated by the infant's head turn towards the side of presentation. The expectation to observe differences in listening times between the conditions in the HPP experiment is based on the assumption that a representation of the familiarization stimulus has been formed during the familiarization phase and that the stimuli presented during the test phase are mapped onto this representation. Establishing this representation requires that at least some memory traces survive the switch to the test phase of the experiment. Such a long-lasting memory component is not involved in the ERP paradigm. Nevertheless, the fact that the six-month-olds show effects in the HPP when the boundary is marked by the full set of boundary cues or by the subset of pause and lengthening suggests that memory requirements alone are not sufficient to explain the null effect in the condition with only lengthening and pitch. Rather, differences in the level of attention might account for the different outcomes across the ERP and HPP measurements as well as for the different outcomes across ages in the behavioral studies. Considering that listening times are an indicator of attention, the change that we observed between the six- and the eight-month-olds in the HPP data may suggest that infants have sharpened their attention to pitch and lengthening, which are functionally relevant cues to German phrase boundaries, at the age of eight months.

6.8 Conclusion

To conclude, the present study provides evidence that German six-month-old infants are able to detect a major prosodic boundary characterized by the three main cues. This ability is crucial for the first steps in language acquisition as it equips the naïve learner with a tool to chunk the continuous speech stream into clauses. In a HPP discrimination task, we found that for six-month-olds pitch and lengthening cues are not sufficient, but they need the pause cue. Boundary detection on the basis of combined relational prosodic cues like pitch changes and final lengthening shows up only by eight months. We argued that this behavioral change displays an enhancement in sensitivity that is reflected in a shift of attention to boundary markings that are functionally relevant in the ambient language. The ability to detect a boundary with the full as well as with a subset of cues enables syntactic parsing of not only major, but also minor, syntactic units. This ability is also necessary for the adult listener, especially in the case of structural ambiguities. Therefore, being able to detect these boundaries with their language-specific markers is essential to becoming an efficient processor of a given language.

Study III

7 Infants‘ use of prosodic phrase boundary cues: Pilot experiments on a segmentation task

Studies I and II have shown that German-learning six-and eight-month-old infants are able to perceive an IPB that is indicated by pitch change, final lengthening, and pause. Moreover, it has been shown that for eight-month-olds even a subset of two cues as indicator is sufficient to discriminate between with- and no-boundary sequences. Building up on the close prosody-syntax interface and infants’ sensitivity to the acoustic correlates of prosody, the third assumption of the prosodic bootstrapping theory proposes that infants use prosodic cues to step into syntax. While the materials and the procedure used in Studies I and II solely tested the discrimination of different prosodic groupings and the relevance of each individual prosodic boundary cue for the perception of such a grouping, Study III aims at developing a more complex experimental setting which demands infants to use the same cues in order to segment and recognize prosodic units.

Previous studies provided evidence that infants indeed make use of prosody to encode information within a prosodic unit and also to segment clauses. In two studies, Mandel and colleagues (1994, 1996) applied the high-amplitude-sucking procedure to investigate whether two-month-old infants can make use of prosodic structuring in order to remember acoustic properties of the speech signal. The first study by Mandel et al. (1994) comprised two experiments. In the first experiment, infants were assigned to one of two experimental groups that was either habituated to a sequence of words that occurred in a complete sentence with natural prosody, for example, “The rat chased white mice.”, or to the same sequence of words that occurred as a list of words produced in isolation, for example, “the, rat, chased, white, mice”. During the test phase, the experimental groups listened to the same words in the same prosodic context (clause or list), but one or two phonemes had changed (one phonetic change, e.g., “the cat chased white mice” or two phonetic changes, e.g., “the cat raced white mice”.) The results revealed that only those infants that were habituated with the clause context detected both, one and two phonetic changes. This was expressed by a significant increase in sucking rate relative to the control group (that listened to exactly the same stimuli during habituation and test phase). The group of infants that had been habituated with a list of isolated word, did not detect the single phonetic change. When stimuli contained two phonetic changes, infants in the list group showed a tendency to detect them (i.e., the sucking rate marginally increased compared to their control group). Hence, at least in the case of the more demanding one-phonetic change condition the clause context was more supportive for remembering the words. From these findings, the authors concluded that prosodic chunking helps infants to encode phonetic details. They argued that “prosody serves as a kind of “perceptual glue” – it has a psychological reality as a packager of information, and facilitates the processing and retention of linguistic information in memory” (Mandel et al., 1994: 166).

In a second experiment, the clause context was compared to a *two-fragment context* that replaced the list of isolated words. This was done to rule out a potential influence of local co-articulatory cues that were present in the sentences, but not in the list of words of the prior experiment. In the two-fragment context, co-articulation between words was present. Hence, different outcomes between the clause and the two-fragment contexts would only relate to rather global prosodic characteristics. The two-fragment context referred to a four-word sequence in which the first and the last two words were prosodically related. Between these two prosodic chunks a clausal boundary occurred, for example, “cats like # park benches” spliced out from the two adjoining clauses “Brigid really knows what cats like. Park benches are their favorite things to climb on.” One group of infants was habituated to such a four-word sequence in the two-fragment context, the other one to the same sequence in the clause context. During the test phase, the four-word sequence was presented in the same prosodic context, either in a condition with one phonetic change or in a condition with two phonetic changes. The single phonetic change occurred either in the verb (e.g., “cats strike. Park benches”), or in the adjective (e.g., “cats like. Dark benches”). The two-phonetic change condition split into a condition in which the two changes occurred within the same fragment (e.g., “cats like. Dark wrenches”), and another condition in which the two changes occurred across both fragments (e.g., “cats strike. Dark benches”). Significant increases in sucking rate compared to the control groups showed up for the one- and the two-phonetic change group in the clause context. In the two-fragment context, neither in the one-phonetic change nor in the two-phonetic change condition infants differed from the control group. Regarding the two-phonetic change group, there was no difference between the group that listened to within-fragment changes and the group that listened to changes across both fragments. Hence, only infants in the clause, but not in the two-fragment context discriminated the phonetic changes between the habituation and test stimuli. Mandel et al.’s (1994) study provided evidence that infants are better able to remember phonetic information when it is embedded in a single well-formed prosodic unit implying that the prosodic packaging of clauses enhances infants’ memory capacities.

In a further study with a similar experimental procedure, Mandel et al. (1996) investigated whether two-month-old infants detect a change in word order, for example, from “cats would jump benches”, to “cats jump wood benches”. Only the group that was habituated with the four words in a clause context detected the change in word order, but not the group that was habituated to the same words in a two-fragment context. Hence, with support from clausal prosody infants are sensitive to serial word order. Mandel et al.’s studies (1994, 1996) were pioneering revealing that infants can make use of prosody for organizing and encoding the speech signal. The authors claimed that prosody keeps the information of important units together which seems necessary when acquiring syntax and thus, is in line with the proposals of the prosodic bootstrapping account. However, their studies did not directly investigate whether infants derive syntactic units from prosodic structuring. Only a couple of years later, this kind of evidence was provided by a study by Nazzi et al. (2000, for a description see Chapter 2.4.1). They showed that infants can recognize a prosodically well-formed clause within a passage of continuous speech, but not the same sequence of words when it straddles a prosodic boundary. This study gave the first hint that infants use prosodic cues to segment the speech stream into syntactic units such as clauses. Soderstrom, Nelson, and Jusczyk (2005) added evidence that clause segmentation is even possible when the same kind of familiarization sequence that Nazzi et al. (2000) used, is embedded in continuous speech throughout the familiarization and the test phase. Again some years later, the series of studies by Seidl (2007), Johnson & Seidl (2008) and Seidl & Cristià (2008) added insights how each individual prosodic boundary cue contributes to infants' clause segmentation (for a description see Chapter 4.1).

Based on the studies by Mandel et al. (1994, 1996), Nazzi et al. (2000) and Soderstrom et al. (2005) that suggest that prosodic packaging enhances infants' recognition and memories, Study III aimed at infants' recognition of a test sequence in relation to the prosodic structuring of the familiarization stimulus. Specifically, it was asked whether infants are better able to segment and recognize a two-name sequence (N1 and N2) from a three-name sequence (N1 and N2 and N3), if the two names are familiarized as one prosodic chunk of one IP [N1 and N2]$_{IP}$ [and N3]$_{IP}$, compared to if the two names are familiarized as a subpart of an IP [[N1 and N2]$_{PP}$ and N3]$_{IP}$, or if the two names are familiarized as straddling a prosodic boundary within the three-name sequence, [N1]$_{IP}$ [and N2 and N3]$_{IP}$.

As in Studies I and II, the stimuli of Study III were sequences of coordinated names only containing sonorant sounds. Three experiments were conducted. Experiments III/1 and III/2 familiarized infants with two combinations of three names that either contained an IPB after the second name or not. Both experiments only differed in whether the test stimuli contained the first or the last two names of the familiarization stimuli and were run in parallel. Based on the results of these experiments, the third experiment was modified such that both familiarization stimuli included an internal IPB within a combination of three names, either after the first or after the second name. Infants were tested on their recognition of the first two names.

The aim of Study III was to find a successful design that uncovers infants' use of the three main boundary cues pitch change, final lengthening, and pause. Based on this design, in future experiments – that are not part of the dissertation – the prosodic boundary information that indicates the IPB in the familiarization stimulus could then be reduced and varied similar to Studies I and II (for example with only pitch and final lengthening cues or pause and lengthening signaling the IPB).

7.1 Experiment III/1

Experiment III/1 tested whether eight-month-old German-learning infants would use the prosodic boundary information given in a familiarization sequence of three names to segment and recognize a sequence of two names out of it. The combination of names were “Moni und Lilli und Manu” und “Leni und Manni und Lola”. In a crossed design, each combination of names was presented in two different prosodic phrasings, either with or without an IPB, resulting in two familiarization versions. Files for Group A contained the sequence [Moni und Lilli]$_{\text{IP}}$ [und Manu]$_{\text{IP}}$ as the prosodic type *With IPB* and the sequence [Leni und Manni und Lola]$_{\text{IP1}}$ as the prosodic type *Without IPB*. Files for Group B included the same sequences of names in the opposite prosodic type, that is, [Moni und Lilli und Manu]$_{\text{IP}}$ as a sequence of the type *Without IPB* and [Leni und Manni]$_{\text{IP}}$ [und Lola]$_{\text{IP}}$ as a sequence of the type *With IPB* (see Tab. 7.1). Infants were randomly assigned to one of the groups. After familiarization, infants were tested on sequences of two names that either contained the first two names of the familiarized sequence of the prosodic type *With IPB* (e.g., “Moni und Lilli”[13]) or the first two names of the familiarized sequence of the type *Without IPB* (e.g., “Leni und Manni”). The test sequences were produced in isolation as a sequence of two names that formed one well-formed prosodic unit, that is, they were not spliced from the three-name sequences.

The assumption underlying this experimental design was that the prosodic structure of the familiarization stimulus *With IPB* (e.g., [Moni und Lilli]$_{\text{IP1}}$ [und Manu]$_{\text{IP2}}$) would support a grouping of the three names into two chunks, one of two names [Moni und Lilli]$_{\text{IP1}}$ and the other a single name [und Manu]$_{\text{IP2}}$. Infants familiarized with the sequence of names with IPB should build up a representation of two IPs of this sequence. Subsequently, the test sequence that contained the first two names of this familiarization stimulus “Moni und Lilli” would display a familiar IP. This test condition is called *IP*. Under the assumption that infants are able to identify the first IP of the familiarization sequence and that this prosodic structuring supports memory capacities, the names of the test condition *IP* should be easier recognized, that is, listening times to this kind of test sequence should differ from those to the other test condition.

In contrast, for the familiarization stimulus *Without IPB* [Leni und Manni und Lola]$_{\text{IP1}}$, infants are assumed to build up a representation of only one single IP. A grouping into IPs within the three names is not supported. The lexically corresponding test sequence “Leni und Manni” would be a part of the single IP. This test condition is hence called *part-IP*. The fact that this test sequence would not be mentally represented as one chunk (rather as a part of a larger chunk of three names), should be a disadvantage in remembering these two names. If infants’ listening times significantly differ between test conditions *IP* and *part-IP* this may indicate that infants make use of the IPB cues provided in the

---

[13] This and all the following examples illustrating the set-up of the stimuli refer to the materials of Group A.

familiarization stimulus to segment the first two names apart from the third which in turn enables them to remember the two lexically corresponding names during the test phase.

Table 7.1 Experimental design of Experiment III/1. Familiarization and test stimuli for Group A and B.

group	familiarization stimuli		test condition	
	with IPB	without IPB	IP	part-IP
A	**[Moni und Lilli]** [und Manu]	*[Leni und Manni* und Lola]	**Moni und Lilli**	*Leni und Manni*
B	**[Leni und Manni]** [und Lola]	*[Moni und Lilli* und Manu]	**Leni und Manni**	*Moni und Lilli*

7.1.1 Participants

Sixteen eight-month-old infants (9 girls) were tested. The mean age was 8 months, 19 days (range: 8 months, 3 days to 8 months, 30 days). All infants were from monolingual German-speaking families, born full-term, and normal-hearing. Twelve additional infants were tested but their data were not included in the analysis for the following reasons: failure to complete the experiment (4), crying or fussiness (3), technical problems (3), experimenter error (2).

7.1.2 Stimuli

The familiarization stimuli for the combination “Moni und Lilli und Manu” were the same recordings that had been used in Studies I and II. Additionally, the sequences of the combination “Leni und Manni und Lola” and the test sequences “Moni und Lilli” and “Leni und Manni” were recorded under the same technical conditions by the same young female speaker from Brandenburg who had produced the recordings for Studies I and II. The speaker produced six repetitions (tokens) of each sequence.

An acoustic analysis revealed clear differences between the familiarization stimuli across the two prosodic groupings with and without IPB, while both combinations of names in the grouping *With IPB* were comparable with respect to the strength of the prosodic cues at the critical boundary positions.

Table 7.2 Means and range of the acoustic correlates of prosodic boundary cues in familiarization stimuli at the same position in sequences with IPB and in sequences without IPB.

acoustic correlate	with IPB		without IPB	
	[Moni und Lilli] [und Manu]	[Leni und Manni] und Lola	[Moni und Lilli und Manu]	[Leni und Manni und Lola]
max F0 in Hz on name 2	397 (371–422)	394 (345-449)	277 (264–293)	263 (246-273)
F0 rise (range) on name 2 … in Hz … in semitones	220 (197–240) 14.0 (12.8–14.6)	220 (165-277) 14.1 (11.3-16.6)	88 (77–110) 6.7 (5.8–8.2)	87 (62-104) 6.9 (5.0-8.1)
duration of final vowel[14] of name 2 in ms	175 (162–186)	178 (144-194)	99 (91–110)	115 (98-138)
duration of pause after name 2 in ms	506 (452–556)	486 (461-517)	0	21 (0-41)

---

[14] Only the steady party of the vowel was measured, that is, without parts of creaky voice or transition to the following conjunction.

Sequences of the prosodic type *With IPB* had a flat tonal contour on the first name, whereas on the second name a large pitch rise of 220 Hz on average occurred, leading to an upstepped high peak of 397/394 Hz[15] on average. In contrast, sequences of the prosodic type *Without IPB* followed a downstep pattern with a smaller pitch rise at the second name of only 88/87 Hz on average. The final vowel of the second name was 1.8/1.6 respectively times longer in sequences with IPB compared to the same vowel in sequences without IPB. Finally, sequences with IPB contained a pause with a mean duration of 506/ 486 ms, whereas in sequences without IPB no or only a minimal pause occurred (see Tab. 7.2). Examples for familiarization sequences are depicted in Figure 7.1(A-D). In sum, the main prosodic boundary cues pitch rise, final lengthening, and a pause showed up in sequences with IPB, but not in sequences without IPB.

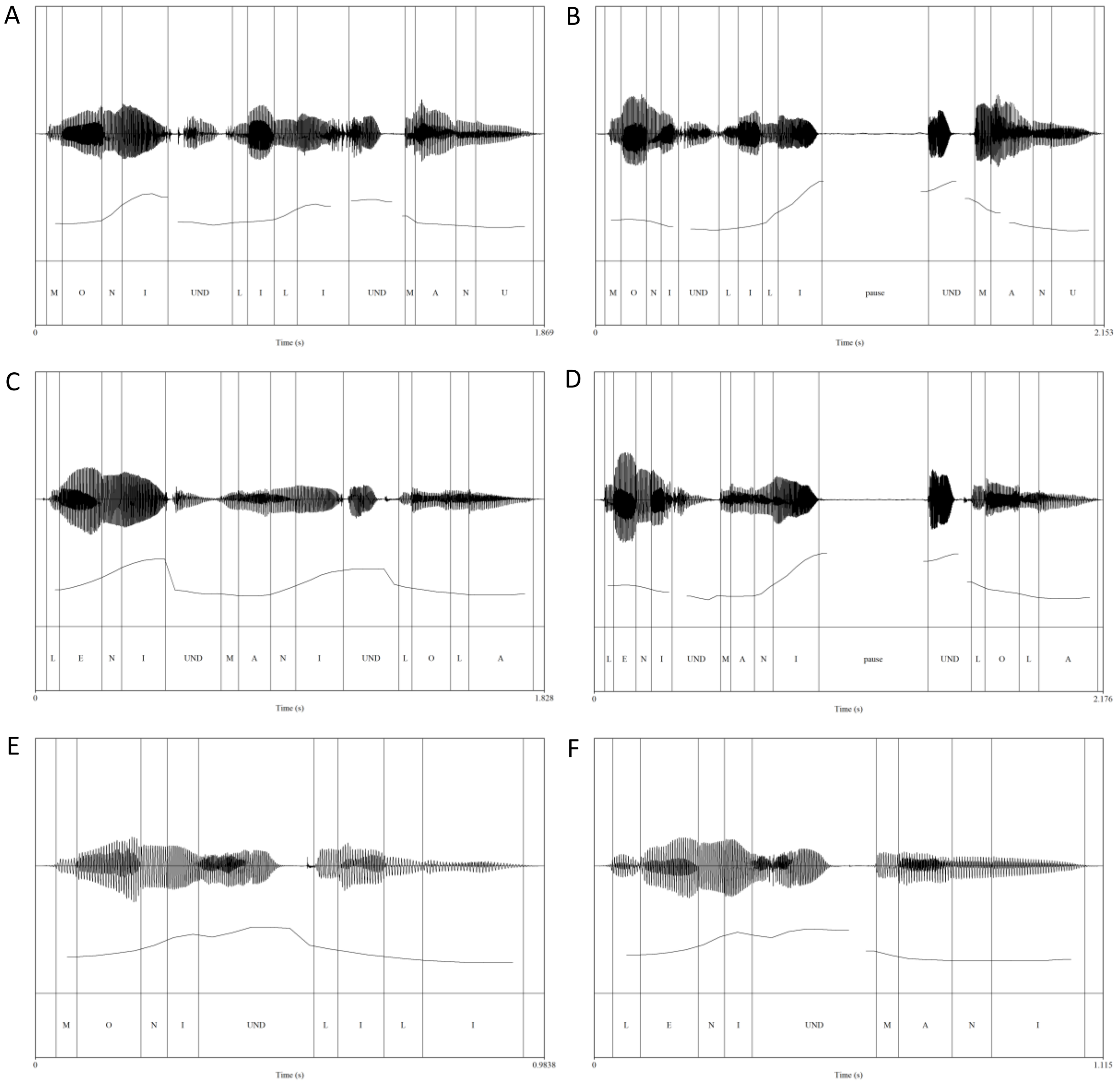

Figure 7.1 Oscillograms and pitch contours aligned to the text. (A) Familiarization sequence “Moni und Lilli und Manu” without an IPB, (B) Familiarization sequence “Moni und Lilli und Manu” with an internal IPB, (C) Familiarization sequence “Leni und Manni und Lola” without an IPB, (D) Familiarization sequence “Leni und Manni und Lola” with an internal IPB, (E) Test sequence “Moni und Lilli”, (F) Test sequence “Leni und Manni”.

---

[15] The first value refers to the name combination “Moni und Lilli und Manu”, the second one to “Leni und Manni und Lola”.

An acoustic analysis was also carried out for the test sequences. It revealed a pitch rise of 7.3 (7.0)[16] semitones on average on the first name. On the following conjunction “und” the pitch contour still rises, leading to a peak of 305 (311) Hz. On the second name the pitch contour is falling by 5.4 (6.4) semitones on average, leading to a low boundary tone of 169 (167) Hz. Within the sequences no pauses occur. The sequence-final vowel yields the longest duration of all phonemes. Comparing the duration of the vowel */i/* at the final position of the first name, that is, sequence-internally, to the same vowel at the sequence-final position (e.g., “Moni̲ und Lilli̲”) reveals final lengthening at the end of the sequence by factor 2.6 (2.3). The low boundary tone and the final lengthening clearly indicate the ending of the two-name sequences. Hence, the test sequences represent a well-formed, full IP. Examples for both test sequences are depicted in Figure 7.1(E,F).

All recordings, familiarization and test sequences, were scaled to 70 dB. From the six repetitions that had been recorded for each familiarization and test sequence, different speech files were constructed to be presented during familiarization and test trials. Each infant listened to only one familiarization file.

Familiarization speech files consisted of 24 sequences in total, equally presenting both prosodic types – *With* and *Without IPB*. The prosodic type was alternating after six tokens of the same sequence (see Tab. 7.3 for the construction). These tokens were randomly concatenated with an interstimulus interval of 1 s. When the prosodic type changed, the interstimulus interval was 2 s. The beginning prosodic type was counterbalanced resulting in two different speech files for Group A and B, each. The total length of sound files, and hence the duration of familiarization, was 72.5 s for Group A and 71.74s for Group B.

Table 7.3 Construction of familiarization speech files

file	familiarization sequence	ISI	familiarization sequence	ISI	familiarization sequence	ISI	familiarization sequence
A1	6x With IPB [Moni und Lilli] [und Manu]	2s	6x Without IPB [Leni und Manni und Lola]	2s	6x With IPB [Moni und Lilli] [und Manu]	2s	6x Without IPB [Leni und Manni und Lola]
A2	6x Without IPB [Leni und Manni und Lola]	2s	6x With IPB [Moni und Lilli] [und Manu]	2s	6x Without IPB [Leni und Manni und Lola]	2s	6x With IPB [Moni und Lilli] [und Manu]
B1	6x With IPB [Leni und Manni] [und Lola]	2s	6x Without IPB [Moni und Lilli und Manu]	2s	6x With IPB [Leni und Manni] [und Lola]	2s	6x Without IPB [Moni und Lilli und Manu]
B2	6x Without IPB [Moni und Lilli und Manu]	2s	6x With IPB [Leni und Manni] [und Lola]	2s	6x Without IPB [Moni und Lilli und Manu]	2s	6x With IPB [Leni und Manni] [und Lola]

For the test phase, twelve speech files were created. Half of them contained the sequence “Moni und Lilli”, the other half the sequence “Leni und Manni”. Each file contained 12 repetitions of the same test sequence (the six recordings were repeated twice), randomly concatenated with an interstimulus interval of 800 ms. The total length of the test speech files containing “Moni und Lilli” was 20.88s, and for the files containing “Leni und Manni” 21.66 s.

---

[16] The first value refers to the sequence “Moni und Lilli”, the one in brackets to “Leni und Manni”.

7.1.3 Procedure

Each infant was assigned to either familiarization group A or B. The experimental procedure followed the usual HPP with a slight modification during the familiarization phase. As usual, the three lamps were fixed inside the booth: a green one on the center wall, and two red ones on each of the side walls. Each experimental trial started with the blinking of the green center lamp. When the infant oriented to the green lamp, the speech stimulus was started immediately and played continuously until the end of the familiarization speech file was reached irrespective of the infant's looking behavior. When the infant turned her head towards one of the red lamp, the green lamp was turned off and one of the red lamps on a side wall started to blink. The blinking of the red lamp stopped when the infant turned her head away for more than 2 s, then the green center lamp started blinking again, followed by one of the red lamps. The modification concerned the familiarization speech file that was continuously playing. This was done to avoid interruptions within one familiarization stimulus that would not correspond to the intended prosodic grouping. Familiarization was finished with the end of the speech file, after about 73 s.

During the test phase, both familiarization groups were presented with the same twelve speech files half of them consisting of "Moni und Lilli" sequences, the other half "Leni und Manni". The test phase was divided into three blocks with four trials each. Two out of these four trials contained "Moni und Lilli" sequences, the others contained "Leni und Manni" sequences. Within each block, the test files were pseudo-randomly ordered – with a maximum of two consecutive trials of the same combination of names – and the side of presentation being counterbalanced for each combination of names. Moreover, within the first two test trials of the experiment, both test conditions were presented, that is, the test condition alternated. For half of the infants the first test trial was of the condition *IP*, followed by a *part-IP* test trial. For the other half of the infants, the first test trial was of the condition *part-IP* and was followed by the condition *IP* in the second test trial.

Based on the infant's head turns the listening time to each test trial was measured. The duration of each experimental session varied between around four and six minutes, depending on the infant's behavior.

7.1.4 Results

Infants (Group A: n=7, Group B: n=9) showed a mean listening time of 7.92 s (*SD* = 2.06 s) to test trials of the condition *IP* and a mean listening time of 7.25 s (*SD* = 2.37 s) to the condition *part-IP* (see Fig. 7.2). A repeated-measures ANOVA with the within-subject factor Test Condition (*IP* vs. *part-IP*) and the between-subject-factor Familiarization Group (A vs. B) revealed no significant main effect of Test Condition, *F*(1,14) = .735, *p* = .406 or Group, *F*(1,14) = .578, *p* = .460 and no significant interaction of Test Condition and Group, *F*(1,14) < .001, *p* > .99.

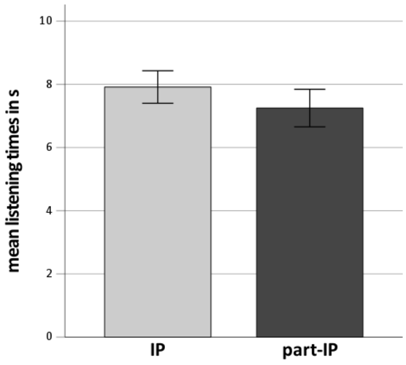

Figure 7.2 Mean listening times in Experiment III/1 to the test conditions IP and part-IP. Error bars indicate +/- 1 SE.

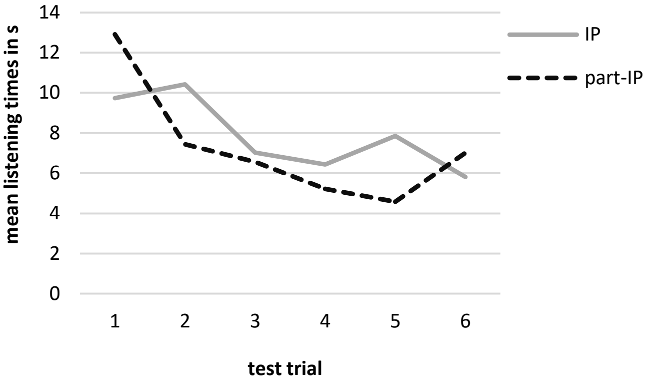

Figure 7.3 Mean listening times per test trial over the course of Experiment III/1.

Looking more thoroughly at this null result it was analyzed whether infants' listening preferences changed over the course of experiment. This was done by comparing the listening times obtained from each of the six test trials of one condition against those of the other condition (e.g., test trial 1 *(IP)* vs. test trial 1 *(part-IP)* etc.). Figure 7.3 shows generally decreasing listening times with a numerical preference for *IP* test sequences in test trials 2, 3, 4, and 5 and the opposite preference pattern in the first and last test trials. To analyze whether these preferences were of significance a further ANOVA including the within-subject-factor Test Trial was calculated. Here, only data from 14 infants were considered, because of missing data points in two infants due to inattentiveness in the last test trial (one listening time was missing for each). This ANOVA with the within-subject factors Test Condition and Test Trial and the between-subject factor Group revealed a significant main effect of Test Trial, $F(5,60) = 7.345$, $p < .001$ and a significant interaction of Test Trial and Test Condition, $F(5,60) = 2.408$, $p = .047$. No other significant main effects or interactions were significant (all p-values >.05). To explore the significant interaction of Test Trial and Test Condition further analyses were conducted. In the first test trial, listening times were especially high, but did not differ significantly between the test conditions, $t(15) = 1.723$, $p = .105$. When excluding the listening times of the first test trial – analyzing only the data from tests 2 to 6 in 14 infants (the 14 complete datasets), infants listened on average 7.48 s to the condition *IP* and 6.22 s to the condition *part-IP*. This difference was marginally significant, $t(13) = 1.890$, $p = .081$.

7.1.5 Discussion

At first glance, the data suggests that infants do not show any difference in their listening times between both combinations of names on the basis of the differing prosodic grouping during familiarization. This could mean that eight-month-old infants are not able to make use of the prosodic cues in the familiarization stimulus to build up an IP and to recognize the lexically corresponding unit during test. Given evidence from other studies that (even younger) infants are able to segment word sequences from complex utterances when prosodic boundary information is present (e.g., Nazzi et al., 2000; Seidl, 2007) this interpretation is rather unlikely to be true. Shortcomings in the materials may explain why infants did not significantly differentiate between test conditions: infants might not discriminate between both test sequences because they are able to recognize both sequences of names irrespective of the familiarized prosodic grouping, purely on the basis of the lexical material. This would mean that the test was too easy – infants merely needed to learn the six names that were presented during familiarization. The materials – two combinations of three disyllabic, trochaic names – were so simple that infants might have been able to memorize all of the six names and hence, do not make a difference between the test conditions simply because both combinations of names are sufficiently familiar. We know that German-learning infants at this age have a trochaic bias, that is, they are able to segment trochees from a speech stream and they prefer to listen to lists of trochees over lists of iambs (Höhle et al., 2009). Hence, infants in this experiment may have used rhythmic information for the segmentation and recognition of the names used in this experiment which enabled recognition of both test sequences likewise.

The more thorough view on the data that takes into account that listening preferences may change over the course of the experiment revealed an interesting finding: in the first test trial, listening times were generally very high compared to the following ones. This might indicate that infants realize the change from the familiarization phase to the test phase, that might simply mean, the change from a longer to a shorter stimulus. Hence, the comparison of the first *IP* versus *part-IP* test trial is somehow special. During this first trial, infants attended numerically longer to the *part-IP* test sequence which could be interpreted as a kind of novelty since the prosody of the *part-IP* test sequence is not congruent with the corresponding familiarization stimulus. In the following trials, infants seem to slightly prefer the test condition that is in prosodic congruence across test and familiarization stimuli as indicated by marginally significant listening preference for *IP* sequences over *part-IP* sequence. From this finding, one may cautiously interpret that infants are able to use prosodic boundary cues to segment and to recognize the first two names out of a three-name sequence when the two names form one IP and are separated from the third name by an IPB. Seeking a clear outcome that evidences infants' ability to use boundary cues for segmentation, the experimental design was modified in Experiment III/3. Prior to that, Experiment III/2 that was designed as a complement and run in parallel to Experiment III/1 will be presented.

7.2 Experiment III/2

Experiment III/2 also tested whether infants' recognition of two names out of a three-name sequence is affected by prosodic structure. In contrast to the previous experiment, here the last (and not the first) two names of this three-name familiarization sequence were tested. Hence, the position within the familiarization stimulus was different resulting in that none of the two test conditions was a familiar IP. The underlying assumption was that a prosodic grouping with an internal IPB between two names would hinder memorization and recognition of these two names.

Basically, the experimental design was similar to the one in Experiment III/1. Infants were split into two familiarization groups and were familiarized with the same speech files as infants in Experiment III/1: infants in Group A listened to the sequence [Moni und Lilli]$_{\text{IP}}$ [und Manu]$_{\text{IP}}$ of the prosodic type *With IPB* and the sequence [Leni und Manni und Lola]$_{\text{IP}}$ of the prosodic type *Without IPB*. Infants in Group B were familiarized with the same sequences of names in the opposite prosodic type, that is, [Moni und Lilli und Manu]$_{\text{IP}}$ as a sequence of the type *Without IPB* and [Leni und Manni]$_{\text{IP}}$ [und Lola]$_{\text{IP}}$ as a sequence of the type *With IPB*. Then, infants were tested on sequences that contained the last two names of the familiarization sequences, hence, “Lilli und Manu” and “Manni und Lola” (see Tab. 7.4). Depending on the familiarization, these sequences formed the test conditions *Spanning IPB* and *part-IP*.

The test condition *Spanning IPB* refers to the test sequence that lexically corresponds to the familiarization stimulus *With IPB* (see Tab. 7.4). For example, the sequence “Lilli and Manu” is the test condition *Spanning IPB* after familiarization with the stimulus [Moni und Lilli]$_{\text{IP1}}$ [und Mani]$_{\text{IP2}}$. In this familiarization stimulus, the last two names are separated from each other by an IPB. Infants are therefore expected to build up a mental representation of two prosodic units for this familiarization stimulus, hence, separating the last two names. Assuming that such a splitting caused by the IPB occurring within the familiarization sequence may impede memorization, recognition of this two-name sequence should be difficult.

The second test condition called *Part-IP* refers to the test sequence that lexically corresponds to the familiarization sequence of the prosodic type *Without IPB*. For example, “Manni und Lola” is the test condition *Part-IP* after familiarization with [Leni und Manni und Lola]$_{\text{IP1}}$. This familiarization stimulus should be represented as one single IP with the last two names being a part of it. Since there is no internal IPB, memorizing and recognizing these two names is assumed to be easier. If these assumptions are correct, listening times to *part-IP* and *Spanning IPB* test sequences should differ.

Table 7.4 Experimental design of Experiment III/2. Familiarization and test stimuli for Group A and B.

group	familiarization stimuli		test sequence	
	with IPB	without IPB	part-IP	spanning IPB
A	[Moni und **Lilli] [und Manu]**	[Leni und *Manni und Lola]*	*Manni und Lola*	**Lilli und Manu**
B	[Leni und **Manni] [und Lola]**	[Moni und *Lilli und Manu]*	*Lilli und Manu*	**Manni und Lola**

7.2.1 Participants

Fifteen eight-month-old infants (7 girls) were tested. The mean age was 8 months, 16 days (range: 8 months, 0 days to 8 months, 29 days). All infants were from monolingual German-speaking families, born full-term and normal-hearing. Ten additional infants were tested but their data were not included in the analysis for the following reasons: failure to complete the experiment (3), crying or fussiness (3), mean listening times of less than 3 seconds per condition (1), technical problems (1), experimenter error (2).

7.2.2 Stimuli

The familiarization stimuli and familiarization speech files were identical to those used in Experiment III/1 (see Fig. 7.1(A-D)). The test sequences, “Lilli und Manu” and “Manni und Lola”, were recorded by the same speaker under the same technical conditions with six repetitions of each sequence. Both test sequences were spoken in isolation and formed one single IP (see Fig. 7.4(A,B)). The test speech files were constructed in the same way as done for Experiment III/1. Half of the test files contained the sequence “Lilli und Manu”, the other half the sequence “Manni und Lola”. Each test speech file contained twice the six recordings of each test sequence (hence 12 per test file), randomly concatenated with an interstimulus interval of 800 ms. The total length of the test speech files “Lilli und Manu” was 22.45 s, and for the files “Manni und Lola” 22.39 s.

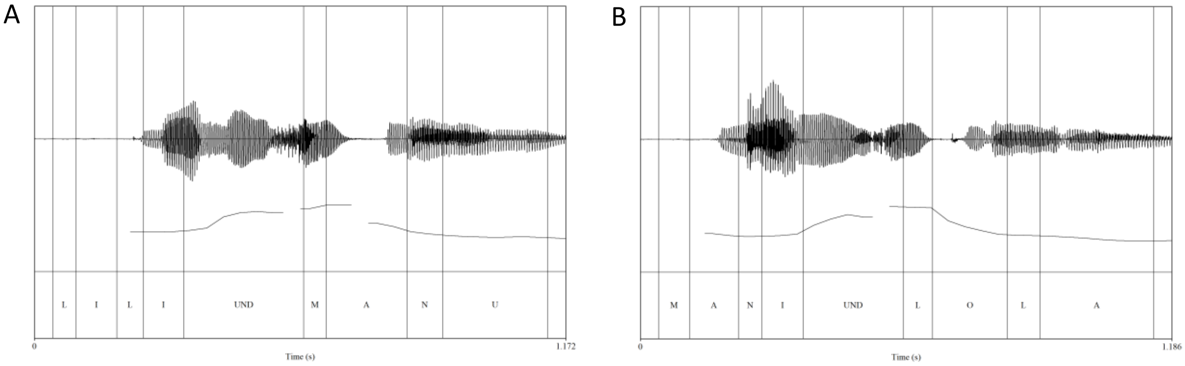

Figure 7.4 Oscillogram and pitch contour aligned to the text for Examples of the two test sequences, (A) ***Lilli and Manu*** **and (B)** ***Manni und Lola*****.**

7.2.3 Procedure

The experimental procedure was identical to the one in Experiment III/1: infants were assigned to one of the familiarization groups (A or B, see Tab. 7.4.). Familiarization was continuously played until the end of the familiarization speech file. Subsequently, infants passed through a test phase of twelve test trials, half containing the test sequence “Manni und Lola” and the other half “Leni und Manni”.

7.2.4 Results

Infants (Group A: n=8, Group B: n=7) showed a mean listening time of 7.07 s (*SD* = 2.21 s) to the test condition *part-IP*, and a listening time of 6.65 s (*SD* = 2.19 s) to the test condition *Spanning IPB* (see Fig. 7.5). A repeated-measures ANOVA with the within-subject factor Test Condition (*part-IP* vs. *Spanning IPB*) and the between-subject-factor Group (A vs. B) revealed no significant main effect of Test Condition, *F*(1,13) = .534, *p* = .478 or Group, *F*(1,13) = .899, *p* = .360, and no significant interaction of Test Condition and Group, *F*(1,13) = .326, *p* = .578.

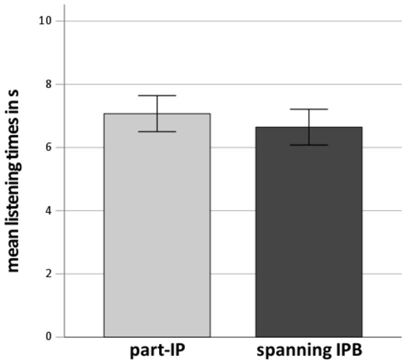

Figure 7.5 Mean listening times in Experiment III/2. Error bars indicate +/- 1 SE.

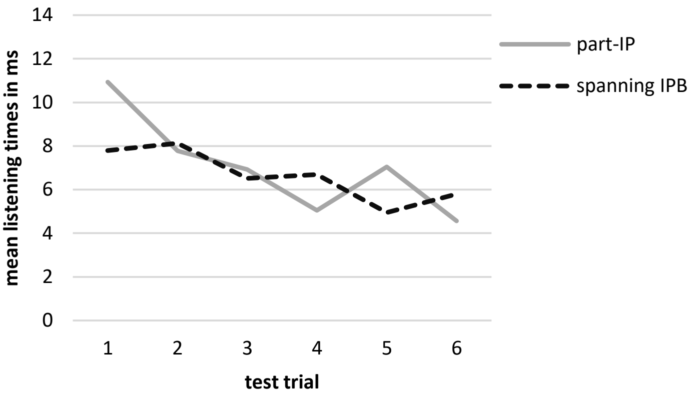

Figure 7.6 Mean listening times per test trial over the course of Experiment III/2.

Visual inspection of the listening times over the course of the experiments (see Fig. 7.6) revealed generally decreasing listening times with no systematic preference for one of the test conditions. While in the first and the fifth test trial the condition *part-IP* was preferred, in the fourth and sixth test trial there was a preference for the condition *Spanning IPB*, in the second and third test trials listening times were similar. In the ANOVA with Test Trial and Test Condition as within-subject factors and Group as between-subject factor the number of data sets decreased to n=14 due to one missing trial in one participant. This analysis did not reveal any significant main effects or interactions (all p-values >.05)

7.2.5 Discussion

As in Experiment III/1, also here, infants' listening time to the test conditions did not differ significantly. This may indicate that eight-month-old infants were not able to use the information on prosodic phrasing given in the familiarization stimulus for the recognition of a test sequence. However, shortcomings regarding the lexical as well as the prosodic characteristics of the materials may have hindered infants to show their ability to use prosodic boundary information.

As already discussed in Experiment III/1, infants might simply have learned all six names based on word learning strategies (such as the trochaic bias) and may not have been influenced by prosodic phrasing at all. Moreover, the differences in the familiarized prosodic status might not have been strong enough between the two test conditions. In the present experiment, one of the test conditions was familiarized as a sequence that is part of an IP and the other one as a sequence that spans an IPB. This means that none of the test conditions referred to a complete prosodic unit in the familiarization sequences. Under the assumption that complete prosodic units support recognition, there might have been not enough support from the familiarized prosodic structure to perform the task.

The previous experiment included such a test condition that referred to a complete IP in the familiarization stimuli. Indeed, this kind of test sequence might have been better recognized– as indicated by marginally longer listening times in the majority of the test trials compared to sequences that were familiarized as a part of an IP. However, recognition ability might turn out better, if the other test condition would not get any support for recognition from prosody. A contrast of *prosodic unit* versus *non-unit* might lead to clear recognition. This could be achieved through a modification of the prosodic types of the familiarization sequences both including an internal IPB resulting in test sequences that were familiarized as a full IP or as a non-IP spanning the internal IPB. With such a rather dichotomous contrast, infants' recognition of the first two names is assumed to show up.

7.3 Experiment III/3

In the third experiment of this study, eight-month-old German-learning infants were tested on their use of prosodic boundary cues for segmenting prosodically differently structured materials. Considering the shortcomings discussed for Experiments III/1 and III/2 some modifications of the materials were carried out concerning the prosodic types used in the familiarization stimuli: two combinations of three names were presented as sequences with an *early IPB* (after the 1st name, e.g., [Moni]$_{IP}$ [und Lilli und Manu]$_{IP}$) opposed to sequences with a *late IPB* (after the 2nd name, e.g., [Leni und Manni]$_{IP}$ [und Lola]$_{IP}$). In the case of the *early IPB*, the first two names were separated from each other by an IPB and hence, were expected to be represented as two different units. Hence, recognition of the first two names would not get support from prosody. The corresponding test condition (a sequence of the same first two names produced as one prosodic grouping) is called *non-unit*. In contrast, in the familiarization type *late IPB*, the first two names formed one complete IP. The representation of this IP built up during familiarization was assumed to support the recognition of the lexically corresponding two-name sequence during test. This test condition is called *unit*.

Another modification concerned the neutralization of the pitch contour in the test sequences. In the familiarization sequence with IPB, the pitch cue indicating the internal IPB was a rising pitch contour leading to a high boundary tone. In contrast, the pitch cue in the natural recordings of the test sequences was a falling pitch contour ending in a low boundary tone. The employment of diverging boundary tones might confuse infants' process of building up mental representations. Therefore, for this experiment, the pitch cue was neutralized in the test sequences.

Finally, using again the same lists of disyllabic names that were phonologically well-controlled, but lexically rather simple and hence, easy to learn, the familiarization duration was shortened to less than one minute to increase task difficulty.

7.3.1 Participants

Twenty eight-month-old infants (11 girls) were tested. The mean age was 8 months, 12 days (range: 8 months, 1 day to 9 months, 0 day). All infants were from monolingual German-speaking families, born full-term, and normal-hearing. Twelve additional infants were tested but their data were not included in the analysis for the following reasons: failure to complete the experiment (1), crying or fussiness (4), mean listening times of less than 3 seconds per condition (1), technical problems (4), experimenter error (1), familiarity with the names of the test sequence (1).

7.3.2 Stimuli

Familiarization sequences were the same combinations of names used in the previous experiments (“Moni und Lilli und Manu”; “Leni und Manni und Lola”) in two different prosodic types, either with an *early IPB*, e.g., [Moni]$_{IP1}$ [und Lilli und Manu]$_{IP2}$) or with a *late IPB*, e.g., [Moni und Lilli]$_{IP1}$ [und Manu]$_{IP2}$. The latter corresponds to the familiarization stimuli of the type W*ith IPB* of Experiments III/1 and III/2. Sequences of the type *early IPB* were recorded under the same technical conditions with the same young female speaker and scaled to 70dB. Using a crossed design, each combination of names occurred in both prosodic types across Group A and B (see Tab. 7.5). Examples of the prosodic properties of the familiarization sequences are depicted in Figure 7.7(A-D).

From these sequences, four speech files were created to be presented during familiarization trials. Each familiarization speech file consisted of 16 sequences in total, equally presenting both prosodic types *early* and *late IPB*. The prosodic type was alternating after a block of four recordings of the same sequence. Within a block of the same prosodic type, recordings were randomly concatenated with an interstimulus interval of 1 s. When the prosodic type changed, the interstimulus interval was 1.5 s.

Table 7.5 Experimental design of Experiment III/3. Familiarization and test stimuli for Group A and B.

group	familiarization sequence		test sequence	
	early IPB	late IPB	unit	non-unit
A	**[Leni] [und Manni** und Lola]	*[Moni und Lilli]* und Manu	*Moni und Lilli*	**Leni und Manni**
B	**[Moni] [und Lilli** und Manu]	*[Leni und Manni]* und Lola	*Leni und Manni*	**Moni und Lilli**

The beginning prosodic type was counterbalanced resulting in two different speech files for Group A and B, each. The total length of sound files, and hence the duration of familiarization, was 51.7 s for Group A and 52.8 s for Group B.

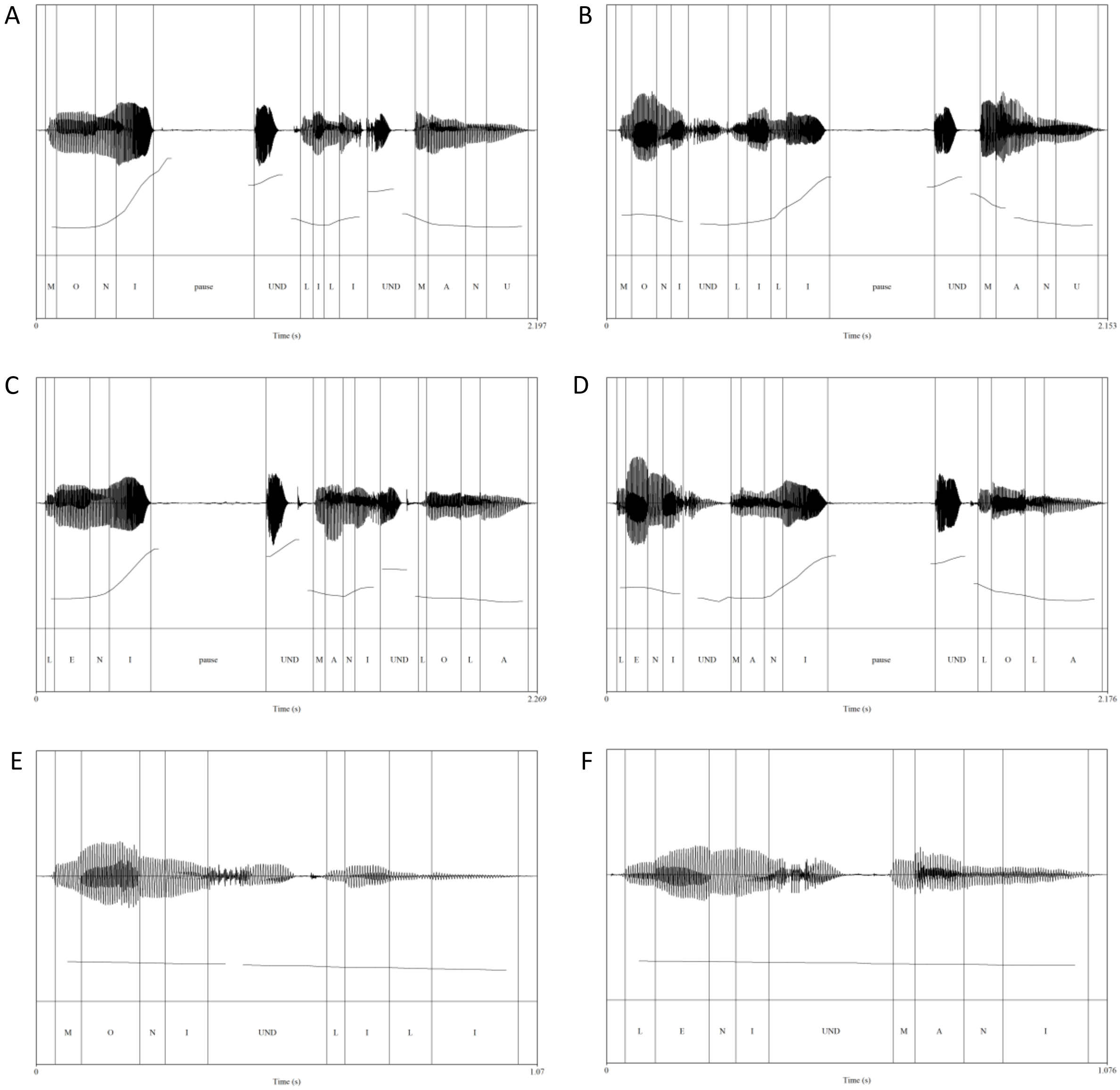

Figure 7.7 Oscillogram and pitch contour aligned to the text. (A) Familiarization sequence "Moni und Lilli und Manu" with early IPB (B) Familiarization sequence "Moni und Lilli und Manu" with late IPB (C) Familiarization sequence "Leni und Manni und Lola" with early IPB, (D) Familiarization sequence "Leni und Manni und Lola" with late IPB, (E) Test sequence "Moni und Lilli" with neutralized pitch contour, (F) Test sequence "Leni und Manni" with neutralized pitch contour.

The test stimuli were the first two names of each combination of names, "Moni und Lilli" and "Leni und Manni" (six repetitions each), hence the same test sequences as used in Experiment III/1 with one modification in the pitch contour: F0 was neutralized to rule out a potential interference from different acoustic correlates of the pitch boundary cue in the test sequence (a low boundary tone) compared to the pitch cue used in the familiarization sequence with IPB (a high boundary tone). Pitch manipulation and resynthesizing was done using PRAAT (Boersma & Weenink, 2010) and similarly to the procedure described in Seidl (2007). The pitch contour was reduced to the first and last pitch points. The first pitch point was set to the mean F0 of each test sequence, the last pitch point was not changed. This way of pitch neutralization was preferred over a total flattening of the pitch contour to result in a more natural sounding. Indeed, test sequences sounded quite monotonous, but not robotic and did no longer reveal a low boundary tone. Examples of the test sequences are depicted in Figure 7.7(E,F).

The test sequences with neutralized pitch served to create test speech files to be used during the twelve test trials. There were six files that contained the sequence “Moni und Lilli”, and six with “Leni und Manni”. Each file contained 12 repetitions of the same combination of names (the six recordings were repeated twice), randomly concatenated with an interstimulus interval of 800 ms. The total length of the test speech files containing “Moni und Lilli” was 20.88s, and for the files containing “Leni und Manni” 21.66 s.

7.3.3 Procedure

The experimental procedure was identical to the one in Experiment III/1. Infants were randomly assigned to one of the familiarization groups and were familiarized for the total length of a familiarization speech file, about 52 s, without interruptions of the sound. Subsequently, infants passed twelve test trials.

7.3.4 Results

Infants (Group A: n=11, Group B: n=9) showed a mean listening time of 7.15 s (*SD* = 2.37 s) to test trials of the condition *unit* and a listening time of 6.92 s (*SD* = 2.15 s) to the condition *non-unit* (see Fig. 7.8). A repeated-measures ANOVA with the within-subject factor Test Condition (*unit* vs. *non-unit*) and the between-subject-factor Group (A vs. B) revealed no significant main effect of Test Condition, *F*(1,18) = .504, *p* = .487, or Group, *F*(1,18) = .038, *p* = .849 and no significant interaction of Test Condition and Group, *F*(1,18) = 1.445, *p* = .243.

Visual inspection of the listening behavior over the course of the experiment revealed generally decreasing listening times and no clear preferences for one of the test conditions (longer listening times for *unit* test sequences in test trials 1,2,3; similar listening times in test trial 5 and longer listening

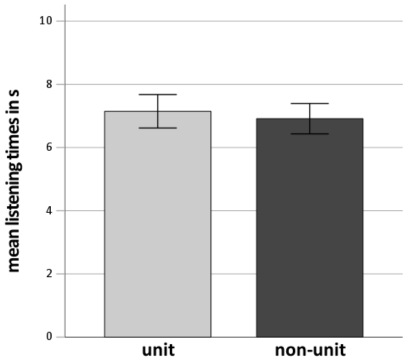

Figure 7.8 Mean listening times in Experiment III/3. Error bars indicate +/- 1 SE.

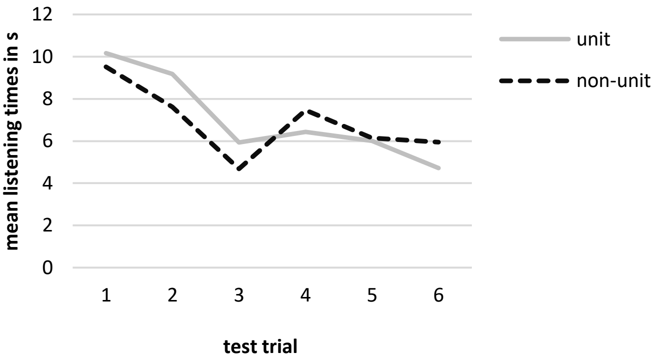

Figure 7.9 Mean listening times per test trial over the course of Experiment III/3.

times for *non-unit* test sequences in test trials 4 and 6). An ANOVA with the within-subject factors Test Condition and Test Trial and the between-subject factor Group revealed a significant main effect of Test Trial, $F(5,85) = 9.237$, $p < .001$, but no other significant main effects or interactions (all p-values >.05).

7.3.5 Discussion

Experiment III/3 tested whether infants' recognition of a sequence of two names depends on the prosodic status in its lexically corresponding familiarization sequence. It was hypothesized that recognition would be easier if the sequence was familiarized as a complete prosodic unit, namely an IP, whereas recognition might not be successful, if the sequence spanned an IPB in the familiarization stimulus. Listening times did not differ between test conditions. Hence, infants seemed not to differentiate between both sequences of names although the prosodic contrast, being a unit versus a non-unit in the familiarized sequences, was stronger in this experiment compared to the two previous ones.

An analysis that investigated listening differences over the course of the experiment did not reveal significant preferences neither. The main effect of Test Trial was presumably driven by the generally decreasing listening times due to infants' decreasing attention.

7.4 General discussion

To sum up, Study III aimed at finding a procedure to provide evidence that infants are not only able to perceive different constellations of prosodic boundary cues – as Studies I and II showed – but that infants also make use of prosodic boundary cues for segmentation. All three experiments used a recognition task that slightly differed with respect to the prosodic phrasing of the familiarization stimuli. While the materials were highly suitable for Studies I and II, for the purpose here, they turned out unsuccessful. Infants did not differentiate the test conditions suggesting that the prosodic grouping of the familiarization stimulus did not play a role in this task. Infants seemed to recognize both types of names.

Previous studies suggest that infants' language processing benefits from prosodic information. For example, Mandel et al. (1994) showed that two-month-old infants are able to detect subtle phonetic differences but only in materials that were presented as a prosodic unit and not in a list of isolated words. In contrast to the study by Mandel et al. (1994), infants here were much older, but the task – recognizing a sequence of two names – was not especially difficult and might be solved without the use of phrasal prosody at this age. Word learning strategies may have come into play and enabled recognition of all the names of the test conditions. Hence, the procedure used here might be effective in much younger infants, at a point in language acquisition when segmentation based on purely rhythmic cues is not yet possible, that is, at about four months (Höhle et al., 2009). However, this is not the stage in acquisition to assume that prosodic boundary cue perception has already undergone perceptual attunement: a language-specific cue development of prosodic boundary perception has been found to develop in German only from six to eight months (Studies I and II, van Ommen et al., 2020). Therefore future research on the use of prosodic boundary cues should focus on an experimental design that works with this age group.

One option for varying the experimental design would be including syntactically and lexically more complex materials: First, complex sentences from which clauses need to be segmented could be used (similar to the studies by Nazzi et al., 2000). Second, the lexical materials within these sentences should vary with respect to lexical rhythm. Importantly, they may not only include disyllabic trochees, but also words with one, three or more syllables (for example pseudo names like "Manuli, Lemanu, …"). Considering these two aspects, infants would presumably not be able to perform successfully simply by using their word segmentation strategies and memory capacities. They would need to take the prosodic phrase boundary information into account. Nevertheless, the materials need to be controlled regarding their phonology. The stimuli used throughout this dissertation had the advantage that their phonological structure was perfectly suited for acoustic analysis and manipulation. The names were of the structure CVCV with sonorant consonants und full vowels and did not contain vowels reduced to Schwa. The second vowel of a name always was /i/ before the conjunction "und" because the vowels /i/ and /u/ are highly differentiable in the spectrogram. This allowed for a clear setting of the segmental boundaries at the possible IP boundary position and hence, a most exact duration measurement of final lengthening. Moreover, with all consonants being sonorants pitch measurements could be carried out without interruptions of the F0 curve as would be caused by for example, voiceless plosives. When boundary cues can be clearly measured, a more thorough acoustic manipulation is possible resulting in most naturally sounding stimuli. Taken together, by using natural sentences with lexically varying materials the segmentation task is assumed to be more age-appropriate and also more similar to the one that infants face in natural language acquisition. Together with a control for segmental phonology within these sentences, materials will be best suited to investigate the role of each prosodic boundary cue in infant syntactic segmentation.

8 The broader picture

Starting from the prosodic bootstrapping theory, many studies in the last four decades have revealed insights into infants’ sensitivity to and use of prosodic information to derive knowledge about other linguistic domains. Prosodic phrasing in relation to clause segmentation has been widely studied from this perspective. Moreover, a great deal of research has been conducted on language-specific speech perception in early language acquisition. This has mainly focused on infants’ sensitivity to phoneme and tonal contrasts as well as word stress. This dissertation has combined both aspects of research, infants’ sensitivity to phrasal prosody and language-specificity in early speech perception, by investigating the role of individual prosodic cues in infant boundary processing in the course of the first year of life.

8.1 Prosodic boundary cue perception in the early acquisition of German: Summary of the main findings

Let us begin drawing the broader picture by summarizing the main findings of Study I and Study II. This will answer the first main question raised in the Introduction: **What role do the acoustic correlates of prosodic boundaries play in boundary perception in early language acquisition? Specifically, which of the individual acoustic features are necessary to trigger boundary perception in German-learning infants and is there a change in infants' sensitivity during the first year of life?**

First, the boundary present in the materials used throughout this dissertation was clearly acoustically signaled. The materials, lists of three coordinated names, were produced in two differently bracketed groupings. While the grouping [N1 and N2 and N3] revealed no internal major prosodic boundary between the names, the other grouping [N1 and N2] [and N3] showed evidence of an internal IPB after the second name. The acoustic characteristics of this boundary comprised the three typical boundary cues, pitch change, final lengthening, and pause. The pitch cue was a rising pitch contour leading to a high boundary tone on the second name followed by a pitch reset at the subsequent conjunction. Final lengthening was expressed by an almost doubled duration of the preboundary vowel compared to the same vowel in the sequence without an internal boundary. The pause had an average duration of about 500 ms. The materials were thus suitable for testing the impact of prosodic boundary cues in infant perception. Infants were faced in several experiments with a familiarization-preference task that required the discrimination of a sequence without a boundary from a sequence with a boundary in materials in which the occurrence and constellation of the three prosodic boundary cues were varied. The following conclusions can be drawn.

The presence of all three cues leads to successful boundary perception in six- and eight-month-old German-learning infants. At six months, two out of the three main prosodic boundary cues are sufficient to detect a prosodic phrase boundary as long as one of these cues is the pause cue. At eight months, a subset of two cues is also sufficient. However, within this subset, pause is no longer a necessary cue; pitch and lengthening in combination provide sufficient information for detecting a prosodic phrase boundary although perceiving a boundary in the absence of a pause is more demanding. Still, similar to what has been observed for adult speakers of German with the same materials (Holzgrefe-Lang et al., 2016), the occurrence of pitch or lengthening as single cues is not sufficient for six- and eight-month-olds. **In sum, Studies I and II suggest a developmental change in German prosodic boundary cue perception from six to eight months. At an initial stage of prosodic boundary perception, infants strongly rely on the pause cue. With increasing exposure to the German language, infants sharpen their sensitivity to the combination of pitch change and final lengthening, and hereby reach adult-like prosodic phrasing.**

Hence, pause develops from a necessary to an optional boundary marker. However, note that not being necessary does not imply perceptual disregard. Evidence from machine learning indicates that whenever a prosodic cue is present, it is likely that the perceptual system takes it into account (Ludusan et al., 2016).

The strong perceptional role of the pause cue can be linked to prosody enhancement in IDS. Fernald and Simon (1984) reported higher pause duration and a higher correspondence of pauses and sentence boundaries, thus a higher reliability of the pause cue, in German IDS compared to ADS. Ludusan et al. (2016, see Chapter 2.2.2) found consistently larger pause durations in IDS than in ADS in their literature review and in their analysis of the RIKEN corpus of Japanese IDS and ADS. Moreover, in a machine learning experiment, they found the learning algorithm better at detecting a boundary in IDS than in ADS when the combination of boundary cues contained a pause (of at least 300 ms). Hence, across all parts of Ludusan et al.'s study, pause proved to be ranked the most important boundary cue. Pauses seem to be an important cue in IDS, but with increasing age infants must learn to cope without pause: in their corpus data, Ludusan et al. (2016) observed a non-significant linear trend for pause duration at the boundary to decrease across the studied ages from 18 to 24 months. Since pause is not a reliable cue in German ADS (Peters et al., 2005) and not at minor boundaries in IDS (Soderstrom et al., 2008), infants need to refine their sensitivity to boundary cues. Attuning to pitch and duration prepares German-learning infants for segmenting the speech signal in the absence of pauses.

8.2 Perceptual attunement in prosodic boundary cue perception: Evidence across languages

Although early language acquisition can be universally regarded as a rather stable process that forms proficient users of a language after an incredibly short period of time, there is some individual variation in performance in infancy that may result from four factors that Cristià (2009) lists in her thesis: first, the linguistic input, that is, the learner's language environment; second, sensory, especially auditory, acuity determining how fine-grained distinctions in the auditory signal can be perceived; third, domain-general skills such as an infant's attention, memory, and processing speed; and fourth, the presence of innate linguistic categories.

The research conducted in this dissertation in relation to studies from other languages sheds light on the first factor, the role of the linguistic input, by evaluating whether and how infant processing mechanisms adapt to the prosodic properties of the native language in prosodic boundary cue perception. This was spelled out in the Introduction as the second main question of this dissertation: **Is the perception of phrasal prosody a matter of perceptual attunement within the first year of life?** To answer this question, in the following I will link the results from Studies I and II to those from studies on other languages.

A limited number of studies have investigated the impact of individual or combined prosodic boundary cues on infant boundary perception: prior to Study I of this thesis, there existed three studies targeting the monolingual acquisition of American English and Dutch (Johnson & Seidl, 2008; Seidl, 2007; Seidl & Cristià, 2008). During the course of this dissertation, further studies reporting on French and European Portuguese were added (Frota et al., 2019; van Ommen et al., 2020). Moreover, slightly different from the abovementioned studies (participants were much older and the task was more complex), two studies with American-English- and Japanese-learning toddlers (Hawthorne & Gerken, 2014; Hawthorne, Mazuka, & Gerken, 2015) investigated the exploitation of the full set of boundary cues for learning syntactic constituents (see Chapter 2.4.1).

All these studies used behavioral methods applying a familiarization-preference procedure, implemented in the HPP or with eye-tracking. Moreover, all these studies presented stimuli in two prosodically different conditions, without an internal prosodic boundary (also referred to as the clausal/well-formed/non-straddling condition) and with an internal boundary (also called the non-clausal/ill-formed/straddling condition). Stimuli differed with respect to the kind and number of prosodic boundary cues that were presented, either testing the influence of all three cues together, of a subset of two cues, of single cues, or of all possible constellations. Table 8.1 gives an overview of the studies' characteristics and the outcomes as indicated by infants' listening or looking preferences.

Table 8.1 Summary of the preferences in listening/looking time observed in behavioral studies on prosodic boundary cue perception. AE = American English, EP = European Portuguese, "–" = this cue or constellation of cues was not tested.

Study	**Characteristics of the study**					**Tested constellation of boundary cue(s) and infants' preferences**					
	Method	**Stimuli**	**Tested language**	**Native language**	**Age**	**pitch lengthening pause**	**pitch lengthening**	**pause lengthening**	**pitch pause**	**lengthening**	**pitch**
Seidl & Cristià (2008)	HPP segmentation	clauses, text passages	AE	AE	4mos	non-clause	none	none	none	-	-
Seidl (2007)					6mos	clause	clause	none	clause/ non-clause[17]	none	none
Johnson & Seidl (2008)			Dutch	Dutch	6mos	clause	none	-	-	-	-
Wellmann et al. (2022)	HPP, discrimination	list of names	German	German	4mos	none	-	-	-	-	-
				German	6mos	novelty	none	novelty	-	-	-
Wellmann et al. (2012)				German	8mos	novelty	familiarity	-	-	none	none
van Ommen et al. (2020)	HPP, discrimination	list of names	French	German	6mos	-	none	-	-	-	-
				German	8mos	-	familiarity	-	-	-	-
				French	6mos	novelty	none	-	-	-	-
				French	8mos	familiarity	none	-	-	-	-
Frota et al. (2019)	eye-tracking, discrimination	delexicalized clauses	EP	EP	9mos	-	familiarity	-	-	-	-
Hawthorne & Gerken (2014)	HPP, constituency learning	non-words, artificial grammar	AE	AE	19mos	familiarity	-	-	-	-	-
Hawthorne, Mazuka, & Gerken (2015)			AE	Japanese	19mos	novelty	-	-	-	-	-
			Japanese	AE	19mos	familiarity	-	-	-	-	-

---

[17] Depending on the procedure for neutralizing final lengthening.

What becomes obvious when comparing infants' preferences across studies is the special role of the pause cue. The pattern in which the impact of the pause as a boundary cue decreases during the first year of life is not unique to German, but is also reflected in the developmental trajectory of infant boundary processing in American English, though somewhat earlier. Moreover, an initial necessity of the pause cue has also been observed for Dutch and French infants; while for the latter group this necessity persists at least until the age of eight months, but not until adulthood, for Dutch, we do not know whether the sensitivity to prosodic boundary cues changes since data from older Dutch infants are lacking.[18]

The most recent study tested European-Portuguese-learning infants aged eight to ten months in discriminating short sentences. These sentences were either of SVO structure, forming one IP, or of a topicalized structure making up two IPs with a sentence-internal IPB marked by final lengthening and pitch change, but no pause. Each sentence type was delexicalized by replacing all vowels by [ɐ] and all consonants by [n] or [ʃ], maintaining the prosodic structure. After familiarization to one of the prosodic conditions, infants displayed a novelty effect. Hence, at about nine months, boundary perception in European Portuguese does not require the pause cue, and is thus already adult-like since adult listeners were shown to perceive IPBs without the pause cue, but signaled by pitch and lengthening (Severino, 2016). Since younger infants were not tested, it remains unclear whether this pattern is the result of a development or is present from the beginning on.

Comparisons across studies with respect to language-specificity need to be taken with caution since differences can also stem from variations in the experimental task and/or the materials used. Thus the phonetic magnitude of the prosodic boundary cues differed between stimuli; for example in Seidl's (2007) English materials the pause at the clause boundary lasted 300/380 ms, whereas in the Dutch materials it was much larger, 630/820 ms. Another main difference also concerns the acoustic manipulation of the prosodic boundary cues: in the Dutch and English materials, cues were neutralized to test the impact of the remaining cues, whereas in the German and French materials the cue(s) under investigation were added to the materials. This implies that the Dutch and English study tested the impact of pitch and lengthening along with a present pause (though ambiguous because the duration was the same in both prosodic conditions), whereas in the studies on German, French, and European Portuguese no pause was present when testing the influence of pitch change and final lengthening. Further research is needed to clarify whether such a difference in cue implementation matters for the perceptual system.

---

[18] Note that Schmitz (2008) hypothesizes that the missing effect in the experiment of pitch and lengthening with Dutch infants (Johnson & Seidl, 2008, Experiment 2) reflects a rejection of unnatural stimuli because pause neutralization was obtained by inserting a pause of 738 ms (the mean of the pauses in the two ill-formed sequences) within the Dutch clause. This could have been too long in relation to the pause of 1s that occurred between sentences. Thus, Dutch infants may have rejected this kind of material because it did not conform to natural speech and not because they needed the pause cue. Hence, given this argument, it is still open whether Dutch six-month-olds can recognize a boundary without the pause cue.

In this regard, three of the listed studies are of special value because they tested learners from different language environments with the exact same materials: these are the study by van Ommen et al. (2020, testing German and French learners on French stimuli) and the studies by Hawthorne and Gerken (2014) and Hawthorne et al. (2015, testing English and Japanese infants on English stimuli). Hence, similarities or differences in the outcomes could be clearly attributed to the influence of linguistic experience. These studies are also indicative of whether native prosodic skills are language-specific or can be transferred to non-native prosody: German-learning infants were tested on similarly structured stimuli in German (Wellmann et al., 2012) and in French (van Ommen et al., 2020), and English-learning infants were tested in English (Hawthorne & Gerken, 2014) and in Japanese (Hawthorne et al., 2015). Since the studies by Hawthorne and colleagues tested older children and presented the full set of naturally occurring boundary cues, that is, no reduced boundary markings without the pause cue were tested, they do not directly contribute to resolving the question of perceptual attunement within the first year of life like all the other studies listed in Table 8.1. They are thus excluded here, although I will return later to Hawthorne et al.'s results when elaborating on the underlying mechanism of prosodic boundary perception (see Chapter 8.3).

Van Ommen et al. (2020) discovered that eight-month-old German-learning, but not French-learning, infants perceive a pitch-lengthening cued boundary, that is, without the pause cue, when listening to French stimuli with or without an internal boundary (e.g., [Loulou et Manu][et Nina] or [Loulou et Manou et Nina]). French infants exhibited the same pattern at six and eight months: they discriminated sequences only with fully cued boundaries. In contrast, German infants' perception underwent a developmental change with respect to the necessity of the pause cue – similar to the outcome with German materials (Wellmann et al., 2012). In adults who were also tested by van Ommen et al. (2022) a difference between French and German listeners was no longer observable. This suggests a language-specific development of the perception of prosodic boundary information in infancy: French infants are slower in developing a sensitivity to the cue combination of only pitch and lengthening. They seem to stick longer to an initial language-general way of boundary processing that requires the pause cue. Since French materials were used across both language groups, this study also revealed that German-learning infants are sensitive to non-native prosody, presumably by drawing similarities to the prosodic structure of the native language. While Seidl and colleagues have proposed a language-specific development in prosodic boundary cue perception based on comparisons of results from different studies, van Ommen et al.'s study provides direct evidence for this view on methodologically proper grounds, based on comparisons of the performance of learners from distinct language environments in the same test language.

Taken together, based on the results from Studies I and II, I draw the conclusion that prosodic boundary perception undergoes perceptual attunement. Differently to the typical pattern most

often reported for other phonological domains, this attunement is not reflected in a decline in sensitivity, that is, infants do not lose their sensitivity to one of the boundary cues. Rather, they expand their sensitivity to the available boundary cues and no longer require each individual cue. Hence, perceptual attunement in prosodic boundary cue perception is a process of enhancement, specifically from an initially strong reliance on the acoustically salient pause cue to a more sophisticated sensitivity to individual and presumably more subtle prosodic cues, such as pitch change and final lengthening.

Regarding the languages investigated so far, the language-specificity in this development is primarily demonstrated as one of age, that is, infants respond to the same cues at different ages: whereas the impact of the pause cue has already declined by six months in the acquisition of American English, for German this does not happen until eight months and for European Portuguese by nine months (whether the pause is a necessary cue before that age is not yet known). For Dutch and French, no longer relying on the pause cue presumably occurs after the age of six and eight months, respectively. In addition, language-specificity is also reflected in a different weighting of individual cues (for example English infants needing the pitch but not the pause cue at a certain age or German infants not requiring the pitch cue but needing the pause cue). The language-specific differences in sensitivity to these cues raise the question of the origin of this variation. In the following, I suggest that these differences relate to the specific distribution of pitch und durational cues in the language that is being acquired.

Firstly, prosodic boundary marking seems to be a rather universal phenomenon with the same acoustic features indicating the presence of a boundary across typologically related and unrelated languages (Vaissière, 1983). Jun (2014: 522) points out that “the IP or ip[19] tonal rhythm tends to vary more within a language (due to the variable size of IP or ip) and vary less across languages (for both size and the type of boundary tone).” Nevertheless there is some crosslinguistic variation, in production (Vaissière, 1983) as well as perception, as evidenced here for infant speech perception. How do the tested languages differ in prosody?

English, Dutch, and German are all stress-timed languages. For stress languages, an early sensitivity to word stress mainly cued by pitch, intensity, and duration with a preference for trochees has been demonstrated (Höhle et al., 2009; Jusczyk, Cutler, & Redanz, 1993; Keij & Kager, 2016), whereas infants learning a language without lexical stress, such as French, exhibit no preference for an iambic or trochaic stress pattern (Höhle et al., 2009). French is a syllable-timed language, with fixed, thus highly predictable, phrase-final stress, which is typically realized as a “rising pitch accent marking simultaneously the head and the edge of an AP” (Jun, 2014: 531). This means that the acoustic marking

---

[19] = intermediate phrase as assumed in the Autosegmental-Metrical model of Intonational Phonology. Jun (1998) explains that in English the intermediate phrase corresponds to the Phonological Phrase in the Prosodic Hierarchy of Nespor & Vogel (1986).

of pitch accent and phrase boundary often falls together, which reduces the overall prosodic variability compared to languages that have variable stress positions.

Like French, European Portuguese is a Romance language, which are usually syllable-timed, but among these, European Portuguese has a special prosodic status since it has a mixed rhythmic profile: a high variation in the intervals between consonants renders it stress-timed, whereas a high proportion of vowels characterizes it as syllable-timed (for a detailed analysis, see Frota & Vigário, 2001). Moreover, in contrast to French, but similar to German, Dutch, and English, European Portuguese has lexical stress along with variable stress positions. However, different to English, in European Portuguese, most stressed syllables are not pitch accented, but mainly marked by duration (and vowel reduction). As for German, Dutch, and English, an early sensitivity to word stress patterns has been shown with European-Portuguese-learning infants (Frota, Butler, Uysal, Severino, & Vigário, 2020).

Hence, there are at least two factors, the presence of lexical stress and predictability of the phrasal stress position, that render German, Dutch, English, and European Portuguese prosodically more variable than French. A high *prosodic variability*, that is, a frequent and flexible occurrence of F0 and durational properties, implying less predictability, might increase infants' attention to prosodic details simply due to their high acoustic salience in the speech signal, similar to the facilitative effect of motherese on language development. Moreover, prosodic variability causes a stronger need to attend to specific acoustic correlates to recognize and disentangle the intended linguistic functions in contrast to prosodic regularity. Thus, increased attention as well as linguistic need might improve infants' overall sensitivity to prosodic details, such as pitch and duration – that is, also when they are used as cues to prosodic phrasing. The earlier ability to cope with pitch-lengthening cued boundaries – as observed with German- and English-learning infants compared to French ones – may thus reflect the exposure to a prosodically more variable language environment (and, vice versa, prosodic regularity as in French might lead to less prosodic sensitivity).

Going beyond word stress and phrasal stress positions, van Ommen et al. (2020) consider Jun's (2014) revised model of *Prosodic Typology* to link a language's prosodic properties, specifically its *macro-rhythm*, to language-specific abilities in the perception of phrase boundaries. In this model, Jun (2014) introduces *macro-rhythm* in addition to the parameters *prominence* and *word prosody* (stress, tone/lexical pitch accent, both, none) to classify languages. Traditionally, the term *rhythm* (also called *micro-rhythm*) refers to the duration-based temporal organization of smaller units (such as syllables and feet). In contrast, *macro-rhythm* refers to the global tonal rhythm, that is, the rhythm created by pitch movements at the phrasal level as conveyed by pitch accents, lexical tones, or boundary tones. The degree of macro-rhythm can be strong, medium, or weak and is determined by three factors: 1) the number of possible tonal patterns in the specific language, that is, the more types of tonal patterns, the more variable the pitch contour and the weaker the macro-rhythm; 2) the type of the most common phrase-medial tonal pattern, that is, when the most common tonal pattern is a complex one (i.e., LH or HL), the macro-rhythm is stronger; conversely, when the most common tonal pattern is a level one (i.e., L or H), the macro-rhythm is weaker; 3) the frequency of tonal patterns, that is, the more often tonal patterns occur (i.e., a pitch accent on every content word), the stronger the macro-rhythm. To sum up, a pitch contour has a strong macro-rhythm if low and high tones alternate at regular intervals and if the tonal units are similar to each other. A strong macro-rhythm is thus associated with high regularity.

According to these parameters, Jun (2014) classifies Dutch, German, and English as having a medium macro-rhythm since 1) they have several pitch accent types, 2) the most common one is a level tone (H*), and 3) pitch accents occur at almost every content word. For European Portuguese, the first two parameters are the same, but 3) pitch accents occur less frequently, which renders it weakly macro-rhythmic. French is classified as a language with strong macro-rhythm, even though the domain of the pitch accent is equal to or larger than a content word – similar to English, Dutch, and German – but differently, it has few types of pitch accents und the most common one is a rising pattern.

Following this classification, van Ommen et al. (2020) argue that French infants' failure to perceive a pitch-lengthening cued boundary without the pause cue reflects a less developed sensitivity to prosodic details due to the strong macro-rhythm and lack of word stress in French – which causes a highly regular and predictable prosodic phrase structure and thus less need to pay attention to prosodic details. In contrast, German with lexical stress and medium macro-rhythm (like English and Dutch) is prosodically more variable and less predictable and thus requires the learner to attend to prosodic details to detect the prosodic structure. This may explain the earlier development in prosodic boundary cue perception. This explanation is also in line with infants' boundary discrimination in European Portuguese, a language with stress and weak macro-rhythm.

8.3 Beyond the crosslinguistic evidence: Implications for the prosodic bootstrapping theory

The previous section presented crucial differences in how languages make use of pitch and duration for prosodic structuring of words and phrases (as expressed in lexical and phrasal stress, stress positions, and macro-rhythm) that are likely to account for the language-specific performance in the available data on prosodic boundary cue perception. However, these data also reveal that infants do not apply language-specific strategies from the beginning on. Rather, the investigated languages have one crucial similarity: the beneficial impact of the pause cue at early stages of speech perception. This lets us return to the third main question raised in the introduction: **Does infants' initial processing of boundary cues more likely relate to a universal processing mechanism or do infants apply language-specific strategies from the beginning on?** In the following, **I will argue that infants' initial processing**

of boundary cues reflects a language-general processing mechanism that primarily focuses on acoustic saliency, while only a few months later infants' sensitivity has adapted to the language-specific prosodic properties.

Hawthorne and colleagues hypothesized that prosody's supportive role in syntax acquisition is just driven by acoustic salience. Seeking the origins of prosodic bootstrapping, in their crosslinguistic studies on constituency learning (Hawthorne & Gerken, 2014; Hawthorne et al., 2015), they consider two hypotheses on the underlying mechanism of prosodic bootstrapping, the *acoustic salience hypothesis* and the *language-specific hypothesis*. The *acoustic salience hypothesis* assumes that prosody is useful simply because the cues in the speech signal are acoustically salient. In this view, cues given by pitch, duration, and amplitude are acoustic information that the mammalian auditory system can easily perceive without any specific prior knowledge of the native language. Hawthorne et al. refer to studies showing that human newborns as well as rats and cotton-top tamarins are able to distinguish languages from different rhythm classes. Here, prosody is linked to general auditory processing. In contrast, the *language-specific hypothesis* assumes that the knowledge that the learner acquires through exposure to her native language is the important basis for using prosodic cues for segmentation. Here, prosody is linked to language processing.

In an artificial grammar study, Hawthorne et al. (2015) tested toddlers' use of non-native prosody for syntactic segmentation and constituency learning. Following Hawthorne and Gerken (2014, see Chapter 2.4.1) they familiarized 19-month-old Japanese-learning toddlers with non-word sentences in English prosody. These were the same strings that had previously been used with English-learning toddlers (Hawthorne & Gerken, 2014). During testing, Japanese learners, similar to English ones, distinguished between the grammatical and ungrammatical re-orderings of the familiarization string (grammatical items preserved the prosodic groupings from the familiarization string, whereas ungrammatical items split the groupings that the familiarization had indicated). Conversely, presented with non-word sentences in Japanese prosody, English learners successfully recognized the grammatical constituents. Hence, even when encountering non-native prosody, toddlers exploited prosodic cues to learn constituents. This was possible even though the native language and the language being tested were prosodically very distinct. Hawthorne et al. conclude that it is possible to generalize the native knowledge on prosodic phrasing to a very distinct new prosodic system (as also evidenced for German infants' boundary perception with French stimuli in van Ommen et al., 2020). Hawthorne et al. (2015) take the application of native prosodic skills in a non-native language as support for the acoustic salience hypothesis of prosodic bootstrapping.

However, I would argue that the transfer of native prosodic knowledge may depend on the occurrence of a pause cue at the boundary. Note that Hawthorne et al. only tested the impact of the full naturally occurring set of boundary cues, that is, pitch, final lengthening, and the acoustically salient pause cue.

With this cue constellation, one should be careful in drawing conclusions on (a lack of) language-specificity since prior infant studies demonstrated pause to be a language-general cue. The most indicative condition in studying boundary perception across languages was a pitch-lengthening cued boundary. Presumably, constituency learning would be more difficult if stimuli contained no pause, but only non-native pitch and final lengthening cues. This has not yet been investigated. Hence, future research should examine whether Japanese infants are able to perceive a prosodic boundary cued by pitch and lengthening only, and if so, whether Japanese toddlers can exploit this kind of boundary for deriving constituents in the native language as well as in a non-native language. Japanese has a strong macro-rhythm (similar to French). Moreover, compared to English, Japanese IDS has been found to have less final lengthening (Fisher & Tokura, 1996). Given these properties, I predict that Japanese-learning infants would fail (similar to French ones) with a prosodic boundary marking that is missing a pause, in both their native and a non-native language.

Moreover, given that we have observed that non-native prosody can in principal be processed (van Ommen et al., 2020) and that English-learning infants do not require the pause cue already by six months (Seidl, 2007), I further expect that English infants would also manage with a pitch-lengthening cued boundary in Japanese stimuli – provided that the pitch and lengthening cues lie in certain ranges of phonetic magnitude that are similar to those in English. If successful performance without the pause was the result in English-learning infants along with a failure in Japanese-learning infants, this would display another case of language-specific differences in prosodic boundary perception (similar to the German–French case). This example illustrates how important and necessary it is to take into account the specific marking of a prosodic boundary. Studying only the constellation of the three boundary cues in combination might point to an auditory processing of prosodic information and thus to the acoustic salience hypothesis. In contrast, testing varying cue constellations, for example with and without the pause cue, might reveal a language-specific processing as assumed in the language-specific hypothesis of prosodic bootstrapping. **The varying perceptual impacts of pitch change, final lengthening, and pause on infant processing demonstrated in this thesis as well as several other studies can be game-changing when disentangling different hypotheses on the origin of prosodic bootstrapping.**

Still, the importance of acoustic saliency is not negligible. In segmenting the speech signal, the presence of a pause is a cue that is independent of context, very local, and thus presumably easy to recognize. The presence of a pause in combination with the native prosodic knowledge is likely to guide the learner to perceive a major prosodic boundary even in a non-native language. Taken together with the evidence from prosodic boundary cue perception, I suggest that relying on the pause cue reflects a universal auditory phenomenon of perceiving highly salient acoustic features.

Studying phonemic contrasts, Narayan, Werker, and Beddor (2010: 418) raise the point of a “relative contribution of native-language experience and acoustic saliency in infancy and beyond.” While perceptual reorganization is commonly reflected in a decline in non-native phoneme discrimination in the absence of language experience, Narayan et al. shed light on the discrimination of certain consonantal contrasts that improve with language experience. In their study, they examined infants’ discrimination of two pairs of Filipino nasal place contrasts, one contrast, [ma] vs. [na], with a larger acoustic distance than the other one, and [na] vs. [ŋa], in which the spectral characteristics of the first and second formant are very similar. The phoneme [ŋ] is native in the onset position in Filipino, but not in English (where its occurrence is constrained to coda positions). Across all tested ages in the first year of life, English- and Filipino-learning infants successfully discriminated the acoustically salient difference of [ma] and [na]. The less salient [na] vs. [ŋa] contrast was only discriminated by the oldest Filipino-learning group. Narayan et al. argue that for this contrast young infants’ performance mainly reflects acoustic sensitivity, whereas the older Filipino-learning infants, who had relevant linguistic experience, were able to sharpen their discrimination abilities regarding finer-grained phonetic contrasts. Hence, what was observable here is an interplay of language experience and acoustic sensitivity in line with Aslin and Pisoni’s (1980) theory on perceptual attunement, which considers experiential and psychoacoustic factors for different ways of maturation. In this case, the sensitivity moves in the direction of enhancement and is thus comparable to the development we have seen in prosodic boundary cue sensitivity. The importance of taking differences in acoustic salience into account has also been observed in the maintenance of certain segmental contrasts. For example, the highly salient place contrast in Zulu clicks remains discriminable for older English infants and adults even in the absence of language experience (Best et al., 1988). Best, McRoberts, and Sithole (1988) argue that these clicks fall far outside the phonological space of English and are thus discriminable by means of (non-linguistic) auditory processing. Also for the domain of tonal contrasts, acoustic salience has been shown to be an influential factor in perceptual reorganization. Liu and Kager (2014) found that Dutch-learning infants from five to 18 months succeed in discriminating a natural Mandarin tone contrast. However, presented with the same contrast in a synthesized version with a weaker acoustic difference in F0, only the youngest and the oldest age groups discriminated the contrast while the intermediate ones failed. Liu und Kager interpreted this U-shaped development as a less robust discrimination after six months that is mediated by acoustic salience and recovers by 18 months as a result of knowledge and perceptual sensitivity to F0. In his review of infant speech perception research on phonemic and tonal contrasts, Narayan (2020) highlights acoustic salience for explaining the variation in discrimination patterns “as a mediating factor in determining the developmental trajectory for the perception of speech contrasts” (Narayan, 2020: 9).

Turning back to the two hypotheses on the origins of prosodic bootstrapping, the acoustic salience and the language-specific hypothesis, favoring one of them would be much too simplistic. Both hypotheses are connected to valid factors of infant speech perception. **For the case of infants' sensitivity to individual prosodic boundary cues, I have pointed out a role of acoustic saliency as well as phonological typology. This role varies depending on the developmental stage and on the specific acoustic features and their phonetic magnitude in the speech signal. The main boundary cues can be split into the rather dynamic and relational cues pitch change and final lengthening, which are subject to crosslinguistic variation, and the more static pause cue, which seems to be related to universal auditory processing.**

8.4 Conclusion

This thesis investigated infants' prosodic boundary cue perception within one language and across age. It provides support for the role of both acoustic salience and the linguistic input in shaping infants' developmental perception. Prosodic boundary cue perception, a phenomenon closely related to the prosodic bootstrapping of syntax, is initially driven by a language-general way of processing based on acoustic saliency. During the course of the first year of life, it seems to develop into a language-specific processing that takes more fine-grained, relational cues, relevant for a more precise analysis of the syntactic structure, more strongly into account. Since in ADS a large proportion of phrase boundaries are not signaled by a pause (Peters et al., 2005), a shift away from the pause cue is even necessary.

So far, only a few studies on Germanic and Romance languages and Japanese have addressed the phenomenon of infant prosodic boundary cue sensitivity. To verify the observed developmental pattern and to provide further insights into the interplay of language experience and acoustic salience, future research is necessary that tests non-native and native listeners of typologically varying languages with the same stimuli in one language. This includes using materials that systematically address the influence of different contexts of a prosodic boundary, the influence of the varying types of a language's tonal patterns, as well as the influence of different phonetic strengths of the individual cues marking the boundary contrast.

REFERENCES

Aasland, W. A., & Baum, S. R. (2003). Temporal parameters as cues to phrasal boundaries: A comparison of processing by left- and right-hemisphere brain-damaged individuals. *Brain and Language*, *87*(3), 385–399. https://doi.org/10.1016/S0093-934X(03)00138-X

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and Language*, *162*, 46–59. https://doi.org/10.1016/j.bandl.2016.08.002

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. H. Kavanagh, & A. Ferguson (Eds.), *Child phonology, Volume 2: Perception* (pp. 67–96). New York: Academic Press.

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*(6), 644–663. https://doi.org/10.1016/0749-596X(91)90030-N

Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46*(2–3), 183–216. https://doi.org/10.1177/00238309030460020701

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360.

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language affects rhythmic grouping of speech. *The Journal of the Acoustical Society of America*, *134*(5), 3828–3843. https://doi.org/10.1121/1.4823848

Bijeljac-Babic, R., Serres, J., Höhle, B., & Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PloS ONE*, *7*(2), 1–8. https://doi.org/10.1371/journal.pone.0030843

Boersma, P., & Weenink, D. (2010). PRAAT: doing phonetics by computer. Version 5.1.25.[software]. *Retrieved from http://www.praat.org/*.

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*(1), 29–49. https://doi.org/10.1207/S15327078IN0201_3

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV.Some effects of perceptual learning on speech product. *Journal of the Acoustical Society of America*, *101*(4).

Butcher, A. (1981). *Aspects of the Speech Pause: Phonetic Correlates and Communicative Functions.* Institut für Phonetik, Universität Kiel, Kiel.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, *95*(3). https://doi.org/10.1121/1.408544

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, *2*(3), 385–394. https://doi.org/10.1207/S15327078IN0203_6

Collins, B., & Mees, I. M. (1981). *The phonetics of English and Dutch*. Leiden: Brill.

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

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed dpeech. *Linguistics and Language Compass*, *7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Cristià, A. (2009). *Individual variation in infant speech perception: Implications for language acquisition theories*. Purdue University, IN.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201. https://doi.org/10.1177/002383099704000203

de Carvalho, A., Dautriche, I., & Christophe, A. (2016). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental Science*, *19*(2). https://doi.org/10.1111/desc.12300

de Carvalho, A., Dautriche, I., Lin, I., & Christophe, A. (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition*, *163*. https://doi.org/10.1016/j.cognition.2017.02.018

Fernald, A. (2000). Speech to infants as hyperspeech: Knowledge-driven processes in early word recognition. *Phonetica*, *57*(2–4), 242–254. https://doi.org/10.1159/000028477

Fernald, A., & McRoberts, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (Vol. 20, pp. 365–388). Mahwah, N.J.: Lawrence Erlbaum. https://doi.org/10.1017/S0142716499212088

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501.

Féry, C., Hörnig, R., & Pahaut, S. (2011). Correlates of phrasing in French and German from an experiment with semi-spontaneous speech. In C. Gabriel & C. Lleó (Eds.), *Intonational Phrasing in Romance and Germanic* (pp. 11–41). John Benjamins.

Féry, C., & Kentner, G. (2010). The prosody of embedded coordinations in German and Hindi. In *Proceedings of the 5th International Conferencepf Speech Prosody, 2010*. Chicago.

Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech : Cross-linguistic evidence. *Child Development*, *67*(6), 3192–3218.

Frank, M. C., Alcock, K. J., Arias-Trejo, N., Aschersleben, G., Baldwin, D., Barbu, S., … Soderstrom, M. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. https://doi.org/10.1177/2515245919900809

Friederici, A. D., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, *17*(14), 1208–1211. https://doi.org/10.1016/j.cub.2007.06.011

Frota, S., Butler, J., Severino, C., Uysal, E., & Vigário, M. (2019). Infant perception of prosodic boundaries without the pause cue: an eye-tracking study. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 3160–3164). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Frota, S., Butler, J., Uysal, E., Severino, C., & Vigário, M. (2020). European Portuguese-learning infants look longer at iambic stress: New data on language specificity in early stress perception. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01890

Frota, S., Butler, J., & Vigário, M. (2014). Infants' Perception of Intonation: Is It a Statement or a Question? *Infancy*, *19*(2). https://doi.org/10.1111/infa.12037

Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Probus*, *13*(2). https://doi.org/10.1515/prbs.2001.005

Gerken, L., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition*, *51*(3), 237–265.

Gervain, J. (2018). Gateway to Language: The perception of prosody at birth. In *Studies in Natural Language and Linguistic Theory* (Vol. 94, pp. 373–384). Springer Science and Business Media B.V. https://doi.org/10.1007/978-3-319-90710-9_23

Gervain, J., Christophe, A., & Mazuka, R. (2020). Prosodic Bootstrapping. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198832232.013.36

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

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

Goldman-Eisler, F. (1972). Pauses, Clauses, Sentences. *Language and Speech*, *15*(2), 103–113. https://doi.org/10.1177/002383097201500201

Gollrad, A. (2013). *Prosodic cue weighting in sentence comprehension: Processing German case ambiguous structures*. University of Potsdam.

Gollrad, A., Sommerfeld, E., & Kügler, F. (2010). Prosodic cue weighting in disambiguation: Case ambiguity in German. In *Proceedings of the 5th International Conference on Speech Prosody*.

Götz, A., Yeung, H. H., Krasotkina, A., Schwarzer, G., & Höhle, B. (2018). Perceptual reorganization of lexical tones: Effects of age and experimental procedure. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fmicb.2018.00477

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*, *51*(4), 548–567. https://doi.org/10.1016/j.jml.2004.07.002

Gussenhoven, C. (1992). Sentence accents and argument structure. In I. Roca (Ed.), *Thematic Structure: Its Role in Grammar* (pp. 79–106). Berlin: Foris.

Hawthorne, K., & Gerken, L. A. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, *133*(2). https://doi.org/10.1016/j.cognition.2014.07.013

Hawthorne, K., Mazuka, R., & Gerken, L. A. (2015). The acoustic salience of prosody trumps infants' acquired knowledge of language-specific prosodic patterns. *Journal of Memory and Language*, *82*, 105–117. https://doi.org/10.1016/j.jml.2015.03.005

Hayashi, A., & Mazuka, R. (2002). Developmental Change in Infants' Preferences for Continuousness in Japanese Speech. In *13th Biennial International Conference on Infant Studies, 2002*.

Herold, B., Höhle, B., Walch, E., Weber, T., & Obladen, M. (2008). Impaired word stress pattern discrimination in very-low-birthweight infants during the first 6 months of life. *Developmental Medicine and Child Neurology*, *50*(9), 678–683. https://doi.org/10.1111/j.1469-8749.2008.03055.x

Hirsh-Pasek, K, Kemler Nelson, D., Jusczyk, P., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*(3), 269–286.

Hirst, D., & Di Cristo, A. (1998). A Survey of Intonation Systems. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems A Survey of Twenty Languages* (pp. 1--44). Cambridge University Press.

Höhle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics*, *47*(2), 359–382. https://doi.org/10.1515/LING.2009.013

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behavior & Development*, *32*(3), 262–274. https://doi.org/10.1016/j.infbeh.2009.03.004

Höhle, B., Schmitz, M., Santelmann, L. M., & Weissenborn, J. (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: Evidence for lexical and structural influences on children's early processing capacities. *Language Learning and Development*, *2*(4), 277–300. https://doi.org/10.1207/s15473341lld0204_3

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Holzgrefe-Lang, J., Wellmann, C., Höhle, B., & Wartenburger, I. (2018). Infants' processing of prosodic cues: Electrophysiological evidence for boundary perception beyond pause detection. *Language and Speech*, *61*(1), 153–169. https://doi.org/10.1177/0023830917730590

Holzgrefe-Lang, J., Wellmann, C., Petrone, C., Räling, R., Truckenbrodt, H., Höhle, B., & Wartenburger, I. (2016). How pitch change and final lengthening cue boundary perception in German: converging evidence from ERPs and prosodic judgements. *Language, Cognition and Neuroscience*, *3798*(April), 1–17. https://doi.org/10.1080/23273798.2016.1157195

Hunter, M. A., & Ames, E. W. (1988). A multifactoral model of infant preferences for novel and familiar stimuli. In L. P. Lipsitt & C. Rovee-Collier (Eds.), *Advances in Infancy Research* (Vol. 5, pp. 69–95). Norwood, N.J.: Ablex.

Huttenlauch, C., de Beer, C., Hanne, S., & Wartenburger, I. (2021). Production of prosodic cues in coordinate name sequences addressing varying interlocutors. *Laboratory Phonology*, *12*(1), 1–31. https://doi.org/10.5334/LABPHON.221

Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A Crosslinguistic perspective. *Infancy*, *13*(5), 440–455. https://doi.org/10.1080/1

Jun, S.-A. (2014). Prosodic typology: by prominence type, word prosody, and macro-rhythm. In *Prosodic Typology II*. https://doi.org/10.1093/acprof:oso/9780199567300.003.0017

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675–687.

Jusczyk, P. W., Hirsh-Pasek, K., Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*(2), 252–293. https://doi.org/10.1016/0010-0285(92)90009-Q

Jusczyk, P. W., & Kemler Nelson, D. G. (1996). Syntactic units, prosody, and psychological reality during infancy. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 389–408). Lawrence Erlbaum Associates, Inc.

Keij, B., & Kager, R. (2016). The development of rhythmic preferences by Dutch learning infants. In J. Heinz, R. Goedemans, & H. van der Hulst (Eds.), *Dimensions of Phonological Stress* (pp. 291–310). Cambridge University Press. https://doi.org/10.1017/9781316212745.011

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*(1), 55–68. https://doi.org/10.1017/S030500090001343X

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, *18*(1), 111–116.

Kohler, K. J. (1983). Prosodic boundary signals in German. *Phonetica*, *40*(2), 89–134. https://doi.org/10.1159/000261685

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608. https://doi.org/10.1126/science.1736364

Lecanuet, J.-P., & Granier-Deferre, C. (1993). Speech stimuli in the fetal environment. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life.* (NATO ASI S, pp. 237–248). Dordrecht: Springer. https://doi.org/10.1007/978-94-015-8234-6_20

Lehiste, I., Olive, J. P., & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, *60*(5), 1199–1202. https://doi.org/10.1121/1.381180

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385–394. https://doi.org/10.1016/j.cognition.2014.06.004

Ludusan, B., Cristia, A., Martin, A., Mazuka, R., & Dupoux, E. (2016). Learnability of prosodic boundaries: Is infant-directed speech easier? *The Journal of the Acoustical Society of America*, *140*(2), 1239–1250. https://doi.org/10.1121/1.4960576

Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. G. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition*, *53*(2), 155–180. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7805352

Mandel, D. R., Kemler Nelson, D. G., & Jusczyk, P. W. (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development*, *11*(2), 181–196. https://doi.org/10.1016/S0885-2014(96)90002-7

Männel, C., & Friederici, A. D. (2008). Event-related brain potentials as a window to children's language processing. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Language Acquisition and Language Disorders: Vol. 44. Developmental Psycholinguistics. On-Line Methods in Children's Language Processing* (pp. 29–72). Amsterdam: Benjamins. https://doi.org/10.1075/lald.44.04man

Männel, C., & Friederici, A. D. (2009). Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. *Journal of Cognitive Neuroscience*, *21*(10), 1988–2006. https://doi.org/10.1162/jocn.2009.21221

Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience*, *5*, 86–94. https://doi.org/10.1016/j.dcn.2013.01.003

Mattock, K., Amitay, S., & Moore, D. R. (2012). Auditory development and learning. In C. Plack (Ed.), *Oxford Handbook of Auditory Science Hearing* (Volume 3, pp. 297–324). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199233557.013.0013

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception : Evidence for perceptual reorganization. *Infancy*, *10*(3), 241–265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, *106*(3), 1367–1381. https://doi.org/10.1016/j.cognition.2007.07.002

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178. https://doi.org/10.1002/dev.21177

Mehler, J., Jusczyk, P., Lamsertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495–500. https://doi.org/10.1016/0163-6383(93)80007-U

Morgan, James L, & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. *Signal to Syntax*. Lawrence Erlbaum Associates.

Narayan, C. R. (2020). An acoustic perspective on 45 years of infant speech perception . II . Vowels and suprasegmentals. *Language and Linguistics Compass*, 1–12. https://doi.org/10.1111/lnc3.12369

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*(3). https://doi.org/10.1111/j.1467-7687.2009.00898.x

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 756–766. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9627414

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*(1), 123–147. https://doi.org/10.1207/S15327078IN0101_11

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Peters, B. (2005). Weiterführende Untersuchungen zu prosodischen Grenzen in deutscher Spontansprache. In K. J. Kohler, F. Kleber, & B. Peters (Eds.), *Prosodic structures in German spontaneous speech (AIPUK 35a)* (pp. 203–345). Kiel: IPDS.

Peters, B., Kohler, K. J., & Wesener, T. (2005). Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In K. J. Kohler, F. Kleber, & B. Peters (Eds.), *Prosodic structures in German spontaneous speech (AIPUK 35a)* (Vol. 35a, pp. 143–184). Kiel: IPDS.

Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenburger, I., & Höhle, B. (2017). Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics*, *61*, 71–92. https://doi.org/10.1016/j.wocn.2017.01.002

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435.

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*(6), 2956–2970.

Pujol, R., Lavigne-Rebillard, M., & Lenoir, M. (1998). Development of sensory and neural structures in the mammalian cochlea (pp. 146–192). https://doi.org/10.1007/978-1-4612-2186-9_4

Ramus, F., Nespor, M., & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition*, *75*(1), 265–292.

Sanderman, A., & Collier, R. (1997). Prosodic phrasing and comprehension. *Language and Speech*, *40*(4), 391–409.

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, *33*(1), 3–11.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *Journal of Cognitive Neuroscience*, *22*(11). https://doi.org/10.1162/jocn.2009.21377

Schipke, C. S. (2012). *Processing mechanisms of argument structure and case-marking in child development: Neural correlates and behavioral evidence*. Max Planck Institute for Human

Cognitive and Brain Sciences, Leipzig.

Schmitz, M. (2008). *The perception of clauses in 6- and 8-month-old German-learning infants : Influence of pause duration and the natural pause hierarchy*. Universtätsverlag Potsdam, Potsdam.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundarya. *Journal of the Acoustical Society of America*, *71*(4), 996–1007. https://doi.org/10.1121/1.387581

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, *57*(1), 24–48. https://doi.org/10.1016/j.jml.2006.10.004

Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, *11*(4), 596–606. https://doi.org/10.1111/j.1467-7687.2008.00704.x

Selkirk, E. O. (1978). On prosodic structure and its relation to syntactic structures. In T. Fretheim (Ed.), *Nordic Prosody II, TAPIR*. Trondheim.

Selkirk, E. O. (1984). *Phonology and Syntax. The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology*, (3), 371–405.

Selkirk, E. O. (2005). Comments on intonational phrasing in English. In S. Frota, M. Vigário, & M. J. Freitas (Eds.), *Prosodies. With special reference to Iberian languages* (pp. 11–58). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197587.1.11

Severino, C. (2016). *Perception of phrasal prosody in the acquisition of European Portuguese*. Universidade de Lisboa.

Shi, R. (2010). Contextual variability and infants' perception of tonal categories. *Chinese Journal of Phonetics*, *2*, 1–9.

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. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6038–6043. https://doi.org/10.1073/pnas.1017617108

Skoruppa, K., Cristià, A., Peperkamp, S., & Seidl, A. (2011). English-learning infants' perception of word stress patterns. *The Journal of the Acoustical Society of America*, *130*(1). https://doi.org/10.1121/1.3590169

Skoruppa, K., Pons, F., Bosch, L., Christophe, A., Cabrol, D., & Peperkamp, S. (2013). The development of word stress processing in French and Spanish infants. *Language Learning and Development*, *9*(1). https://doi.org/10.1080/15475441.2012.693881

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., … Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*(6), 914–919. https://doi.org/10.1111/j.1467-7687.2009.00835.x

Soderstrom, M., Nelson, D. G. K., & Jusczyk, P. W. (2005). Six-month-olds recognize clauses embedded in different passages of fluent speech. *Infant Behavior and Development*, *28*(1), 87–94. https://doi.org/10.1016/j.infbeh.2004.07.001

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*(2), 249–267. https://doi.org/10.1016/S0749-596X(03)00024-X

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*(2), 191–196. https://doi.org/10.1038/5757

Strangert, E. (1991). Where do pauses occur in text read alound? In H. A. Sigurdsson (Ed.), *Papers from the Twelfth Scandinavian Conference of Linguistics* (pp. 403–414). Reykjavik: Linguistic Institute, University of Iceland.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*, *64*(6), 1582–1592.

Sundara, M., Ngon, C., Skoruppa, K., Feldman, N. H., Onario, G. M., Morgan, J. L., & Peperkamp, S. (2018). Young infants' discrimination of subtle phonetic contrasts. *Cognition*, *178*. https://doi.org/10.1016/j.cognition.2018.05.009

Terken, J., & Collier, R. (1992). Syntactic influences on prosody. In Y. Tokhura, E. Vatikiotis-Bateson, & Y. Sagisaki (Eds.), *Speech perception, production and linguistic structure* (pp. 427–438). Amsterdam: IOS Press.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry*, *30*, 219–255.

Truckenbrodt, H. (2002). Upstep and embedded register levels. *Phonology*. https://doi.org/10.1017/S095267570200427X

Truckenbrodt, H. (2005). A short report on intonation phrase boundaries in German. *Linguistische Berichte*, *203*, 273–296.

Truckenbrodt, H. (2007a). The syntax-phonology interface. In P. de Lacy (Ed.), *The Cambridge handbook of Phonology* (pp. 435–456). Cambridge, UK: Cambridge University Press.

Truckenbrodt, H. (2007b). Upstep on edge tones and on nuclear accents. In C. Gussenhoven & T. Rias (Eds.), *Tones and Tunes. Volume 2: Experimental Studies in Word and Sentence Prosody* (pp. 349–386). Berlin: Mouton. https://doi.org/10.1515/9783110207576.2.349

Truckenbrodt, H. (2016). Focus, intonation, and tonal height. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 463–482). Oxford, UK: Oxford University Press.

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 53–66). Berlin: Springer.

Vaissière, J., & Michaud, A. (2006). Prosodic constituents in French: A data-driven approach. In I. Fónagy, Y. Kawaguchi, & T. Moriguchi (Eds.), *Prosody and Syntax* (pp. 47–64). Amsterdam: Benjamins.

van Ommen, S., Boll-Avetisyan, N., Larraza, S., Wellmann, C., Bijeljac-Babic, R., Höhle, B., & Nazzi, T. (2020). Language-specific prosodic acquisition: A comparison of phrase boundary perception by French- and German-learning infants. *Journal of Memory and Language*, *112*. https://doi.org/10.1016/j.jml.2020.104108

Venditti, J. J., Jun, S.-A., & Beckman, M. E. (1996). Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English. In James L. Morgan & K. Demuth (Eds.), *Signal to syntax: bootstrapping from speech to grammar in early acquisition* (pp. 287–311). Lawrence Erlbaum Associates.

Wagner, M. (2005). *Prosody and Recursion*. Massachusetts Institute of Technology.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7). https://doi.org/10.1080/01690961003589492

Weijer, J. Van De. (1997). Language Input to a Prelingual Infant. In A. Sorace, C. Heycock, & R. Shillcock (Eds.), *Proceedings of the GALA '97 conference on language acquisition* (pp. 290–293). Edinburgh University Press.

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012). How each prosodic boundary cue matters: evidence from German infants. *Frontiers in Psychology*, *3*, 580. https://doi.org/10.3389/fpsyg.2012.00580

Wellmann, C., Holzgrefe-Lang, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (forthcoming) Developmental changes in prosodic boundary cue perception in German-learning infants. In: F. Schuboe, S., Hanne, S., Zerbian, & I. Wartenburger (Eds.) *Prosodic boundary phenomena*. Studies in Laboratory Phonology. Berlin: Language Science Press.

Wells, B., & Peppé, S. (2003). Intonation abilities of children with speech and language impairments. *Journal of Speech, Language, and Hearing Research : JSLHR*, *46*(1), 5–20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12647884

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

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*(3), 1707–1717.

Willems, N. (1982). *English intonation from a Dutch point of view*. Dordrecht: Foris.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, *68*(2). https://doi.org/10.1016/j.jml.2012.09.004

Zhang, X. (2012). *A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese*. University of Michigan.

ERKLÄRUNG

Hiermit erkläre ich, Caroline Wellmann, dass ich die vorliegende Dissertation selbstständig und ohne Hilfe Dritter verfasst habe. Bei der Abfassung der Arbeit wurden alle Regelungen guter wissenschaftlicher Standards eingehalten. Ich erkläre ferner, dass ich diese Arbeit ohne Benutzung anderer als der angegeben Quellen und Hilfsmittel verfasst habe.

Potsdam, den 30.6.2022 ____