



**HAL**  
open science

# Language-specific prosodic acquisition: A comparison of phrase boundary perception by French-and German-learning infants

Sandrien van Ommen, Natalie Boll-Avetisyan, Saioa Larraza, Caroline Wellmann, Ranka Bijeljac-Babic, Barbara Höhle, Thierry Nazzi

► **To cite this version:**

Sandrien van Ommen, Natalie Boll-Avetisyan, Saioa Larraza, Caroline Wellmann, Ranka Bijeljac-Babic, et al. Language-specific prosodic acquisition: A comparison of phrase boundary perception by French-and German-learning infants. *Journal of Memory and Language*, 2020, 10.1016/j.jml.2020.104108 . hal-02619164

**HAL Id: hal-02619164**

**<https://hal.science/hal-02619164v1>**

Submitted on 25 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Language-specific prosodic acquisition: A comparison of phrase boundary perception by French- and German-learning infants**

Sandrien van Ommen<sup>a</sup>

Natalie Boll-Avetisyan<sup>b</sup>

Saioa Larraza<sup>ac</sup>

Caroline Wellmann<sup>b</sup>

Ranka Bijeljac-Babic<sup>a</sup>

Barbara Höhle<sup>b</sup>

Thierry Nazzi<sup>a</sup>

<sup>a</sup> CNRS, Université Paris Descartes, France

<sup>b</sup> Universität Potsdam, Germany

<sup>c</sup> University of the Basque Country, Spain

**Cite as:** 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*, 104108. <https://doi.org/10.1016/j.jml.2020.104108>

### **Highlights**

1. Cross-linguistic infant study of general to language-specific prosody perception
2. Identical stimuli reveal language-specific development in French and German
3. Higher sensitivity to French prosody in German-learning than French-learning infants
4. Structural variability in input enhances early perception of phonetic features
5. By adulthood, enhanced prosodic sensitivity less apparent

## **Language-specific prosodic acquisition: a comparison of phrase boundary perception by French- and German-learning infants**

### **Abstract**

This study compares the development of prosodic processing in French- and German-learning infants. The emergence of language-specific perception of phrase boundaries was directly tested using the same stimuli across these two languages. French-learning (Experiment 1, 2) and German-learning 6- and 8-month-olds (Experiment 3) listened to the same French noun sequences with or without major prosodic boundaries ([Loulou et Manou][et Nina]; [Loulou et Manou et Nina], respectively). The boundaries were either naturally cued (Experiment 1), or cued exclusively by pitch and duration (Experiment 2, 3). French-learning 6- and 8-month-olds both perceived the natural boundary, but neither perceived the boundary when only two cues were present. In contrast, German-learning infants develop from not perceiving the two-cue boundary at 6 months to perceiving it at 8 months, just like German-learning 8-month-olds listening to German (Wellmann, Holzgrefe, Truckenbrodt, Wartenburger, & Höhle, 2012). In a control experiment (Experiment 4), we found little difference between German and French adult listeners, suggesting that later, French listeners catch up with German listeners. Taken together, these cross-linguistic differences in the perception of identical stimuli provide direct evidence for language-specific development of prosodic boundary perception.

### **Keywords**

Prosody, Acquisition, Language-specific, Perception, Infant, Prosodic boundaries

### **CRedit author statement**

**Sandrien van Ommen:** Conceptualization, formal analysis, investigation, writing (original draft), data curation; **Natalie Boll-Avetisyan:** Conceptualization, formal analysis, investigation, writing (original draft); **Saioa Larazza:** Conceptualization, investigation, writing (review and editing); **Caroline Wellmann:** Methodology, writing (review and editing); **Ranka Bijeljic-Babic:** Conceptualization, funding acquisition, resources; **Barbara Höhle:** Conceptualization, funding acquisition, resources, supervision, writing (review and editing); **Thierry Nazzi:** Conceptualization, resources, supervision, project administration, writing (review and editing).

## Introduction

A major challenge of language acquisition is for infants to specify which differences between sounds and sound patterns are linguistically relevant, and which differences are not. In the first six months of life, infants perceive phonetic differences in a language-general way, but with increasing linguistic experience, perception is filtered through the phonological system of the native language. This process of perceptual attunement has been well-studied for segmental phonology (consonants and vowels), but less is known about the development of the perception of prosody (the rhythm and melody of speech). However, prosody – like segmental information – is linguistically highly relevant and is subject to cross-linguistic variation. In languages such as English and German, for instance, pitch, duration, and amplitude mark lexical stress, but also phrasal stress, phrase boundaries, and syntax. Other languages, such as Korean and French, also mark phrase boundaries and syntax with prosody but, in contrast, do not have lexical stress, and phrasal stress by default coincides with the phrasal boundary. Thus, infants have to learn the weights and interactions of these acoustic cues for different linguistic functions that are specific to their native language. Hence, just like for segmental information, infants are expected to develop language-specific sensitivity to prosodic information.

In the present study, we focus on the development of infants' sensitivity to prosodic boundary cues, which are one of the earliest cues available to infants for segmenting and structuring language. Given the links between prosodic and syntactic units, prosodic boundary cues (like pauses, pitch changes and final lengthening) may serve as bootstrapping cues to the syntactic structure of language (Gleitman & Wanner, 1982; Hawthorne & Gerken, 2014). Sensitivity to this kind of prosodic information is present from early on. Even newborns perceive phrase boundary cues, as shown by French-learning newborns' ability to discriminate between disyllables with and without an internal prosodic boundary both in French (Christophe, Dupoux, Bertoncini, & Mehler, 1994) and in Spanish (Christophe, Mehler, & Sebastián-Gallés, 2001). From around seven months of age, infants prefer phrases containing pauses at prosodic boundaries over those with pauses within prosodic units, and this has been shown in both American English (Gerken, Jusczyk, & Mandel, 1994; Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989) and German (Schmitz, 2008). Regarding the actual use of prosodic boundaries for speech segmentation, studies have shown that by 6 months, prosody helps English-learning infants to segment and memorize syllable sequences delimited by prosodic boundaries (Nazzi,

Nelson, Jusczyk, & Jusczyk, 2000; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). Furthermore, phonological phrase boundaries restrict lexical access in English-learning 10- to 13-month-olds (Gout, Christophe, & Morgan, 2004). These studies establish that from early on, infants perceive and use prosodic phrases for processing and memorization. However, even though the same set of acoustic correlates is used across languages to mark prosodic boundaries, the weighting of the individual cues in their relevance as boundary markers may vary across languages (Johnson & Seidl, 2008; Seidl & Cristia, 2008), and the use of these same correlates for other linguistic functions differs per language, too. Therefore, the questions arise whether infants perceptually attune to the phrasal prosody of their native language and, if so, when.

### **Language-specificity in development**

Given the language-general and language-specific characteristics of phrasal prosody, prosodic boundary perception can be hypothesized to follow a two-step developmental trajectory. At first, infants may use language-general salient acoustic characteristics of prosody such as pauses and pitch peaks to get a handle on higher order structure (language-general prosodic processing). Then, through exposure to their native language, they would learn which prosodic dimensions are functionally important in their native language and which acoustic properties should be treated as redundancy, variation, or noise. Convergent evidence from different subdomains of phonological acquisition suggests that this process of perceptual attunement (Aslin & Pisoni, 1980) proceeds mostly during the first year of life (for consonants: Best, McRoberts, & Sithole, 1988; Werker & Tees, 1984; for vowels: Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; for word stress: Höhle et al., 2009; Skoruppa et al., 2013, for tone: Mattock & Burnham, 2006). Perceptual attunement is usually associated with a perceptual decline in non-native contrast discrimination; for example, English-learning infants lose the ability to discriminate between dental and retroflex Hindi stops (Werker & Tees, 1984), a contrast that is not phonemic in their native language. However, perceptual attunement also results in enhancement or maintenance of perceiving specific sound contrasts. These latter two developmental processes have initially been considered to affect *native* sound discrimination (Kuhl et al., 2006) but recent studies have shown that both can also appear in *non-native* tone (Chen, Liu, & Kager, 2015; Liu & Kager, 2014) and vowel perception (Mazuka, Hasegawa, & Tsuji, 2014). Which direction perceptual attunement takes depends on acoustic salience of the contrast and its relation to native

categories, attesting that it is reflecting acquisition of the native categories (e.g., Best & McRoberts, 2003; Best et al., 1988; Shi, Santos, Gao, & Li, 2017).

Whether phrasal prosody attunement is a process of loss, enhancement or maintenance is not well understood yet, because there are no cross-linguistic studies on this question. Earlier work studied single languages on the relative perceptual importance of three phonetic cues that concur in marking prosodic boundaries: pitch, lengthening and pause.<sup>1</sup> These studies followed up on a study by Nazzi et al. (2000) that established infants' sensitivity to natural prosodic boundaries. Using the Headturn Preference Paradigm (HPP), English-learning 6-month-olds were familiarized with full clauses forming a prosodic unit such as [leafy vegetables taste so good] and non-clauses with an internal prosodic boundary such as [leafy vegetables][taste so good]. They were then tested on their recognition of these clauses and non-clauses embedded in longer sentences. Infants preferred listening to the sentences containing the familiarized full clause in comparison to the non-clause. Later studies investigated in more detail the relative roles of pitch, duration and pausing as phonetic cues in prosodic boundary perception across development. For English, it was found that 4-month-olds needed the combination of all cues to perceive a boundary (Seidl & Cristià, 2008), while 6-month-olds needed only a subset of phonetic cues, with pitch being necessary, but pause not (Seidl, 2007). A similar study was conducted with Dutch-learning infants (Johnson & Seidl, 2008), showing that, in contrast, the presence of pauses was crucial for the 6-month-olds to be able to perceive a boundary. Related studies were conducted in German (Höhle, Wellmann, & Holzgrefe-Lang, 2013; Wellmann et al., 2012). In these studies, infants were also tested with an HPP paradigm with a familiarization and a test phase, but the stimuli were different from the previous studies. The materials were simple concatenations of first names, so they were highly restricted in their lexical and syntactic structure, but both sequences with- and without prosodic boundary were prosodically well-formed. Two conditions were compared: one in which prosodic phrasing indicated one group of three people ([Moni und Lilli und Manu]) and one in which it indicated two groups, one of two and one of one ([Moni und Lilli][und Manu]). Results revealed that, like in Dutch, 6-month-olds rely on the presence of pauses to

---

<sup>1</sup> Phrase boundaries are additionally realized by articulatory and acoustic segmental fortition (Fougeron & Keating, 1997), gradually so depending on the level of the phrase in the prosodic hierarchy. It is widely assumed that pitch, duration and pause are the most important cues for boundary perception regardless of language-specific differences in weighting, which was the reason for Wellmann, Holzgrefe, Truckenbrodt, Wartenburger, & Höhle (2012) and, hence, the current study, to study these cues in isolation in perception.

perceive a boundary, whereas discrimination in 8-month-olds did not depend on the presence of a pause, but instead relied on pitch and final lengthening.

One clear generalization from this previous evidence is that pausing is crucial for prosodic boundary detection for younger but not for older infants. This suggests that prosodic perceptual attunement may be a process of enhancement: through development, infants become more sensitive to individual or subtle prosodic cues. The available findings also signal an important role for pausing, which is mirrored in infants' typical speech input: it is a reliable and salient prosodic cue in infant directed speech (IDS) across languages (Ludusan, Cristia, Martin, Mazuka, & Dupoux, 2016). However, pausing is not reliably used as a prosodic boundary marker in adult directed speech (Nespor & Vogel, 1986; Peters, Kohler, & Wesener, 2005; Vaissière, 1983). Hence, while the pause might be a universally useful and necessary cue for boundary detection at a very young age, reliance on pause needs to be later refined based on the weighting of cues that are most important for prosodic boundary marking in the specific language in acquisition.

Note that evidence for such an attunement in the development of boundary processing is still scarce and restricted to very few languages. Previous research suggests that English-learning infants do not need a pause for boundary perception at an age at which Dutch and German infants still rely heavily on pause information (Johnson & Seidl, 2008; Seidl, 2007; Wellmann et al., 2012). Moreover, whereas pitch seems to be a necessary cue for boundary detection for English-learning 6-month-olds, this is not the case for German-learning 6-month-olds (Höhle et al., 2013; Seidl, 2007). However, these between-study comparisons have to be interpreted with caution. While differences in the results could reflect language-specific prosodic boundary perception, they could also arise from differences in the stimuli used in these studies. It is possible that the phonetic cues in the English, German and Dutch stimuli were not equally salient, irrespective of whether their realization was representative of natural phrase boundary realization in each of the languages. Moreover, the languages tested so far were all prosodically similar Germanic languages, limiting the generalizability of the results. The present study extends this research in three directions. First, it directly studies language-specific effects on perception, by comparing how infants learning two different languages (German and French) process the same materials. Second, it broadens the cross-linguistic perspective by providing data from a Romance language, namely French. Third, it investigates the potential role of phonological prosodic regularity, in this case on the level of

macro rhythm, on the development of prosodic boundary perception. We present more details on macro rhythm and other prosodic characteristics of French and German below.

### **Language-specific prosody**

French and German differ markedly in their structures and realizations of prosodic phrases (Delattre, 1965). French is characterized as a phrase language (Féry, 2017) or boundary language (Vaissière & Michaud, 2006), having phrase stress but neither word stress nor tone, (Jun, 2005). The smallest prosodic phrase is the accentual phrase which often but not always corresponds to one word. French is classified as a language with a strong macro rhythm (Jun, 2014), which means that it has a high regularity in tonal patterns on the phrasal level. While rhythm is usually defined in terms of duration-based temporal organization, macro rhythm is a tonal rhythm defined at the phrasal level. The tonal regularity can be composed of word stress, word tone, phrase boundary tones or any combination of these, and it can be perceived to be more regular both when the timing of the tones is more regular and when the tonal repertoire is more restricted. Accordingly, French is classified as a language with strong macro rhythm because both its timing and its restricted tonal repertoire result in a highly regular tonal macro rhythm. Production studies show that French prosody is characterized by a regular rising and lengthening pattern at each phrase boundary, strictly aligned to the final syllable; typically followed by a final pitch fall at the end of major clauses (Vaissière & Michaud, 2006). Perception studies with adults have shown the importance of final lengthening for the identification of minor prosodic phrases and a combination of lengthening and pitch for major prosodic boundaries (Michelas & D'Imperio, 2015).

German, in contrast, has positionally variable stress. It has a variety of pitch and duration patterns at the phrasal and lexical levels, not strictly aligned to the final syllables of phrases, but rather to lexically-stressed syllables, thereby overruling strictly boundary-driven phrasal regularities. German is described as a language with medium macro rhythm (Jun, 2014), because it has a relatively large tonal repertoire and the domain at which it places tones is notably larger than a content word, making the intonation pattern overall less rhythmic. Production studies show that German speakers produce a close correspondence of pitch and lengthening cues to mark embedded phrase structure (Féry & Kentner, 2010). Perception studies show that German adult listeners use pitch (Gollrad, 2013; Petrone et al., 2017) or a combination of pitch and lengthening cues to perceive major boundaries (Holzgreffe-Lang et al., 2016), and lengthening to identify minor prosodic boundaries (Gollrad, 2013; Petrone et



al., 2017). Comparing French and German, Féry, Hörnig, & Pahaut (2011) found greater pre-boundary lengthening in French than in German. We summarize some of the relevant differences between the two languages in Table 1.

Table 1. Structural prosodic differences between French and German

	<b>French</b>	<b>German</b>
<b>Macro rhythm</b>	Strong	Medium
<b>Word-stress</b>	No	Yes
<b>Tonally marked embedded phrase structure</b>	No	Yes
<b>Pitch accent (phrase stress position)</b>	Final	Variable
<b>Number of pitch accents in inventory</b>	3 to 4	6
<b>Final lengthening</b>	Strong	Weak

The present study investigates prosodic boundary perception using both a cross-linguistic (comparing two prosodically different languages, French and German) and a developmental (comparing 6- and 8-month-olds) perspective, in order to determine whether a change from language-general to language-specific perception can be observed in this age range. We used the same method as Wellmann et al. (2012) with highly comparable material to allow for between-studies comparisons of (a) native prosodic boundary perception in French- or German-learning infants, (b) native and non-native prosodic boundary perception in German-learning infants, and (c) a direct (within-study) comparison of French- and German-learning infants' perception of French prosodic boundaries. Experiment 1 targets developmental changes in the perception of natural French prosodic boundaries by 6- and 8-month-old French-learning infants. Experiment 2 targets the perceptual role of pausing in French-learning 6- and 8-month-olds. Experiment 3 directly targets the language-specificity of the results of Experiment 2 by testing German-learning 6- and 8-month-olds with the same (French) stimuli. Finally, in light of the results, Experiment 4 investigates boundary perception in the same stimuli in French- and German-speaking adults.

### **Predictions**

If the previously documented reliance on pause as a boundary cue is a general developmental trait, no cross-linguistic differences should surface in our experiments. In contrast, if prosodic boundary perception changes from being language-general to being language-specific between 6 and 8 months, cross-linguistic differences should be found in the older age group,

given the prosodic differences between French and German reviewed above. The regularity of the strong macro rhythm in French bound to the phrasal boundary tones should lead to better predictability of the occurrence of phrase boundaries, with the implication that perceiving the local acoustic cues marking the boundary may be less important in French than in German. More specifically, by 8 months, we expect German-learning infants to rely heavily on the crucial acoustic markers for phrase boundaries in German (pitch and final lengthening) while French-learning infants may be less responsive to these acoustic boundary cues given the high degree of rhythmicity and, thereby, predictability of phrase boundaries in French. We make this prediction in analogy to the stress “deafness” hypothesis. This hypothesis was first proposed to explain cross-linguistic data from studies on perception and sensitivity to lexical stress. Adult speakers of French have been found to have reduced sensitivity to word stress as compared to speakers of lexical stress languages such as Spanish, a pattern further extended to other languages with predictable stress placement such as Hungarian and Finnish (Dupoux, Pallier, Sebastian, & Mehler, 1997; Dupoux, Peperkamp, & Sebastian-Galles, 2001; Peperkamp, Vendelin, & Dupoux, 2010). Moreover, there is converging evidence that these effects are set into place early in development, between 6 and 9/10 months of age: at 6 months, learners of both Spanish and French showed similar word stress discrimination abilities (only when there was no token variability), and by 9 months, learners of Spanish but not French were able to discriminate word stress under high token variability (Skoruppa et al., 2013, 2009). Similar findings were found in studies comparing French monolingual and bilingual infants (Abboub, Bijeljac-Babic, Serres, & Nazzi, 2015; Bijeljac-Babic, Serres, Höhle, & Nazzi, 2012). A second domain in which French listeners show a different or weaker use of stress is that of prosodic grouping. Indeed, French adult listeners have been found to be less consistent in their application of the Iambic/Trochaic Law (Hayes, 1985, 1989) than German listeners when processing high-variability stimuli (Bhatara, Boll-Avetisyan, Agus, Höhle, & Nazzi, 2016; Bhatara, Boll-Avetisyan, Unger, Nazzi, & Höhle, 2013). However, such cross-linguistic differences were not yet found at 7 months of age, as infant learners of these two languages still displayed language-general prosodic grouping (Abboub, Boll-Avetisyan, Bhatara, Höhle, & Nazzi, 2016). Thus, the current study tested the hypothesis that at the phrase level, high prosodic regularity also leads to lower sensitivity to specific acoustic information. Experiment 1 investigates whether French-learning 6- and 8-month-olds detect French natural prosodic boundaries and whether there is a developmental change in their ability to do so.

## Experiment 1

### Method

#### *Participants*

Forty 6-month-old (21 girls,  $M_{\text{age}} = 6.30$  months; range: 5.85–6.93) and 40 8-month-old (18 girls,  $M_{\text{age}} = 8.22$  months; range: 7.85–8.61) French-learning monolingual infants were tested at the babylab of the Laboratoire Psychologie de la Perception, Paris (currently the Integrative Neuroscience and Cognition Center). An additional 11 6-month-olds were tested and excluded due to fussiness (9) or having two consecutive trials with insufficient ( $< 1500$  ms) orientation times (2). An additional 19 8-month-olds were tested and excluded due to fussiness (10), having two consecutive trials with insufficient orientation times (7), technical problems (1), or parental interference (1). All infants were without apparent health problems, not at high familial risk of developmental or linguistic disorders, and they were born after at least 37 weeks of gestation. They were recruited from birth lists obtained through the Paris city hall archives or through a database of families who had previously participated in speech perception studies in our laboratory. Informed written consent was obtained from all parents.

#### *Stimuli*

Following Wellmann et al. (2012), the stimuli consisted of a sequence of three coordinated, in this case French, names. The stimuli were different tokens of a single sequence (Loulou et Manou et Nina) produced by a native French speaking mother in IDS (with her baby being present) Recordings were made in a sound-attenuated booth with a TACSAM DR-07-MKII recorder at a sampling rate of 48000 with 24-bit resolution. She read the sequences in two different prosodic groupings as indicated by different bracketing as below.

No boundary:            [Loulou et Manou et Nina]

With boundary:        [Loulou et Manou] [et Nina]

The prosodic implementation of the bracketing in the speakers' productions was subsequently confirmed by two independent judges, both native speakers of French. Measurements of some relevant acoustic properties of the stimuli, given in Table 2, also confirm the intended prosodic differences.

Table 2. Means and standard deviations of the phonetic correlates of prosodic boundaries in the French stimuli. Measures refer to the vowels in the second name, and the pause after the second name.

<b>Condition</b>	<b>Duration</b> $V_{\text{final}}^a$ (ms)	<b>Pitch rise</b> ( $\max V_{\text{final}} - \min V_{\text{prefinal}}$ ) (semitones)	<b>Pause duration</b> (ms)
No-boundary	141 (19)	6.76 (1.00)	0
With-boundary	184 (27)	12.8 (1.36)	633 (126)

a.  $V_{\text{final}}$  = vowel in last syllable of the second name,  $V_{\text{prefinal}}$  = vowel in prefinal syllable of the second name

For each condition (with-boundary versus no-boundary), we created two sound files, each containing two repetitions of four different tokens of each phrase. The two sound files differed only in the order of the tokens. All sound files had the same duration of 25.34 s and contained the same number of tokens. Given the longer duration of the tokens in the with-boundary condition, the ISI between the single tokens was extended in the no-boundary condition to achieve the same duration of the files.

#### *Procedure, apparatus and design*

We used the Headturn Preference Procedure (HPP) (Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989) including a familiarization and a test phase. This is a classic experimental paradigm to measure infants' discrimination of a priori equally familiar items. Through the familiarization phase, the infant becomes more "familiar" with one type of stimulus. The other type of stimulus will first be presented during the test phase, and perceived as "novel" if infants can discriminate between the two types of stimuli. If not, there will be no difference in orientation time to the two types. Familiarization condition was counterbalanced: one group of infants heard no-boundary sequences, and the other heard with-boundary sequences. At test, they heard both types of sequences. Infants were randomly assigned to a familiarization condition, with 20 infants per age group and condition.

The experiment took place in a dimly-lit sound-attenuated booth, and the stimuli were played at 70dB. The infant was held on a caregiver's lap in the center of the test booth. The experimenter sat outside the test booth and observed the infant using a camera which was positioned underneath a green light. Both the caregiver and the experimenter wore earplugs as well as headphones with a music/speech mask. A green light was directly in front of the infant and there was a red light at the same height on each side of the booth. Loudspeakers were

hidden behind the red lights. Stimulus presentation and the blinking of the lights were controlled by the experimenter by means of a button box with a button for each of the lights. Each trial began with the green center light blinking until the infant had oriented to it. Then, the experimenter pressed the associated button and one of the red side lights began to flash. As soon as the infant oriented to the red light, the experimenter pushed the associated button and a sound file began to play. Each sound file played until completion, or until the infant looked away for longer than 2 seconds. Then, the green light started to flash again. The time spent looking away was automatically subtracted from the total orientation time. If the infant oriented shorter than 1.5 seconds during a trial, this trial was discarded and replayed right away. If the infant also oriented for less than 1.5 seconds to the repeated trial, the infant was excluded from analysis.

Infants were familiarized until they reached a familiarization criterion of 60 seconds of orientation time, 30 seconds per sound file. If there was still a sound file playing when the familiarization criterion was reached, it played until completion or until the infant looked away. After the familiarization phase, the test phase started, consisting of 12 trials distributed across three blocks, each block containing each of the 4 sound files. The stimuli were pseudo-randomized on the condition (no condition more than two times in a row) and the side from which the sound files were presented (no side more than two times in a row), with a different randomization per infant.

### *Analysis*

Results were analyzed with a mixed-model using R (R Core Team, 2016), lme4 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2016), plots were made using ggplot2 (Wickham, 2009). We used a predefined model reflecting our experimental design, and we kept the model structure the same across experiments. In this model, infants' orientation times in seconds were predicted by familiarity (novel versus familiar type), familiarization condition (familiarization with no-boundary versus with-boundary stimuli) and age (6 versus 8 months), taking into account covariance of the familiarization time in seconds (scaled) and trial order (centered), interacting with familiarity<sup>2</sup>. Familiarization time in seconds was included because it varies

---

<sup>2</sup> An analysis without the covariates led to the same results. Covariates are entered in the analysis to yield higher precision in the estimates, i.e. to get a better approximation of the amount and kind of variation that can be attributed to the experimental variables.

per infant as a result of their orientation behavior during familiarization, and might influence infants' orientation to familiar and novel trials during the test (e.g., long familiarization can flip a familiarity preference to a novelty preference (Hunter & Ames, 1988). Trial order was included because trials are not independent measures (they are related in time) and infants' behavior during HPP experiments can change over time. The model additionally included individual participant intercepts and slopes of trial order (centered) and familiarity, for maximum random effects structure (Barr, Levy, Scheepers, & Tily, 2013) in order to account for individual variation. Levels of all categorical factors were centered, so the intercept of the model represented the grand mean.

*Formula: orientation time ~ (familiarity \* familiarization condition \* age) + (familiarity \* trial order \* familiarization time) + (1 + trial order \* familiarity | participant)*

Our main hypothesis is tested with the variable “Familiarity.” Hence, the results sections will focus on effects involving familiarity. All other effects can be found in the tables.

To assess what inferences can be made from null results, we additionally ran each model over the standardized dependent variable, i.e., the z-scores of the dependent variable. The resulting standardized beta and standard errors of the beta indicate whether an effect size can reasonably be considered a true null effect. If the confidence interval of standardized beta includes zero, we conclude that there is no reason to reject the null hypothesis, that is, we conclude that the results indicate no difference between conditions. In the tables and text, the beta of the original models is referred to with “raw coefficient,” and the beta of the models ran over the standardized dependent variable is referred to with “standardized beta.” The values of  $t$  and  $p$  are the output of the original models.

## Results

The results of the first model are presented in Table 3; the description of categorical variables in the running text additionally lists estimated marginal means (EM, the means based on the model).

There was an interaction of familiarity by age: the 8-month-olds showed longer orientation times to the familiar (EM = 8.83) than the novel (EM = 7.60), but the 6-month-olds did not (EM familiar: 9.57; EM novel: 10.35). An additional unexpected finding is the interaction of familiarization condition by age: the 6-month-olds showed longer orientation times after the no-boundary (EM = 10.91) than after the with-boundary (EM = 9.02)

familiarization, but the 8-month-olds did not (EM no boundary: 7.54; EM with boundary: 8.90).

Table 3. Raw coeff., st.  $\beta$ , standard error of st.  $\beta$ , t and p for the fixed parameters, variance and standard deviation of random intercept and slopes, of the main model of Experiment 1.

<b>Fixed Parameter<sup>a</sup></b>	<b>Raw coefficient</b>	<b>Standardize d <math>\beta</math></b>	<b>Std. Error of st. <math>\beta</math></b>	<b>t</b>	<b>p<sup>b</sup></b>
Intercept (grand mean)	9.09	0	0.05	28.92	<.001 ***
Familiarity (familiar; novel)	-0.22	-0.03	0.05	-0.64	.520
Familiarization condition (no-boundary; with boundary)	-0.26	-0.04	0.10	-0.41	.683
Age (6; 8)	-1.74	-0.27	0.10	-2.59	.012 *
Trial order	-0.49	-0.08	0.00	-8.89	<.001 ***
Familiarization time	1.05	0.16	0.05	3.04	.003 **
Familiarity * familiarization condition	-1.20	-0.19	0.11	-1.69	.092 .
Familiarity * age	-2.01	-0.31	0.12	-2.67	.008 **
Familiarization condition * age	3.25	0.51	0.19	2.60	.011 *
Familiarity * Trial order	0.14	0.02	0.02	1.36	.178
Familiarity * familiarization time	-0.48	-0.07	0.06	-1.24	.217
Trial order * familiarization time	-0.16	-0.02	0.01	-2.81	.006 **
Familiarity * familiarization condition * age	2.06	0.32	0.22	1.47	.141
Familiarity * trial order * familiarization time	-0.03	-0.00	0.02	-0.30	.762
<b>Random intercept and slopes</b>		<b>Variance</b>		<b>Std. Deviation</b>	
Participant – intercept		5.50		2.35	

.. – trial order slope	0.05	0.21
.. – familiarity slope	0.08	0.28
.. – trial order * familiarity	0.07	0.27

slope

- Categorical variables (two levels) are contrast coded as -.5 and .5 in the order of the levels given. A positive st.  $\beta$  thus indicates a negative value for the first level, and a positive value for the second level.
- p-values for this and the following experiments were calculated with lmerTest (Kuznetsova et al., 2016) which uses Satterthwaite approximations to degrees of freedom. Significance codes: ‘\*\*\*’ <0.001; ‘\*\*’ < 0.01; ‘\*’ < 0.05; ‘.’ < 0.1; ‘ ’ < 1.

Given the interaction of familiarity by age, we ran separate models per age. The analysis for the 6-month-olds only revealed an interaction trend of familiarity and familiarization condition ( $t = -1.95$ ,  $p = .053$ , Raw coeff. = -2.10; st.  $\beta = -0.30$ , CI st.  $\beta = -0.46$  to -0.15; Figure 1, left panel) originating from significant longer orientation times during the novel (EM = 11.97) than during the familiar trials (EM = 10.34;  $t = 3.15$ ,  $p = .002$ , Raw coeff. = 1.63; st.  $\beta = 0.23$ , CI st.  $\beta = 0.13$  to 0.34) for infants familiarized in the no-boundary condition but not for those familiarized in the with-boundary condition (EM familiar: 9.79; EM novel: 9.32). The analysis for the 8-month-olds only revealed a main effect of familiarity ( $t = -2.22$ ,  $p = .028$ , Raw coeff. = -1.05; st.  $\beta = -0.19$ , CI st.  $\beta = -0.27$  to -0.10), with longer orientation times during the familiar (EM = 8.36) than the novel trials (EM= 7.31; Figure 1, right panel).



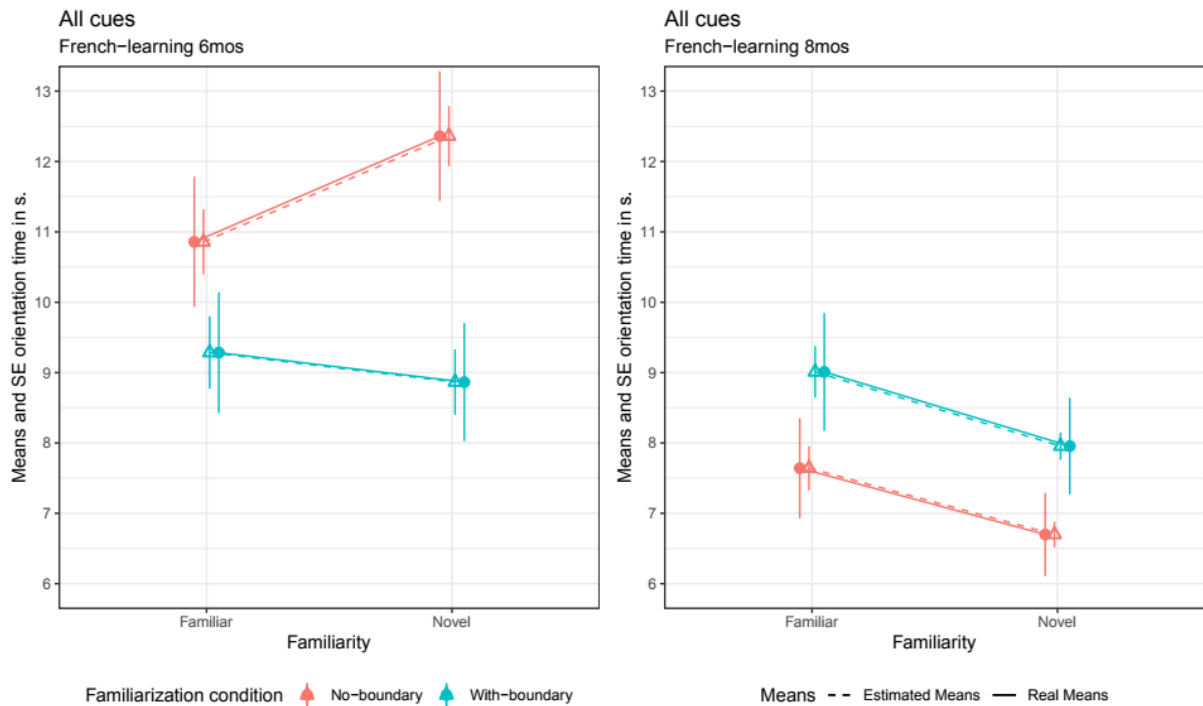


Figure 1. French-learning infants' orientation times to the familiar versus novel all cue stimuli broken down by age (left: 6-month-olds; right: 8-month-olds)

## Discussion

The results of Experiment 1 establish that both 6- and 8-month-old French-learning infants can discriminate between sequences of names, both those that include and those that do not include a major prosodic boundary, when all the boundary cues are presented. This extends previous findings on early sensitivity to major prosodic boundaries in infants learning Germanic languages (for English: Seidl & Cristià, 2008; for German: Wellmann et al., 2012; for Dutch: Johnson & Seidl, 2008) to infants learning a Romance language.

Our results further demonstrate developmental differences in the performance of French-learning 6- and 8-month-olds. The 6-month-olds showed a mixed pattern of effects: They displayed a novelty effect after familiarization to the no-boundary sequences, but there was no effect after familiarization to the with-boundary sequences. The former result suggests that they discriminate the two types of sequences. However, the lack of an effect in the latter condition indicates that their behavior in the test phase cannot be exclusively attributed to the familiarity of stimuli. There must be additional factors that modulate the orientation times, such as a spontaneous preference for some intrinsic properties of the with-boundary stimuli, or an asymmetry caused by their order of presentation. For example, the with-boundary

stimuli might be more interesting after listening to the no-boundary sequences than vice versa. We will return to this issue in the general discussion. In contrast, the 8-month-olds showed a main effect of familiarity (not interacting with familiarization condition). The difference between ages could indicate that the 8-month-olds have overcome a spontaneous preference for attending to with-boundary stimuli.

The present results are compatible with the results of prior studies with German-learning 6- and 8-month-olds (Höhle et al., 2013; Wellmann et al., 2012), given that both age groups in both languages perceive a difference between the conditions. However, the results are also different across the two languages: both German-learning age groups displayed a novelty preference only after familiarization with no-boundary stimuli – the pattern that was also observed in the French 6-month-olds – while the French-learning 8-month-olds showed a familiarity preference after both familiarizations. This constitutes a potential cross-linguistic difference in prosodic boundary processing at 8 (but not 6) months of age. To further explore these differences, Experiment 2 investigated the role of pausing in the discrimination of prosodic boundaries by French-learning 6- and 8-month-olds, since German-learning 8- but not 6-month-olds were previously shown to be able to detect the boundary marked by pitch and lengthening alone.

## **Experiment 2**

### **Method**

#### *Participants*

Forty 6-month-old (21 girls, mean age = 6.47 months; range: 5.78 – 6.87) and 40 8-month-old (16 girls, mean age = 8.57 months; range: 8.05 – 9.03) French-learning monolingual infants were tested at the babylab of the Laboratoire Psychologie de la Perception, Paris (currently the Integrative Neuroscience and Cognition Center). An additional 15 6-month-olds were tested and excluded due to fussiness (12), technical problems (2), or parental interference (1). An additional 12 8-month-olds were tested and excluded due to fussiness (11) or technical problems (1).

#### *Stimuli*

Stimuli were constructed based on those of Experiment 1. Importantly, as opposed to Experiment 1, there was no pause cue in the with-boundary sequences, and the only cues

distinguishing the two types of stimuli were pitch and duration of the final vowel of the second name. This resynthesis was chosen to ensure maximum comparability with Wellmann et al. (2012) study that used the same resynthesis for their German stimuli without pauses with German-learning infants. Stimuli were created in exactly the same way as in Wellmann et al. (2012): the no-boundary stimuli of Experiment 1 were phonetically resynthesized for two new conditions using Praat (Boersma & Weenink, 2016). For the no-boundary condition, the pitch contour was stylized to make it comparably synthetic to that of the resynthesized with-boundary condition. For the with-boundary condition, final lengthening and pitch rise (with values corresponding to the naturally produced stimuli from Experiment 1, see Table 2) were added to these no-boundary stimuli on the second name, that is, the name before the boundary.

#### *Procedure, apparatus, design, and analysis*

As in Experiment 1.

### **Results**

Model results are shown in Table 4. In contrast to Experiment 1, there were no main effects or interactions involving familiarity (see Figure 2). To ensure that the null result of all parameters involving familiarity are true null effects, we inspected the confidence interval (CI) of st.  $\beta$ . We found that the null result of familiarity is a true null effect, since the CI of st.  $\beta$  includes zero (CI st.  $\beta$  = -0.08 to 0.03). For the same reason, the interaction of familiarity and familiarization condition (CI st.  $\beta$  = -0.21 to 0.01) and familiarity, familiarization condition and age (CI st.  $\beta$  = -0.14 to 0.31) are also true null effects. The confidence interval of st.  $\beta$  for the interaction of familiarity and age does not include zero (CI st.  $\beta$  = 0.02 to 0.25), which means that an effect could potentially surface if the study were higher-powered. Hence, we split the model by age. We found that there was no significant effect of familiarity for either the 6- (t = -1.16, p = 0.25, Raw coeff. = -0.49, st.  $\beta$  = -0.09, CI st.  $\beta$  = -0.17 to -0.01) or the 8-month-olds (t = 0.62, p = 0.54, Raw coeff. = -0.25, st.  $\beta$  = 0.05, CI st.  $\beta$  = -0.03 to 0.13), and that the CI of the st.  $\beta$  for the 8-month-olds includes zero, but that for the 6-month-olds does not.

Table 4. Raw coeff., st.  $\beta$ , standard error, t and p for the fixed parameters, variance and standard deviation of random intercept and slopes, of the main model of Experiment 2.

<b>Fixed Parameter<sup>a</sup></b>	<b>Raw</b>	<b>st. <math>\beta</math></b>	<b>Std.</b>	<b>T</b>	<b>P<sup>b</sup></b>
------------------------------------	------------	-------------------------------	-------------	----------	----------------------

	coeff.		Error of		
			st. $\beta$		
Intercept (grand mean)	7.53	0	0.05	26.89	<.001 ***
Familiarity (familiar; novel)	-0.12	-0.02	0.06	-0.41	.684
Familiarization condition (no-boundary; with boundary)	-0.59	-0.11	0.11	-1.06	.292
Age (6; 8)	0.68	0.13	0.11	1.16	.249
Trial order	-0.33	-0.06	0.01	-6.33	<.001 ***
Familiarization time	1.08	0.21	0.06	3.66	<.001 ***
Familiarity * familiarization condition	-0.51	-0.10	0.11	-0.87	.384
Familiarity * age	-0.70	0.13	0.12	1.15	.249
Familiarization condition * age	-0.09	-0.02	0.21	-0.08	.934
Familiarity * trial order	0.00	0.00	0.02	0.01	.990
Familiarity * familiarization time	-0.26	-0.05	0.06	-0.84	.399
Trial order * familiarization time	-0.17	-0.03	0.01	-3.26	.002 **
Familiarity * familiarization condition * age	0.43	0.08	0.22	0.36	.716
Familiarity * trial order * familiarization time	0.02	0.00	0.02	0.29	.774
<b>Random intercept and slopes</b>			<b>Variance</b>	<b>Std. Deviation</b>	
Participant – intercept			4.61	2.15	
.. – trial order slope			0.09	0.29	
.. – familiarity slope			0.17	0.42	
.. – trial order * familiarity slope			0.00	0.05	

- Categorical variables (two levels) are contrast coded as -.5 and .5 in the order of the levels given. A positive st.  $\beta$  thus indicates a negative value for the first level, and a positive value for the second level.
- p-values for this and the following experiments were calculated with lmerTest (Kuznetsova et al., 2016) which uses Satterthwaite approximations to degrees of freedom. Significance codes: ‘\*\*\*’ <0.001; ‘\*\*’ < 0.01; ‘\*’ < 0.05; ‘.’ < 0.1; ‘’ < 1.

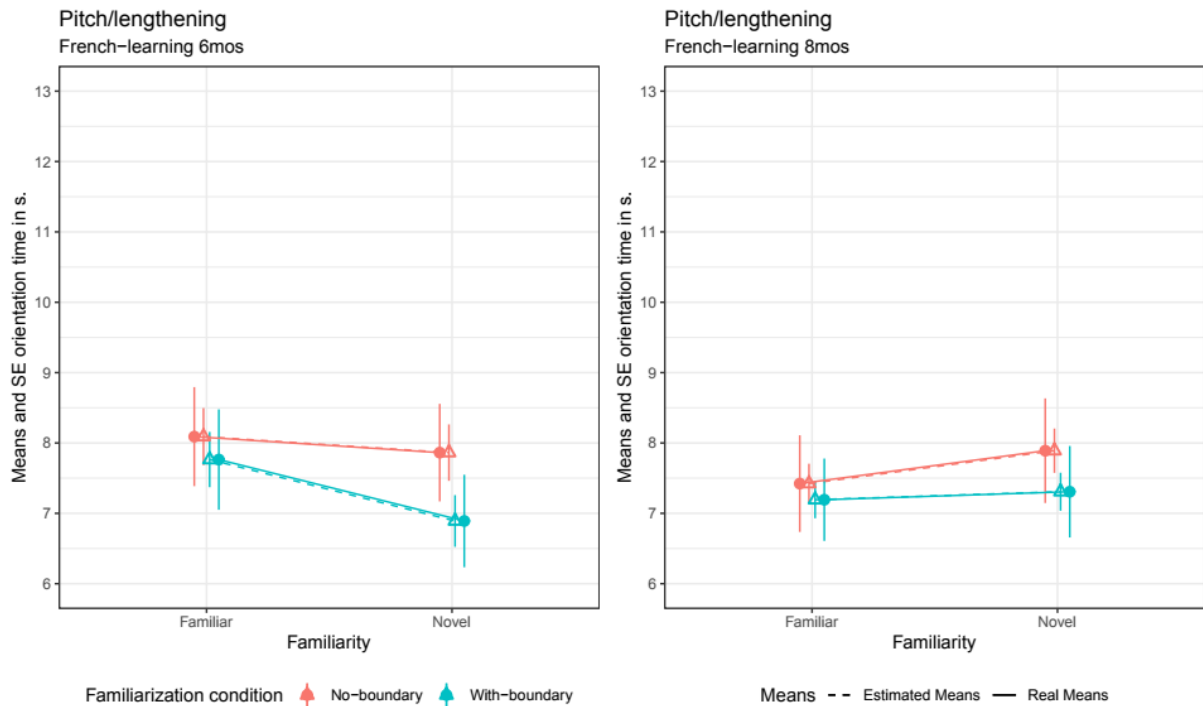


Figure 2. French-learning infants' orientation times to the familiar versus novel pitch/final lengthening stimuli (Experiment 2), broken down by age (left: 6-month-olds; right: 8-month-olds)

## Discussion

Experiment 2 failed to show that either 6- or 8-month-old French-learning infants discriminate our “pitch/final lengthening” stimuli from stimuli without boundary: there was no significant effect or interaction involving familiarity. To investigate whether we can maintain the null hypothesis (i.e., whether our data present no reason to reject it), we inspected the confidence intervals of our standardized effects. This inspection revealed no reason to reject the null hypothesis that our French-learning 8-month-olds respond similarly to the familiar and the novel test items. The null results of the familiarity-age interaction as well as the lack of a main effect of familiarity for the 6-month-olds may raise the question of whether a higher-powered study would find a result. However, in comparison, Wellmann et al. (2012) found a familiarity effect in a highly comparable study with 8-month-olds ( $d_z = 0.90$ ) and an evaluation of this effect using GPower (Faul, Erdfelder, Lang, & Buchner, 2007) indicates that an effect of this size would have been found with 0.99 probability with our sample size of 20. Hence, we can maintain the null hypothesis that there is no difference in orientation times between the two types of test items, and furthermore, that any potentially

missed effect in the 6-month-olds is considerably smaller than the previously reported effect for the German-learning 8-month-olds.

At this point, there are alternative explanations for why the French-learning 8-month-olds did not detect the prosodic difference between the test items while the German-learning 8-month-olds did. The cross-linguistic difference in infants' performance could be evidence of language-specific perceptual specialization. On the other hand, acoustic properties of the language-specific boundary implementation in the stimuli could lead to differences in their perceptibility. To assess this, we compared the prosodic salience of the prosodic boundary cues of the French stimuli used in the current study and the German ones used by Wellmann et al. (2012). We measured four acoustic correlates of interest in the natural and the pitch-and-lengthening-only stimuli using Praat (Boersma & Weenink, 2016), see Table 5.

Table 5. Acoustic correlates measured on the final and prefinal syllable of the phrase.

<b>Acoustic dimension</b>	<b>Correlate</b>	<b>Implementation</b>
Pitch	<b>Rise</b>	Semitone difference between the minimum and the maximum pitch on prefinal and final vowel in Name 2.
	<b>Reset</b>	Semitone difference between the maximal pitch in final vowel of Name 2 and the minimal pitch in the following conjunction (et/und) in case of a fall, and the maximum pitch in the same following conjunction in case of a rise
Duration	<b>Lengthening</b>	Durational difference in ms between final and prefinal vowel of Name 2
	<b>Pause</b>	Silence duration after Name 2

Figure 3 and Table 6 show that the major prosodic boundary is marked in similar ways in the naturally produced German and French stimuli. Each of the measured cues contributes to the contrast between conditions. In the natural recordings, the stimuli of both languages have a higher rise on Name 2<sup>3</sup>, a lower reset, more lengthening and a pause in the with-

<sup>3</sup> Visual inspection of Figure 3 additionally suggests that in the French stimuli, there is also a pitch rise on Name 1 in both the no-boundary and the with-boundary conditions (the latter being smaller), while a similar pitch rise is only present for the no-boundary condition in German. This prosodic pattern might additionally

boundary condition as compared to the no-boundary condition. The mean values of these measures however, suggest that the difference between conditions is somewhat larger in the German than in the French stimuli in rise, reset and lengthening, while pausing is longer in the French than in the German stimuli. In sum, in both languages the two conditions have become phonetically less distinctive after the resynthesis, but the acoustic difference between the phrase type conditions in the stimuli without pauses seems to be bigger in the German than in the French stimuli.

Table 6. Mean values of phonetic measures of stimuli in each condition.

<b>Language</b>	<b>Experiment</b>	<b>Phrase type</b>	<b>Rise</b>	<b>Reset</b>	<b>Lengthening</b>	<b>Pause</b>	
French	All cues	No-boundary	6.76	-1.42	14.49	0	
		With-boundary	11.68	-4.7	49.21	652.38	
	Pitch and lengthening	No-boundary	6.21	-0.69	14.49	0	
		With-boundary	12.46	-2.16	64.06	0	
	German	All cues	No-boundary	6.47	0.67	23.99	0
			With-boundary	14.01	-4.13	99.4	499.19
Pitch and lengthening		No-boundary	5.16	2.13	25.93	0	
		With-boundary	13.38	-7.04	115.64	0	

---

increase the perceptibility of the difference between the no- and with-boundary conditions in the German stimuli compared to the French ones.

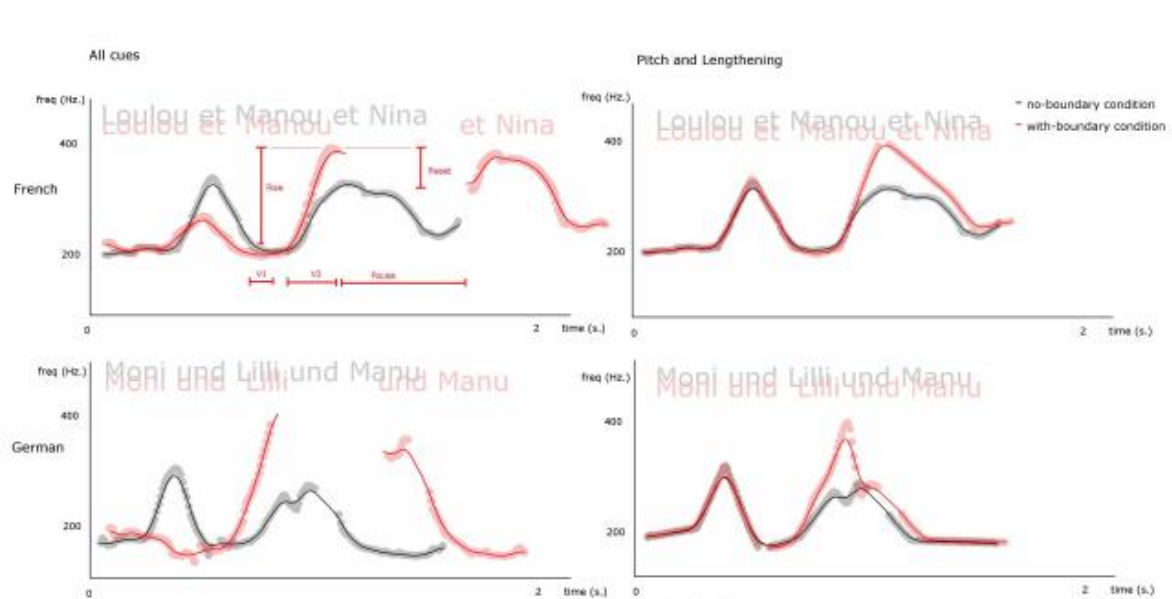


Figure 3. Pitch tracks of each of the prosodic conditions in each language (top panels: French; bottom panels: German) and experiment (left panels: Exp.1/all cues; right panels: Exp.2/pitch and lengthening). The figure was made using Mausmooth (Cangemi, 2015) in Praat (Boersma & Weenink, 2016). The French stimuli were used in the current study, the German stimuli were used in Wellmann et al. (2012).

To investigate whether the differences in discrimination performance between French- and German-learning infants 1) result from these subtle measurable phonetic differences between the sets of stimuli or 2) relate instead to language-specific perception, we presented the French stimuli to German-learning infants. Thus, in Experiment 3, German-learning 6- and 8-month-olds were tested with the same French stimuli as in Experiment 2. At 6 months, both language groups are expected to behave similarly, as language-specific perception of prosody is not expected at this age yet. At 8 months, one of two different outcomes can be expected. If the differences in behavior between our French-learning infants and the German-learning infants in Wellmann and colleagues' studies were caused by stimuli-specific differences (i.e., because boundaries were more saliently produced in the German stimuli), then German-learning 8-month-olds should behave like the French-learning 8-month-olds when presented with the same French materials. If, however, this difference in behavior is related to a language-specific sensitivity to prosodic detail, then German-learning infants should show better discrimination performance than the French-learning infants when hearing the French stimuli.

### Experiment 3



## Method

### *Participants*

Forty 6-month-old (21 girls,  $M_{\text{age}} = 6.52$  months; range: 6.00 – 7.07) and 40 8-month-old (19 girls,  $M_{\text{age}} = 8.41$  months; range: 7.95-9.03) German-learning monolingual infants were tested at the Potsdam BabyLAB. An additional 9 6-month-olds were tested and excluded due to fussiness (6), having two consecutive trials with insufficient ( $< 1500$  ms) orientation times (1), or technical problems (2). An additional 10 8-month-olds were tested and excluded due to fussiness (8), having two consecutive trials with insufficient orientation times (1), or technical problems (1). All infants were without apparent health problems and not at higher risk of developmental or linguistic disorders, and they had been born after at least 37 weeks of gestation. They were recruited from birth lists obtained through the Potsdam city hall archives or through a database of families who had previously participated in speech perception studies in the laboratory. Informed written consent was obtained from all parents.

### *Stimuli*

The same as in Experiment 2.

### *Procedure, apparatus and design*

As in Experiment 2, except for the following differences. Trials were not repeated when orientation times were under 1500 ms. and familiarization was stopped manually after about 60 seconds.

### *Analysis*

As in Experiment 2.

## Results

Model results are given in Table 7. There is a main trend ( $p = .07$ ) of familiarity, with longer orientation times to familiar ( $EM = 7.44$ ) than to novel ( $EM = 6.92$ ) items, and a trend ( $p = .052$ ) of an interaction of familiarity and age in the direction of longer looks to the familiar ( $EM = 7.88$ ) than the novel ( $EM = 6.78$ ) items for the 8-month-olds but not the 6-month-olds ( $EM$  familiar: 7.00;  $EM$  novel: 7.06). A simple-slope comparison confirmed that the effect of familiarity was significant for the 8-month-olds ( $t = -2.65$ ,  $p < 0.01$ , Raw coeff. = -1.09, st.  $\beta$

= 0.23, CI st.  $\beta$  = -0.31 to -0.14), but not the 6-month-olds ( $t = 0.15$ ,  $p < 1$ , Raw coeff. = 0.06, st.  $\beta$  = 0.01, CI st.  $\beta$  = -0.07 to 0.01). The fact that the confidence interval of beta at 6 months includes zero suggests that our results do not give us reason to reject the null hypothesis. See Figure 4 for an illustration of this effect.

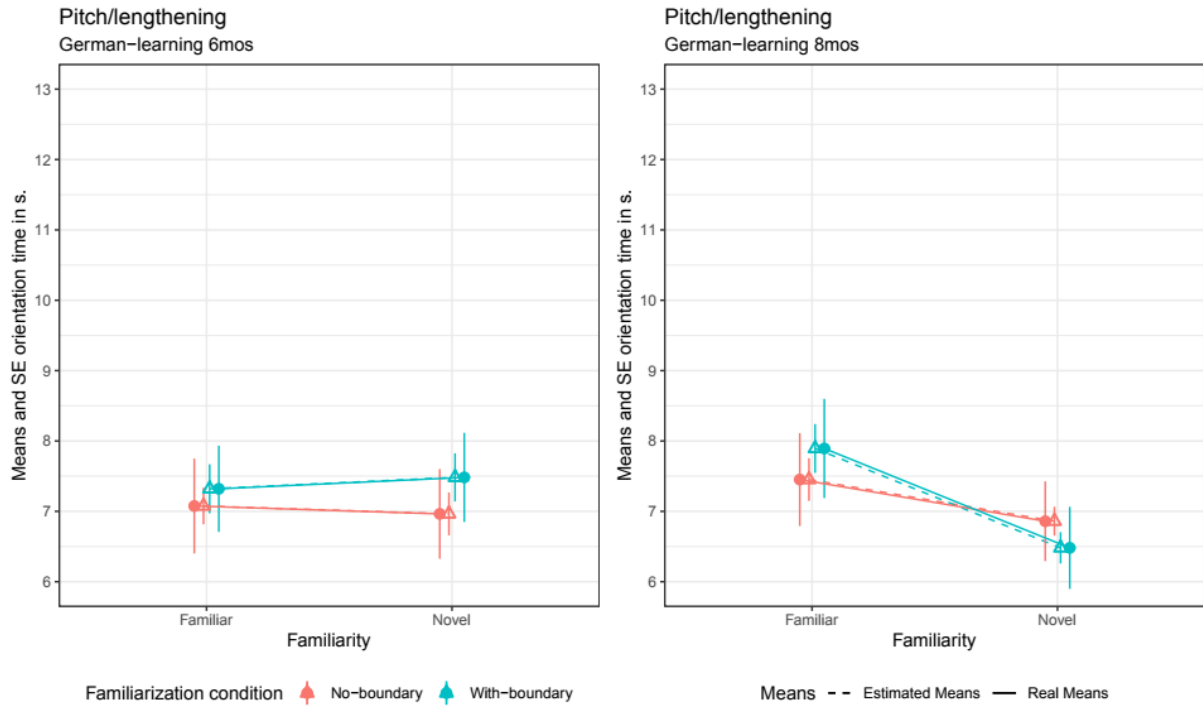


Figure 4. German-learning 6- and 8-month-olds' orientation times to the familiar versus novel stimuli in the "pitch/final lengthening" experiment.

Table 7. Raw coeff., st.  $\beta$ , standard error,  $t$  and  $p$  for the fixed parameters, variance and standard deviation of random intercept and slopes, of the main model of Experiment 3.

Fixed Parameter <sup>a</sup>	Raw coeff.	st. $\beta$	Std. Error of st. $\beta$	$t$	$p^b$
Intercept (grand mean)	7.18	-0.00	0.05	28.59	<.001 ***
Familiarity (familiar; novel)	-0.51	-0.11	0.06	-1.82	.070 .
Familiarization condition (no-boundary; with boundary)	0.03	0.01	0.10	0.07	.943
Age (6; 8)	0.30	0.06	0.10	0.60	.553
Trial order	-0.30	-0.06	0.01	-6.12	<.001 ***
Familiarization time	0.65	0.14	0.06	2.44	.017 *
Familiarity * familiarization	-0.27	-0.06	0.12	-0.47	0.638

condition					
Familiarity * age	-1.15	-0.24	0.12	-1.95	.052 .
Familiarization condition * age	0.31	0.06	0.20	0.33	.744
Familiarity * trial order	0.00	0.00	0.02	0.03	.975
Familiarity * familiarization time	-0.27	-0.06	0.06	-0.89	.375
Trial order * familiarization time	-0.15	-0.03	0.01	-3.02	.003 **
Familiarity * familiarization condition * age	-1.09	-0.22	0.23	-0.96	.339
Familiarity * familiarization condition * age	-0.05	-0.01	0.02	-0.59	.556
condition * familiarization time					
Random intercept and slopes	Variance			Std. Deviation	
Participant – intercept	0.150			0.387	
.. – trial order slope	0.003			0.053	
.. – familiarity slope	0.015			0.121	
.. – trial order * familiarity slope	0.001			0.030	

- Categorical variables (two levels) are contrast coded as -.5 and .5 in the order of the levels given. A positive st.  $\beta$  thus indicates a negative value for the first level, and a positive value for the second level.
- p-values for this and the following experiments were calculated with lmerTest (Kuznetsova et al., 2016) which uses Satterthwaite approximations to degrees of freedom. Significance codes: ‘\*\*\*\*’ <0.001; ‘\*\*\*’ < 0.01; ‘\*\*’ < 0.05; ‘.’ < 0.1; ‘ ’ < 1.

According to these results, German-learning 6-month-olds did not discriminate the two conditions in the test while the 8-month-olds did. Such a difference between groups was not observed between 6- and 8-month-old French-learning infants (Exp. 2). To clarify whether the results of the German- versus French-learning infants are indeed different, we ran a model with the data from both Experiments 2 and 3. The structure of the model was the same as that of all previous models, except for the added predictor of ‘language’ (French/German).

This model revealed a significant interaction ( $p = .03$ ) of familiarity, age and language. To further inspect this three-way interaction, we calculated two separate models over subsets of the data split by age and found a significant interaction of familiarity and language for the 8-month-olds ( $p = .012$ ), but not the 6-month-olds ( $p = .92$ ). At 8 months, German-learning infants had longer orientation times to the familiar (EM = 8.10) than the novel (EM = 6.71) items, while French-learning infants did not (EM familiar: 6.88; EM novel: 7.51). A simple-

slope comparison confirmed that the difference between conditions was significant for the German-learning 8-month-olds ( $t = -2.75$ ,  $p < 0.01$ , Raw coeff. =  $-1.38$ , st.  $\beta = 0.28$ , CI st.  $\beta = -0.38$  to  $-0.18$ ), but not for the French-learning 8-month-olds ( $t = -1.28$ ,  $p < 0.5$ , Raw coeff. =  $0.63$ , st.  $\beta = 0.13$ , CI st.  $\beta = 0.03$  to  $0.23$ ). Figure 5 illustrates these results.

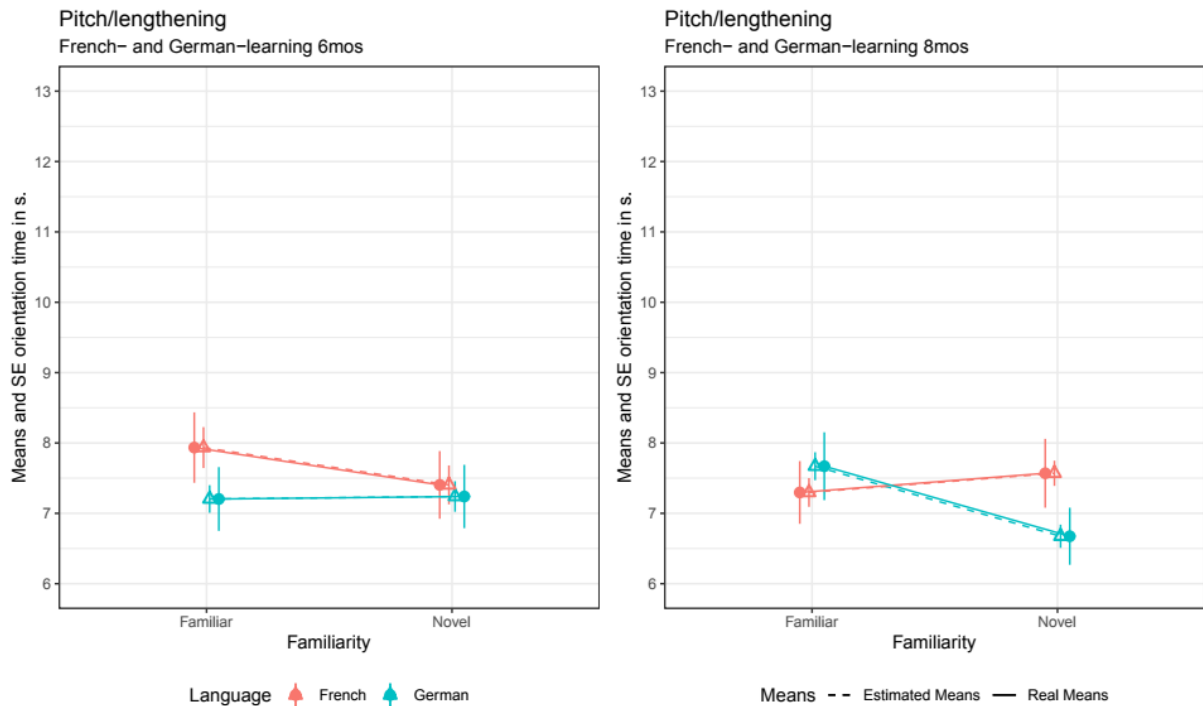


Figure 5. French- and German- learning 6- and 8-month-olds' orientation times to the familiar versus novel stimuli in the "pitch/final lengthening" experiment.

## Discussion

This experiment with French stimuli shows that German-learning 8-month-olds, but not 6-month-olds, discriminate between sequence types in which prosodic boundaries are marked only by pitch and final lengthening. The CI of st.  $\beta$  for the 6-month-olds confirmed that we have no reason to reject the null hypothesis.

More importantly, the cross-linguistic comparison of German- and French-learning infants in one model confirmed the cross-linguistic difference in the expected direction: At 6 months, there was no difference between French- and German-learning infants, and neither of the two groups showed discrimination of the two sequence types when boundaries were marked only by pitch and final lengthening. However, at 8 months, there was a difference between French- and German-learning infants, and only the German- and not the French-learning infants discriminated between the two sentence types. This establishes that German-learning infants are able to process major phrase boundary information in a foreign language

with different prosodic properties. It also suggests that the non-responsiveness of the French-learning infants to boundaries only marked by pitch and final lengthening was not due to weaker boundary marking in the French stimuli compared to the German. The cross-linguistic difference that this result constitutes is in line with our expectations: French-learning infants acquire a regular phrasal prosodic system for which it is not necessary to pay close attention to phonetic/prosodic detail, which may cause a weaker perceptual discrimination of this detail. At this point, however, it is not clear whether the French-learning infants' development of the ability to discriminate this subtle phonetic difference is merely delayed, or whether a low sensitivity is also found in French-speaking adults. Hence, Experiment 4 tested German- and French-speaking adults' discrimination of the same French stimuli.

## **Experiment 4**

### **Method**

#### *Participants*

Twenty French (mean age = 23.1, range = 20-34) and twenty German listeners (mean age = 24, range = 19-32) participated in the experiment voluntarily or for course credits. All listeners were native speakers of French or German, respectively, and had learned any second or third languages only after the age of four years. Participants had normal hearing and reported no language disorders and (corrected to) normal vision. French listeners were tested in Paris, and German listeners in Potsdam.

#### **Stimuli**

The same four tokens of each of the no-boundary and the with-boundary condition as created for Experiment 2 (only pitch and lengthening as boundary markers) were used, resulting in 8 stimuli overall, each containing only one repetition of the name sequence.

#### *Procedure, apparatus and design*

We used an AXB paradigm, in which listeners heard three stimuli consecutively; Stimulus A from one condition, Stimulus B from the other condition, and Stimulus X from either condition A or B. The three stimuli in one trial were never identical tokens. The participant was asked to respond whether Stimulus X was more similar to A or to B. Response condition and reaction time were measured.

The order of presentation of the stimuli within a trial was counterbalanced, resulting in 96 unique trials. Trials were presented in randomized order with EPrime software. The experiment started with 4 practice trials in which Stimulus X was the same token as either A or B. On the screen, participants saw three circles in a line which lit up during stimulus presentation: the left circle lit up during A, the middle during X and the right during B. The response screen had three circles on the left and three circles on the right. On the left of the screen, the middle and left circle were lighter ( o o o ) to depict the response that X was more like A, and B was the odd one out. Correspondingly, on the right of the screen, the middle and right circle were lighter ( o o o ) to depict the response option that X was more like B, and A was the odd one out. Participants pressed the left arrow key on the keyboard to indicate that X was more like A, and the right arrow key to indicate that X was more like B.

Participants were seated alone in a quiet room, and stimuli were presented over headphones. During the instructions and the practice phase, the experimenter was present to make sure that the instructions were understood and correctly executed. The experiment took around 20 minutes (without a break).

### *Analysis*

Results were analyzed with a (generalized) linear mixed-model analysis using R (R Core Team, 2016), lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2016), and plots were made using ggplot2 (Wickham, 2009). Listeners' responses (no-boundary; with-boundary) are predicted by participant language (French; German), the category of Stimulus X (no-boundary; with-boundary stimuli) and the category of Stimulus A (no-boundary; with-boundary stimuli) and all interactions. The Stimulus A variable was included to control for the possibility that order of presentation affects listeners' responses. The model additionally included individual participant intercepts and token intercepts, as well as participant slope of the category of Stimulus X and Stimulus A and their interaction, in order to account for individual variation. Levels of all categorical factors were centered, so the intercept of the model represented the grand mean.

*Formula: Responses ~ language \* stimulus X \* stimulus A) + (1 + stimulus X \* stimulus A / participant) + (1 | token X)*

### **Results**

Mean error rate for German listeners was 18.8% (range 1.0 to 49.0%), versus 16.4% for French listeners (range 1.0 to 46.9%).

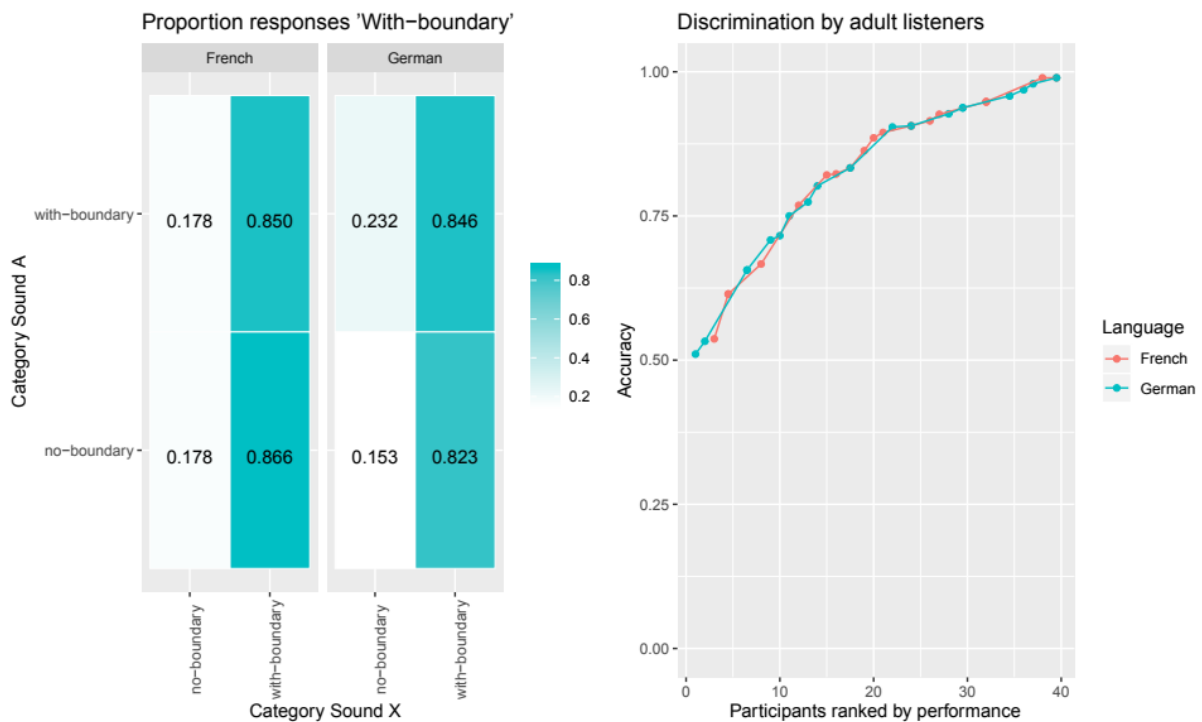


Figure 6 (Left) Proportion of “with-boundary”-responses by Category of sound A and sound X and native language; (Right) Individual variability in French and German adult perception and categorization of prosodic boundaries marked with pitch and duration.

Figure 6 shows the proportion of with-boundary responses per stimulus combination and language (left) and the range of proportion correct by participants and language (right). First, there was no significant main effect of language (Raw coeff. = 0.07, SE = 0.12,  $t = 0.574$ ,  $p = .567$ ), which suggests that there is no difference in preferred response (with-boundary or no-boundary sequence) between the German and French participants. Second, there was no significant interaction of language with category of Stimulus X (Raw coeff. = -0.14, SE = 0.76,  $t = -0.18$ ,  $p = .855$ ), suggesting no difference in accuracy between the two language groups. However, there was a significant interaction of language and category of Stimulus A (Raw coeff. = 0.63, SE = 0.29,  $t = 2.141$ ,  $p = .032$ ). Upon closer inspection by splitting the models by language, this effect reflects a higher number of with-boundary responses by German listeners when Stimulus A was a with-boundary stimulus type (Raw coeff. = 0.39, SE = 0.13,  $t = 3.04$ ,  $p = .002$ ). There was no such effect for French listeners (Raw coeff. = -0.07, SE = 0.13,  $t = -0.55$ ,  $p = .583$ ).

## **Discussion**

This experiment revealed no difference between German and French adult listeners in discrimination of the subtle prosodic difference between sequence types. Hence, it is likely that the language-specific differences in discrimination ability found at 8 months of age may be indicative of a slower development in sensitivity to prosodic detail in French-learning infants. However, two interesting observations arise from the adult results. First, the level of discrimination in adults in both languages is highly variable, ranging from around 50% (chance level) to 100% correct responses in both languages. This can be seen in Figure 7. So, not all adult listeners are sensitive to prosodic detail, which indicates that other factors like phonetic and musical acuity, second-language experience, and musical experience may influence their perceptual abilities (see Boll-Avetisyan, Bhatara, & Höhle, 2017; Boll-Avetisyan, Bhatara, Unger, Nazzi, & Höhle, 2016, where such factors accounted for individual variability in speech rhythm perception).

Second, the interaction with order of presentation found for German listeners could indicate a language-specific perceptual asymmetry. As we also found an asymmetry in infants' perception depending on the familiarization condition, we present a discussion of the present asymmetries further below.

## **General discussion**

The current study tested the hypothesis that phrase-level prosody perception is subject to language-specific perceptual attunement. To achieve this, we compared the perception of prosodic boundaries by French- and German-learning infants, that is, learners of two prosodically distinct languages. The results are summarized in Table 8.

At 6 months, French-learning infants behaved like the German-learning 6-month-olds in Höhle et al. (2013) and the ones in the present study. This confirms our hypothesis that in the first months of life, prosodic boundaries are perceived in a language-general way. At 8 months, in contrast, French-learning infants did not discriminate between sequences with and without a prosodic boundary when this boundary was not marked by a pause, whereas German-learning infants were able to discriminate between these sequences, both in their native language (Wellmann et al., 2012) and in the same French stimuli for which French infants failed. By means of this highly-controlled cross-linguistic design, we establish for the



first time that changes in the processing of prosodic boundaries across language development reflect perceptual attunement to language-specific prosody.

Table 8. Summary of the orientation time preferences across the current experiments and Experiment 1 and 2 of Wellmann et al. (2012) and Höhle et al. (2013).

		All cues		Pitch and lengthening	
		French stimuli (current study)	German stimuli (Wellmann et al., 2012; Höhle et al., 2013) <sup>a</sup>	French stimuli (current study)	German stimuli (Wellmann et al., 2012; Höhle et al., 2013) <sup>a</sup>
<b>6mos</b>	<b>French-learning</b>	With-boundary <sup>b</sup>		None	
	<b>German-learning</b>		With-boundary <sup>c</sup>	None	None
<b>8mos</b>	<b>French-learning</b>	Familiarity		None	
	<b>German-learning</b>		With-boundary <sup>b</sup>	Familiarity	Familiarity <sup>c</sup>

- a. All results with German stimuli except for the 8-month-olds in the all-cues experiment are based on the no-boundary familiarization condition only.
- b. Preferred only after no-boundary familiarization condition, no effect after with-boundary familiarization condition
- c. Only tested after no-boundary familiarization condition.

In line with previous literature and our hypothesis, French-learning infants discriminated between utterances with and without major prosodic boundaries at 6 and 8 months when a pause was one of the cues to the major prosodic boundary (Experiment 1), but not when the boundaries were cued by pitch and lengthening alone (Experiment 2). This lack of discrimination for the 6-month-olds is in line with previous studies showing that younger infants typically depend heavily on the pause in the discrimination of phrase boundary perception (for German-learning 6-month-olds: Höhle et al., 2013; for English-learning 4-month-olds: Seidl & Cristià, 2008; for Dutch-learning 6-month-olds: Johnson & Seidl, 2008). However, the fact that the French-learning 8-month-olds did not discriminate in the pitch and lengthening experiment is in contrast with the German-learning 8-month-olds in Wellmann et al. (2012) and also with the English-learning 6-month-olds in Seidl (2007). Based on the

outcome of Experiment 3 that showed that German 8-month-olds can detect the prosodic boundary cues in the French materials without a pause, we conclude that our findings are direct evidence for a language-specific development in the perception of prosodic boundaries.

Regarding the developmental angle of our study, we show that the process of perceptual attunement in prosodic boundary perception is one of enhancement in German-learning infants. As was already implicit in previous findings, we see that between 6 and 8 months of age, infants develop an enhanced sensitivity to phrasal prosody: 8-month-olds hear a subtle difference that 6-month-olds do not. Our findings provide explicit evidence for this, and in addition show for the first time that this enhanced sensitivity is part of perceptual attunement and, as such, language specific. These findings are in concordance with other prosodic developmental processes. Enhancement has previously been documented for native tone perception (Tsao, 2017) and word stress discrimination (Weber, Hahne, Friedrich, & Friederici, 2004).

The fact that the French-learning infants did not discriminate the two prosodic structures in this study does not mean that they would not in another context or at another age. Regarding context, it is possible that discrimination would be found in a study focusing on acoustic discrimination (as opposed to phonological generalization as the current study). Because in the current study infants listened to different phonetic tokens of the same prosodic structure, some generalization had to be made to differentiate the conditions. In contrast, a paradigm without acoustic variability could directly test infants' lower-level perceptual abilities. Furthermore, it is possible that other (behavioral or neurophysiological) measures could uncover a more subtle perceptual sensitivity. However, this is not a limitation for the current study. Because we directly compared perception of the same stimuli between infants of two different languages using the same task/measure, we can confidently conclude that French-learning infants are significantly less sensitive than German-learning infants to phrasal prosodic detail at 8 months. Whether this means that they cannot discriminate these details or that they would discriminate them in another paradigm does not affect this conclusion.

Regarding the development after 8 months of age, it is possible that French-learning infants will learn to use pitch/lengthening cues and are simply delayed compared to German-learning infants. The results of our adult experiment point to this direction. While there was a great range of individual variation between listeners, there was no systematic difference between native speakers of German or French. From this simple experiment that attests

discrimination in both language groups, we tentatively conclude that, by adulthood, listeners of both languages reach a similar level of sensitivity to the prosodic detail we investigated, and individual differences are a more important source of variation than native language. Note, however, that it could still be the case that cross-linguistic differences in processing prosodic boundaries by French- vs. German-speaking adults would become evident with other stimuli. For comparability with the infant task, the adults in the present study heard the same stimuli as the infants. We could imagine that group differences may have arisen if, for example, German sequences were used, or if the task demanded higher cognitive load than the present study's stimuli. With higher cognitive load, pure acoustic-phonetic perception would be more difficult.

Regarding the language-specific angle of our study, we also found what we predicted. Our findings provide evidence for the emergence and enhancement of language-specific processing abilities. More specifically, we predicted that the higher prosodic variability in German relative to French would be a richer basis from which to develop sensitivity to prosodic detail. We expected French-learning infants to have poorer discrimination of subtle prosodic phrase boundaries, and this is indeed what we found. Thus, our current findings relate to other findings showing that adult speakers of French have difficulty lexically encoding word stress (Dupoux, Pallier, Sebastian, & Mehler, 1997; Peperkamp, Vendelin, & Dupoux, 2010) and using word prosody for rhythmic grouping of speech according to the ITL (Bhatara, Boll-Avetisyan, Agus, Höhle, & Nazzi, 2016; Bhatara, Boll-Avetisyan, Unger, Nazzi, & Höhle, 2013; Hayes, 1985, 1989) than speakers of languages that lexically encode prosodic information. In line with our findings, these cross-linguistic differences were not yet present in infants up to an age of 6-7 months (Abboub, Boll-Avetisyan, Bhatara, Höhle, & Nazzi, 2016; Bijeljac-Babic et al., 2012; Skoruppa et al., 2013, 2009).

One possibility is that prosodic regularity on a certain level induces less prosodic sensitivity at all levels (word stress, phrase level, prosodic grouping). Another possibility is that prosodic regularity on a certain level induces less prosodic sensitivity at that particular level only. In the prosodic typology of Jun (2014), languages are classified according to their prominence type, that is, their word-level prosody (micro rhythm) and their phrase-level regularity (macro rhythm). As outlined in the introduction, in this typology, French is a language without word stress and with strong macro rhythm, while German is classified as a language with word stress and medium macro rhythm. Based on the data showing modulation

of prosodic sensitivity across languages, it can be hypothesized that listeners/learners of a language with a strong macro rhythm (like French) have less need to pay attention to prosodic detail at the phrase level, while listeners/learners of a language with more variable prosody on several levels (like German) learn to become more attentive to prosodic detail, possibly earlier in development. To specify the different possibilities of how prosodic regularity at a given level affects prosodic sensitivity at that or other levels, we will need to extend the current work to more languages, contrasting those that vary in rhythmic timing, prominence type, and word- and phrase-level regularity, separately. The fact that English-learning infants could discriminate prosodic phrase boundaries without the pause at 6 months of age (Seidl, 2007) is in agreement with this hypothesis, because English, like German, is a language with medium macro rhythm (Jun, 2014).

Our findings additionally provide evidence that prosodic boundary cues can also be processed in non-native stimuli. In this first study testing infant non-native prosodic boundary perception, German-learning 8-month-olds could discriminate prosodic phrase boundaries in French, even when no pause was present. This indicates their ability to apply language-specific prosodic perception routines to non-native speech, either because German-learning 8-month-olds have become more sensitive to prosodic cues in general, or because the French prosodic boundaries were similar enough to German prosodic boundaries for the infants to perceptually assimilate them.

Our study and the work of Wellmann and colleagues covers the most important cells in Table 8 to be able to reach conclusions on development and language-specificity. We argue that the results from the empty cells can be predicted with our current results. First, the empty cells in the left half of the table concern 6- and 8-month-old infants' listening to non-native natural stimuli (including pauses). For these, we predict that for each cell there will be discrimination, since all studies involving pause cues to prosodic boundaries have shown that infants of all ages can detect them. Second, the empty cells in the right half of the table concern French-learning infants listening to German pitch- and lengthening stimuli. For these, we predict that neither 6- nor 8-month-olds would be able to discriminate these non-native stimuli. The reasoning is clear: if these infants did not develop sufficient sensitivity to prosodic detail in their own language at 8 months, it is highly unlikely that they will have this sensitivity in another language. The possibility remains that boundaries in the German stimuli are more discriminable than in the French stimuli, as suggested by our acoustic analysis. In

this case, 8-month-old infants may benefit from this (but not 6-month-olds, as even German-learning 6-month-olds did not discriminate these stimuli; Höhle et al., 2013). In any case, both outcomes would support the conclusion from our current study: at 8 months, German-learning infants are found to be more sensitive to prosodic detail than French-learning infants.

Before concluding, we would like to discuss one unexpected aspect of our results, namely the perceptual asymmetry in favor of the pause-cued boundary by French-learning 6-month-olds in Experiment 1. The same asymmetry was found by Höhle et al. (2013) and Wellmann et al. (2012) for German-learning infants tested with German materials, and interpreted as an asymmetrical preference for a more complex prosodic structure when it follows a simpler one, but not the other way around. However, we did not find the same asymmetry in the pitch and lengthening experiment for the German 8-month-olds, though the prosodic structure is the same. An alternative interpretation may be that the preference reflects an interaction of a natural preference for the with-boundary stimuli and a novelty preference. This would explain why a preference for the with-boundary stimuli does not surface when they are presented in familiarization. However, generally, novelty preferences are expected when experiments are less taxing for the infant, either because of lower task demands or because the infant is older (Hunter & Ames, 1988; Kidd, Piantadosi, & Aslin, 2012), and in this case we found this pattern only in the younger infants. Another possibility is that an acoustic asymmetry in perception differentially affects discrimination. Bharucha & Pryor (1986), for instance, found that two sequences are more discriminable when the first is rhythmic and the second is non-rhythmic than two sequences presented in the opposite order. It can be argued that rhythmic sequences build up concrete expectations that the following sequences will be equally rhythmic. If non-rhythmic sequences are less likely to build up similar expectations, this might cause an asymmetry by order of presentation. In the case of our prosodic boundary stimuli, the no-boundary stimulus can be argued to be more rhythmic than the with-boundary stimulus, since in the latter the rhythm is disrupted by a pause of around half a second. This order of events, that is, a (rhythmically disrupted) with-boundary stimulus after familiarization with the (rhythmic) no-boundary stimuli, might draw infants' attention more strongly than the opposite order. Carrying this idea over to a natural setting, exaggerating the pauses at phrase boundaries in IDS (Ludusan et al., 2016) might then be a way for adult speakers to attract learners' (infants') attention to these pauses, and thus facilitate the acquisition of linguistic structures aligned to the pauses, as proposed by prosodic bootstrapping theories (e.g., Seidl & Johnson, 2006). A second possibility is suggested by the

finding that adult German listeners also show a perceptual asymmetry. In a sequence of three stimuli, they were more likely to have difficulties in discrimination when the third had no boundary. In the adult experiment, the sequences did not notably differ in rhythmicity, as there was no pause cue in the sequence with boundary. An alternative explanation could thus be that stimuli with boundaries (irrespective of whether they contain a pause cue or not) are presenting a phonetically or phonologically more specified prosodic structure (as opposed to a more default structure), prompting a perceptual asymmetry as was found in earlier works on word stress and phonemic contrasts (Altwater-Mackensen, van der Feest, & Fikkert, 2014; Friederici, Friedrich, & Christophe, 2007; Hestvik & Durvasula, 2016). This speculative explanation would need to be tested in further research.

To conclude, the current findings show a developmental change from a language-general perception of prosodic boundaries, heavily relying on pausing at 6 months, to a language-specific difference in perception of these boundaries at 8 months. Importantly, they reveal differences in sensitivity to prosodic cues at the phrase level at 8 months between learners of French and German. These differences constitute the first direct evidence of cross-linguistic differences in phrase boundary perception in infancy and are interpreted to be due to differences in regularity in the prosodic systems of the two languages. Furthermore, the asymmetrical preference for with-boundary stimuli after a no-boundary familiarization raises questions concerning infants' attention to prosodic rhythmicity and regularity of timing. Along the same lines, the perceptual asymmetry in German adults also warrants further investigation. Further research should be aimed at these issues. We suggest three possible ways to continue this investigation: First, a strong, language-general interest for pauses could be scrutinized from a rhythmic perspective. Is the use of pausing in IDS a way to structurally direct infants' attention by use of basic rhythmic expectations? Second, a study on adult perception of prosodic phrase structures differing in phonological specificity could shed further light on perceptual asymmetries. Lastly, the influence of the prosodic regularity of the native language on the perception of prosodic detail could be further investigated from a structured typological perspective, to disentangle different levels of regularity, such as word stress, rhythm class, and macro rhythm discussed above. For example, cross-linguistic research comparing two languages differing only on macro rhythm (e.g., Spanish: word stress, strong macro rhythm and Portuguese: word stress, weak macro rhythm) would help us understand the developmental determinants of the cross-linguistic differences found here between French (no word stress, strong macro rhythm) and German (word stress, medium

macro rhythm), by working toward disentangling the different prosodic levels at play in prosodic perceptual tuning.

## Data availability

The data that support the findings of this study are available upon request, by contacting the corresponding author SvO. This procedure complies with the requirements of the institutional ethics approval, protecting the privacy of the participants in the study. The materials and scripts are available on <https://doi.org/10.17605/OSF.IO/RGHSK>.

## Acknowledgements

This research was supported by ANR-DFG grant “Multilevel prosodic processing in a crosslinguistic perspective” (ANR-13-FRAL-0010 to RB-B and TN; DFG Ho 1960/15-1 to BH). The authors wish to thank Nawal Abboub, Sylvie Margules, Chiara Boila, Elisabeth Metz, Maïke Riegel, Ragna Krug and Vanessa Löffler for their help with recruitment and testing, and all our speakers, parents and infants who have come to our labs and participated in this study.

## References

- Abboub, N., Bijeljac-Babic, R., Serres, J., & Nazzi, T. (2015). On the importance of being bilingual: Word stress processing in a context of segmental variability. *Journal of Experimental Child Psychology*, *132*, 111–120. <https://doi.org/10.1016/j.jecp.2014.12.004>
- Abboub, N., Boll-Avetisyan, N., Bhatara, A., Höhle, B., & Nazzi, T. (2016). An exploration of rhythmic grouping of speech sequences by French- and German-learning Infants. *Frontiers in Human Neuroscience*, *10*. <https://doi.org/10.3389/fnhum.2016.00292>
- Altvater-Mackensen, N., van der Feest, S. V., & Fikkert, P. (2014). Asymmetries in early word recognition: The case of stops and fricatives. *Language Learning and Development*, *10*(2), 149–178.



- Aslin, R. N., & Pisoni, D. B. (1980). Effects of early linguistic experience on speech discrimination by infants: A critique of Eilers, Gavin, and Wilson (1979). *Child Development, 51*(1), 107. <https://doi.org/10.2307/1129596>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Best, C. C., & 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. C., 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.
- Bharucha, J. J., & Pryor, J. H. (1986). Disrupting the isochrony underlying rhythm: An asymmetry in discrimination. *Perception & Psychophysics, 40*(3), 137–141. <https://doi.org/10.3758/bf03203008>
- 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), e30843. <https://doi.org/10.1371/journal.pone.0030843>
- Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer [Computer program]. Version 6.0.13, retrieved 4 February 2016 from <http://www.praat.org/>*. Retrieved from <http://www.praat.org/>

- Boll-Avetisyan, N., Bhatara, A., & Höhle, B. (2017). Effects of musicality on the perception of rhythmic structure in speech. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1).
- Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., & Höhle, B. (2016). Effects of experience with L2 and music on rhythmic grouping by French listeners. *Bilingualism: Language and Cognition*, 19(05), 971–986. <https://doi.org/10.1017/S1366728915000425>
- Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of Mandarin tone sandhi. *Language Sciences*, 48, 62–69.
- 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 Acoustical Society of America*, 95(May), 1570–1580. <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](https://doi.org/10.1207/S15327078IN0203_6)
- Delattre, P. (1965). *Comparing the prosodic features of English, German, Spanish and French: An interim report*. Heidelberg: Julius Groos Verlag.
- Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A distressing “deafness” in French? *Journal of Memory and Language*, 36, 406–421. <https://doi.org/10.1006/jmla.1996.2500>
- Dupoux, E., Peperkamp, S., & Sebastian-Galles, N. (2001). A robust method to study stress “deafness.” *J Acoust Soc Am*, 110(3), 1606–1618. <https://doi.org/10.1121/1.1380437>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>

- Féry, C. (2017). *Intonation and prosodic structure*. <https://doi.org/10.1017/9781139022064>
- 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.), *Hamburg Studies on Multilingualism* (Vol. 10, pp. 11–41). <https://doi.org/10.1075/hsm.10.03fer>
- Féry, C., & Kentner, G. (2010). The prosody of embedded coordinations in German and Hindi. *Proceedings of Speech Prosody, 5th International Conference 2010*, (1), 1–4.
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, *101*(6), 3728–3740. <https://doi.org/10.1121/1.418332>
- 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.
- 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. [https://doi.org/10.1016/0010-0277\(94\)90055-8](https://doi.org/10.1016/0010-0277(94)90055-8)
- Gleitman, L. R., & Wanner, E. (1982). *Language acquisition: The state of the art*. Cambridge University Press.
- Gollrad, A. (2013). *Prosodic cue weighting in sentence comprehension: Processing German case ambiguous structures* (PhD Thesis). Potsdam University.
- 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>
- Hawthorne, K., & Gerken, L. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, *133*(2), 420–428.

- Hestvik, A., & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain and Language*, *152*, 28–43.
- Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. J. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*(3), 269–286. [https://doi.org/10.1016/S0010-0277\(87\)80002-1](https://doi.org/10.1016/S0010-0277(87)80002-1)
- Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, *32*(3), 262–274. <https://doi.org/10.1016/j.infbeh.2009.03.004>
- Höhle, B., Wellmann, C., & Holzgrefe-Lang, J. (2013, June). *Behavioral and ERP responses to prosodic boundaries in German-learning infants: Evidence for an early adult-like cue weighting*. Poster presented at the Workshop on Infant Language Development (WILD), Donostia - San Sebastian.
- Holzgrefe-Lang, J., Wellmann, C., Petrone, C., Råling, 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*, *31*(7), 904–920. <https://doi.org/10.1080/23273798.2016.1157195>
- Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in Infancy research* (5th ed., pp. 69–94). Norwood, NJ: Ablex.
- 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. (2005). *Prosodic typology*. <https://doi.org/10.1093/acprof:oso/9780199249633.001.0001>

- Jun, S.-A. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In *Prosodic Typology II: The phonology of intonation and phrasing* (pp. 520–539). <https://doi.org/10.1093/acprof:oso/9780199567300.003.0017>
- Jusczyk, P. W., Hirsh-Pasek, K., Kemler 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](https://doi.org/10.1016/0010-0285(92)90009-Q)
- 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>
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, *7*(5), e36399.
- 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.
- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). *lmerTest: Tests in linear mixed effects models*. Retrieved from <https://cran.r-project.org/package=lmerTest>
- Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385–394.

- 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>
- Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, *10*(3), 241–265.
- Mazuka, R., Hasegawa, M., & Tsuji, S. (2014). Development of non-native vowel discrimination: Improvement without exposure. *Developmental Psychobiology*, *56*(2), 192–209.
- Michelas, A., & D'Imperio, M. (2015). Prosodic boundary strength guides syntactic parsing of French utterances. *Laboratory Phonology*, *6*(1). <https://doi.org/10.1515/lp-2015-0003>
- Nazzi, T., 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](https://doi.org/10.1207/S15327078IN0101_11)
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
- Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430. <https://doi.org/10.1016/j.wocn.2010.04.001>
- Peters, B., Kohler, K. J., & Wesener, T. (2005). Phonetische Merkmale prosodischer Phrasierung in Deutscher Spontansprache. *Prosodic Structures in German Spontaneous Speech*, 143–184.
- Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgreffe-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>

- R Core Team. (2016). *R: A language and environment for statistical computing*. Retrieved from <https://www.r-project.org/>
- 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* (PhD Thesis). Potsdam University.
- 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>
- Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science*, *9*(6), 565–573. <https://doi.org/10.1111/j.1467-7687.2006.00534.x>
- Shi, R., Santos, E., Gao, J., & Li, A. (2017). Perception of similar and dissimilar lexical tones by non-tone-learning infants. *Infancy*, *22*(6), 790–800. <https://doi.org/10.1111/infa.12191>
- 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), 88–104. <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., 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](https://doi.org/10.1016/S0749-596X(03)00024-X)
- Tsao, F.-M. (2017). Perceptual improvement of lexical tones in infants: Effects of tone language experience. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00558>
- Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (Vol. 14, pp. 53–66). [https://doi.org/10.1007/978-3-642-69103-4\\_5](https://doi.org/10.1007/978-3-642-69103-4_5)
- Vaissière, J., & Michaud, A. (2006). Prosodic constituents in French: A data-driven approach. In I. Fónagy, Y. Kawaguchi, & T. Moriguchi (Series Ed.), *Prosody and Syntax* (pp. 47–64). John Benjamins.
- Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cognitive Brain Research*, 18(2), 149–161. <https://doi.org/10.1016/j.cogbrainres.2003.10.001>
- 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. <https://doi.org/10.3389/fpsyg.2012.00580>
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech-perception—Evidence for perceptual reorganization during the 1st year of life. *Infant Behavior and Development*, 7(1), 49–63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.