Six-month-old infants recognize phrases in song and speech

Laura E. Hahn [1,2], Titia Benders [3], Tineke M. Snijders [4,5], Paula Fikkert [1]

[1]Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

[2]International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands

[3]Department of Linguistics, Macquarie University, North Ryde, Australia

[4]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[5]Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Radboud University, Nijmegen, The Netherlands

Word count: 9917

Link to online materials for this study:
https://osf.io/4zvad/?view_only=4cab4b2f090b4000a3425042422385de

Correspondence
Laura E. Hahn, *Centre for Language Studies, Radboud University,* Postbus 9103, 6500 HD, Nijmegen, the Netherlands. E-mail: l.hahn@let.ru.nl, phone: 0031 24 361 2564

Keywords
prosody – song – phrase – headturn preference paradigm

Acknowledgments
The authors would like to thank Lisa Rommers, Maaike ten Buuren, Katharina Menn and Mareike Geiger for their help in scheduling and testing the infants, Katharina Menn and Emma van Kampen for their reliability coding and Annelies van Wijngaarden for being our speaker and singer for the stimuli. The manuscript also benefitted from the critical comments of three anonymous reviewers. The authors declare no conflicts of interest regarding the funding source for this study.

Abstract

Infants exploit acoustic boundaries to perceptually organize phrases in speech. This prosodic parsing ability is well-attested (Nazzi et al., 2000; Johnson & Seidl, 2008) and is a cornerstone to the development of speech perception and grammar. However, infants also receive linguistic input in child songs. This study provides evidence that infants parse songs into meaningful phrasal units and replicates previous research for speech. Six-month-old Dutch infants (n = 80) were tested in the song or speech modality in the Headturn Preference Procedure. First, infants were familiarized to two versions of the same word sequence: one version represented a well-formed unit, the other contained a phrase boundary halfway through. At test, infants were presented two passages, each containing one version of the familiarized sequence. The results for speech replicated the previously observed preference for the passage containing the well-formed sequence, but only in a more fine-grained analysis. The preference for well-formed phrases was also observed in the song modality, indicating that infants recognize phrase structure in song. There were acoustic differences between stimuli of the current and previous studies, suggesting that infants are flexible in their processing of boundary cues while also providing a possible explanation for differences in effect sizes.

Six-month-old infants recognize phrases in song and speech

Across the globe, caregivers sing for their infants (Trehub & Trainor, 1998). Infant-directed (ID)-song has a universal acoustic shape (Mehr et al., 2019) and is a distinct communicative modality that is recognized across cultures (Mehr et al., 2019; Trehub, Unyk, & Trainor, 1993). Young children are exposed to songs in a highly ritualized and repetitive fashion (Bergeson & Trehub, 2002; Custodero & Johnson-Green, 2008; Custodero, Rebello Britto, & Brooks-Gunn, 2003; Ilari, 2005), and ID-song clearly serves social-emotional functions: to soothe or stimulate the infant and to strengthen the bond between caregiver and infant (Cirelli, Jurewicz, & Trehub, 2019; Corbeil, Trehub, & Peretz, 2016; Shenfield, Trehub, & Nakata, 2003).

Yet, songs often contain language in the song lyrics and thereby entail the possibility of language learning for the infant listener. Previous research provides evidence that infants are sensitive to small phonological units like syllables and words in song lyrics: for example, already before their first birthday infants recognize changes in the syllable order in songs (François et al., 2017; Lebedeva & Kuhl, 2010; Suppanen, Huotilainen, & Ylinen, 2019; Thiessen & Saffran, 2009), differentiate between rhyming and non-rhyming songs (Hahn, Benders, Snijders, & Fikkert, 2018), and learn novel words recurring in the song lyrics (Snijders, Benders, & Fikkert, 2020) .

Language learning, however, entails more than the recognition of single syllables or words. Infants also need to establish hierarchical relationships between smaller units of an utterance. One way for infants to deduce syntactic structure from the input is to tune into phrasal prosody (the melody and rhythm of speech), as boundaries between prosodic constituents typically overlap with boundaries between syntactic constituents (even though

the reverse is not always true) (Nespor & Vogel, 2007; Shattuck-Hufnagel & Turk, 1996). Speakers signal prosodic boundaries by altering prosodic cues like pitch, duration and pauses at constituent edges (Wagner & Watson, 2010). Recognizing these prosodic cues in their input aids infants in determining the edges of syntactic constituents (e.g. de Carvalho, Dautriche, & Christophe, 2017; Hawthorne & Gerken, 2014) and stimulates their morpho-syntactic development (Morgan & Demuth, 1996). The current study asks whether infants also exploit the melodic phrase structure of ID-songs to perceptually organize the linguistic input in songs.

Infants' Recognition of Phrase Structure in Speech and Song

Over the course of the first year of life, infants develop sensitivity to the instantiation of prosodic boundaries in their native language (Johnson & Seidl, 2008; Wellmann, Holzgrefe, Truckenbrodt, Wartenburger, & Höhle, 2012). For infants, just like for adults, the prosodic packaging of speech provides a perceptual filter (Jusczyk et al., 1992) that eases recognition, segmentation and memorization of linguistic elements (Frazier, Carlson, & Clifton, 2006; Hochmann, Langus, & Mehler, 2016; Johnson, Seidl, & Tyler, 2014; Mandel, Jusczyk, & Kemler Nelson, 1994; Seidl & Johnson, 2006; Shukla, White, & Aslin, 2011).

To date it is not known whether infants also recognize phrasal units of caregiver singing. However, many acoustic cues to phrase structure are the same in melodies and speech (Deutsch & Feroe, 1981; Heffner & Slevc, 2015; Lehrdahl & Jackendoff, 1985; Riemann, 1912; Trainor & Adams, 2000) and prosodic phrase segmentation is not bound to the listeners' native language or spoken modality: English-speaking adults do segment words from unfamiliar languages if these words are placed at prosodic phrase boundaries (Endress & Hauser, 2010; Langus, Marchetto, Bion, & Nespor, 2012) and American-English

infants can exploit the prosody of non-native languages, e.g. Japanese (Hawthorne, Mazuka, & Gerken, 2015), Polish (Jusczyk, 2003) and even American Sign Language (Brentari, González, Seidl, & Wilbur, 2010), to recognize phrases. This combination of observations informs the hypothesis that infants might also be able to perceptually organize songs into phrases.

Various aspects of ID-singing may be particularly beneficial for infants' recognition of phrasal song structure: ID-songs have a predictable canonical form (Mehr et al., 2019) and are produced at a rather slow tempo, with multiple visual and auditory cues to phrasal boundaries (Delavenne, Gratier, & Devouche, 2013; Falk & Kello, 2017; Leong & Goswami, 2015; Longhi, 2009), and with particularly salient acoustic boundary cues (Falk & Kello, 2017; Trainor, Clark, Huntley, & Adams, 1997). Songs thus provide infants with ample time to process an acoustic stimulus that is rich in structural cues. Moreover, songs grab infants' attention at least as effectively as ID-speech (Corbeil, Trehub, & Peretz, 2013; Costa-Giomi, 2014), and possibly more (Nakata & Trehub, 2004; Tsang, Falk, & Hessel, 2016).

The arguments provided so far all suggest that infants might be able to parse phrasal units from ID-song. Yet, there are also differences in the acoustic instantiation of boundary cues between song and speech (see e.g. references in Merrill et al., 2012). For infants to recognize phrase structure in songs, they thus need flexible representations of phrase boundaries which adjust to the song modality. So far, the available literature does not provide conclusive evidence for this flexibility.

Investigating infants' ability to recognize phrase structure in songs is also relevant in light of recent evidence that the recognition of phrasal structure in linguistic or musical play is related to grammar development in typically developing preschoolers (Politimou, Dalla

Bella, Farrugia, & Franco, 2019) and children with developmental language disorder (Richards & Goswami, 2019). As prosodic parsing is a pre-cursor to syntactic development (Morgan & Demuth, 1996), these studies raise the possibility that caregivers' language play, including ID-singing, contributes to the development of prosodic parsing. However, the work suggesting a relationship between phrase perception and grammar development focused on (pre)school children (Politimou et al., 2019; Richards & Goswami, 2019), whereas prosodic parsing already develops within the first year of life (Carvalho, Dautriche, Millotte, & Christophe, 2018), and a direct test of children's ability to segment phrases from songs or other forms of language play is still poignantly lacking. The current study thus aims to provide evidence for infants' recognition of the phrasal building blocks of ID-songs.

Infants' Recognition of Phrase Structure in Music

Currently, the literature on infants' perception of melodic phrase structure in music and song is sparse. In a seminal study by Jusczyk and Krumhansl (1990), 6-month-olds differentiated between excerpts from Mozart Minuets with pauses at natural (phrase boundary) positions and excerpts with unnatural pauses (within phrases). A follow-up study with American-English infants extended this finding to melodies from non-western (Japanese) child songs (Jusczyk, 2003), indicating that melodic phrase structure perception does not require extensive experience with a musical tradition. Crucially though, none of these previous studies required infants to encode and process melodic phrase structure. Instead, infants were provided with a pre-segmented stimulus that was reminiscent of their daily musical experience (naturally segmented) or rather odd (unnaturally segmented). It thus remains unclear whether infants chunk native songs into meaningful units and recognize subcomponents of the songs, despite these being the type of musical stimulus

infants are exposed to on a daily basis. Recent evidence from Dutch infants (Hahn et al., 2018) also only indirectly supports the notion of song structure being accessible: 9-month-olds differentiated rhyming (and thus more natural) songs from non-rhyming (and thus less natural) songs, but the study did not test whether this differentiation has implications for the processing of the linguistic content. In the current study, we will provide infants only with natural native child songs and will explicitly test their ability to recognize familiarized song phrases.

Extending the Prosodic Parsing Paradigm

A good starting point for an investigation of infants' encryption of the inherent structure of ID-song, is transferring a reliable paradigm from infant speech perception research to the song modality while also replicating previous research for ID-speech. Such a paradigm was provided by Nazzi and colleagues (2000), showing that 6-month-olds used prosodic phrase structure to segment clauses from continuous speech. In this Headturn Preference study, infants were familiarized to two versions of the word sequence *leafy vegetables taste so good* (Nazzi et al, 2000, Experiment 1). One version of the sequence was prosodically well-formed, carrying phrase boundaries at the edges, and sounded like a coherent clause: *[Leafy vegetables taste so good].* The other version of the word-sequence contained a phrase boundary halfway through, sounding more like snippets of two adjacent clauses: *leafy vegetables] [Taste so good*. In the subsequent test phase, infants heard two spoken passages of three sentences each. The well-formed sequence from the familiarization phase reoccurred as a coherent clause of one passage. The ill-formed sequence from the familiarization phase reoccurred as a subcomponent of two adjacent clauses of the other passage. Infants listened longer to the passage containing the well-

formed compared to the ill-formed word sequence, indicating that they capitalized on the prosodic structure of the passage to recognize the familiarized well-formed word sequence therein. This paradigm has been adopted in numerous subsequent studies (Seidl, 2007; Seidl & Cristia, 2008; Soderstrom, Kemler Nelson, & Jusczyk, 2005). Critically for the present study, Dutch 6-month-olds also showed the same preference for the passage containing the well-formed word sequence (Johnson & Seidl, 2008).

Whether infants' prosodic parsing ability extends to the musical modality has already been explored in two short reports (Hawthorne & Gerken, 2013; Nazzi et al., 2000, see general discussion). Both studies applied the paradigm described above to melodies from a musical instrument. The preliminary results suggest that infants recognized the familiarized well-formed tone-sequence within a longer musical piece.

The Current Study

The current study investigates infants' recognition of the phrasal building blocks of ID-song and replicates earlier studies on infants' recognition of phrases in ID-speech. We will use the paradigm described above that has successfully revealed infants' phrase segmentation of ID-speech (for Dutch: Johnson & Seidl, 2008; the original study for English: Nazzi et al., 2000) with a new sample of Dutch 6-month-olds and a new version of the Dutch stimuli and extend the paradigm to ID-song, using natural song material that matches the ID-speech stimuli in content and syntactic structure. Our approach significantly extends previous work in two ways: First, infants' processing of song lyrics has so far been limited to smaller phonological and lexical building blocks (François et al., 2017; Hahn et al., 2018; Lebedeva & Kuhl, 2010; Snijders et al., 2020; Suppanen et al., 2019; Thiessen & Saffran, 2009). We will extend the scope of this research to phrases, cognitive units which are

relevant not only for the perception of song structure but also for lexical and syntactic development in infants' native language. Secondly, we will build upon the previous work on infants' auditory grouping in polyphonic instrumental music (Jusczyk & Krumhansl, 1993; Krumhansl & Jusczyk, 1990), monophonic melodies (Nazzi et al., 2000; Hawthorne & Gerken, 2013) and non-native child songs (Jusczyk, 2003), employing the type of musical stimulus that possibly best represents infants' musical input (Volkova, Trehub, & Schellenberg, 2006), namely native child songs. By extending the paradigm of Nazzi and colleagues (2000) we will also move beyond mere preferences for naturally phrased melodies. Instead, our study requires infants to incrementally process and organize ecologically valid native song input and match this input to memorized song fragments.

Method

Participants

A sample of 95 6-month-old infants (mean age in days: 184, range: 167-209 days, *SD* = 9.02, 53 girls) from monolingual Dutch households was tested of which 12 infants were excluded, because they fussed or cried during the experiment (*n*=11) or grew up in a bilingual household (*n*=1). Three more infants were excluded from part of the analysis because they did not contribute trials in both experimental conditions for the critical data set (see Analysis Section), resulting in a final dataset of n=80 or n=83 infants depending on the respective analysis.

Participants were recruited from the Baby and Child Research Center at Radboud University, Nijmegen, the Netherlands. According to their caregivers, infants were born full-term, had normal hearing, and no familial history of language or reading problems. The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each infant before any assessment or data collection. Ethical approval for the study was obtained from the Ethiek Commissie Faculteit der Sociale Wetenschappen (ECSW) at Radboud University in Nijmegen, Netherlands. Caregivers had the choice between 10€ or a book as a reward for their participation. The results of a questionnaire on musical exposure confirmed that all participants were regularly exposed to songs and music from electronic devices and human singers (the results of the questionnaire are summarized in the online materials).

A power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), based on Experiment 1 of Johnson and Seidl (2008) (estimated correlation between groups set to 0.5, Cohen's $d_{Experiment\ 1}$ = 0.35) resulted in a required minimum sample of 52 infants to detect the

phrase-segmentation effect in each modality (80% power in one-sided *t*-test with α = .05). We thus aimed for usable data of 104 infants in total. Due to time and resource limitations, however, data collection was terminated after 95 participants.

Materials

Materials, design and procedure of the current study closely followed the study by Johnson and Seidl (2008, Experiment 1, henceforth "J&S"). Stimuli were novel spoken and sung recordings of the J&S stimuli, complemented by a second stimulus set. The basis of all stimulus materials was a pair of text passages from J&S, both consisting of three sentences, separated by two phrase boundaries (see Table 1, Pair 1).

INSERT TABLE 1 AROUND HERE

Within both passages, the same sequence of words occurred (e.g., *koude pizza smaakt niet zo goed*), but one passage contained the sequence as a single phrase, i.e. *phrase-internal* (e.g., [*koude pizza smaakt niet zo goed]*) and the other passage contained the sequence with a phrase boundary in the middle, i.e. phrase-straddling (e.g., … *koude pizza] [smaakt niet zo goed …*). The two passages were used for the test phase of the experiment. The *phrase-internal* and *phrase-straddling* sequences extracted from the passages were used for familiarization. All stimuli were recorded in a spoken as well as a sung version. Passage pair 1 was based on the Dutch stimuli of J&S, with a slight change to fit the melody[1].

Passage pair 2 was created in analogy to pair 1: The number of syllables, word stress,

[1] We slightly modified one phrase: original wording: "Hun zus vindt dat lekker." '*Their sister likes that'.* Novel wording: "Hun opa vindt dat wel erg lekker." '*Their grandpa really likes that.'*

and phrase structure were identical, and the lyrics had the same assumed familiarity (all content words of both pair 1 and pair 2 appeared in the Dutch N-CDI (Zink & Lejaegere, 2002) and the words had a similar mean log raw frequency of 3.6 (pair 1) and 3.8 (pair 2) in the Dutch Celex corpus (Baayen, Piepenbrock, & Gulikers, 1995). Both passage pairs are provided in Table 1.

Song stimuli

Both pairs of passages were set onto the melodies of child songs (Figure 1). Passage 1 was set onto melody 1 ("Sea Saw Margery Daw", originally from England) and passage 2 was set onto melody 2 ("Vine Melcul Suparat", originally from Romania), with one syllable per musical note and stressed syllables on strong metrical positions within each melody. The position of sentence boundaries in the passages was aligned with the position of melodic phrase boundaries in the melodies.

INSERT FIGURE_1 AROUND HERE

Three listeners (two amateur – one professional musician, two Dutch – one English native speaker) who were kept naïve to the purpose of the study, judged the quality of the resulting melodies. All three found them to resemble typical children's songs.

Recording

The same female Dutch speaker was recorded for the spoken and sung stimuli and was kept naïve to the purpose of the study. Only after the recording it became apparent that the same person's voice had been recorded for the original J&S stimuli. The singer/speaker was instructed to speak and sing in a lively, child-directed manner while looking at the photo of a toddler from her family. She chose a speaking and singing tempo and a pitch height that were convenient to her. Recording took place in a sound attenuated

booth and further processing was done using Praat (5.3.49, (Boersma & Weenink, 2014)) and Audacity (2.1.0): Pauses between phrases were set to silence but kept at their original duration. Two sequences were cut from each passage: one internal and one straddling (see Figure 3), resulting in 8 sequences for the sung and 8 sequences for the spoken modality.

INSERT FIGURE 2 AROUND HERE

Acoustic analysis

Acoustic measures were obtained around the internal boundary in the straddling sequence (e.g. … pizza] [smaakt …) and compared to the same sequence without a boundary in the phrase-internal sequence (e.g. [… pizza smaakt ….]) using Praat (5.3.49, (Boersma & Weenink, 2014) (sound files and corresponding text grids can be found in the online materials).

Comparison of Song and Speech Stimuli within the Current Study. Phrase boundaries in both song and speech stimuli were expressed by longer pauses and longer pre-boundary vowels at the phrase boundary in the straddling sequence compared to the corresponding internal sequence. In the song stimuli, the pitch rose after the boundary in the straddling sequences. In the spoken sequences the opposite pattern was found: pitch increased at the final vowel and then decreased at the first vowel of the following phrase. The speaker thus used a rising boundary tone to mark the end of her spoken phrases.

Comparison of Stimuli from Johnson and Seidl (2008) and the Stimuli for the Current Study. Despite the fact that we used the same words and the same speaker, the speech stimuli of the current study were substantially slower than the stimuli by J&S (see Table 4 in the online materials). The longer pauses between sentences together with the overall slower speech rate of the current stimuli resulted in large differences in onset time

of the critical sequence (see Table 5 in the online materials). These differences will be taken into account in the analysis of the looking-time data (see section Mixed-effect model analysis below). Furthermore, the pitch reset at the phrase boundaries of the straddling sequences of the J&S study was less pronounced in the stimuli of the current study (see Table 6 in the online materials), probably due to the slightly different intonation and the rising boundary tone the speaker used for the current study.

Stimulus pre-test

Three Dutch native speakers judged the intelligibility of the sung and spoken sequences and were asked to judge the *straddling/internal* manipulation of the sung and spoken sequences. The three judges were first asked to listen once to each of the 16 sequences and immediately write out the text orthographically, as they understood it. All three participants wrote down the correct texts without mishearing. They were then presented with the *phrase-internal* and *phrase-straddling* versions of every sequence as a pair and were asked to indicate which of the two sequences sounded more coherent. All three participants judged all *phrase-internal* sequences to sound more coherent than their *phrase-straddling* counterparts.

Procedure

The experiment was run using the Head-turn Preference Procedure. Three lights were placed within a three-sided booth at infant eye-level: a blue light in the center and red lights on the right and left walls of the booth. A camera was hidden below the center light to observe infant behavior from outside. Stimuli were presented via loudspeakers below the red lights. The infant and caregiver were seated in the middle of the booth, directly opposite the blue center light, exactly in between the left and right red lights. Stimulus presentation

was controlled from outside the test booth by the experimenter, using the stimulus presentation software Look! (Meints & Woodford, 2008). The experimenter was blind to trial number and trial condition and coded the looking behavior of the infant (left, right, center) using assigned keys. The same procedure was used for both familiarization trials and test trials. The entire session was video-recorded for offline reliability coding (see section "reliability coding" in the online materials).

Design

Infants took part in either the song or the speech version of the experiment, and were tested on their ability to segment either songs or speech into phrases (effect of modality, between subjects). Following Nazzi's (2000) Headturn Preference Procedure, infants were first familiarized with two sequences of the same words, one uttered as phrase internal, carrying phrase boundaries at the edges, e.g. *[Koude pizza smaakt niet zo goed]* ("Cold pizza doesn't taste so well"), the other uttered as phrase straddling, carrying a phrase boundary halfway, e.g. *koude pizza] [smaakt niet zo goed* ("cold pizza. Doesn't taste so well"). The internal sequence thus represented a well-formed acoustic unit, the straddling sequence was ill-formed. Apart from this acoustic difference, the exact same words were occurring in the sequences used in the familiarization phase. In the test phase, infants were presented with two passages of three sentences each: one passage contained the phrase straddling sequence, the other the phrase internal sequence (Table 1). For the analysis, looking times to the passages were assessed. Which passage functioned as the internal and which as the straddling passage was determined by the content of the respective sequence used during familiarization (effect of condition, within-subjects).

Counterbalancing and randomization

The four pairs of passages (Table 1) were distributed across eight lists (four lists for speech, four lists for song). Within each list, one pair of passages was used and presentation side of the first stimulus (left, right) was counterbalanced and the same presentation side and the same condition were restricted to occur maximally two times in a row.

Experimental session

Caregivers were first briefed about the experimental procedure and filled out the music exposure survey (see the online materials for an English translation of the questionnaire). At the start of the experiment, infants were seated on their caregiver's lap in the center of a three-sided test booth. Both caregiver and experimenter wore headphones throughout the experiment and listened to masking music (samba music played simultaneously with spoken text from various female speakers). Testing started with a familiarization phase during which infants heard alternations of the phrase-internal and phrase-straddling sequence and accumulated a minimum of 30 sec of looking time for each sequence (in accordance with J&S). Within the test phase, each infant was presented with two passages. One passage contained the *phrase-internal*, the other the *phrase-straddling* sequence from the familiarization phase. Which passage acted as phrase-straddling or phrase-internal during test depended on which sequence a particular infant was familiarized to. A single test trial consisted of repetitions of a passage for the same condition (internal/straddling). Trials alternated in condition (internal/straddling). Passages were presented in 12 trials distributed over three blocks. Within every block, each passage was presented once from the left and once from the right side.

The full experimental session lasted about five minutes, depending on the number of familiarization trials an infant required to reach the 30 sec familiarization criterion. Sessions

were aborted earlier if the infant fussed. Data from aborted test sessions were not analyzed. After the experiment, caregivers were debriefed about the research question of the experiment.

Results

All data preprocessing and analyses have been performed using R for windows (R Development Core Team, 2012). All raw data and analysis scripts are available in the online materials.

Mixed-effect model analysis

Linear mixed-effect models were used to analyze differences in looking times between the internal and straddling passages in the test phase of the experiment. Two models were fit, one to the full dataset of all trials from all children (N = 83, n = 41 in song; 996 trials, 492 trials in song) and a second model starting from trials during which infants had attended long enough to be presented with the first 500 ms of the critical sequence within the test passage (n = 80, n = 39 in song, 680 trials, 295 trials in song). The second model on this Critical Sequence dataset was considered warranted given the overall slower speech rate and longer pauses in the present compared to the J&S stimuli, as described above. Note that three subjects were excluded from the second model because they did not contribute trials for both conditions in this dataset. The remaining 80 infants contributed an average of 4 trials per condition (range: 1-6 for both conditions). The fixed effects of both models were 1) boundary condition (internal vs. straddling, coded as an orthogonal contrast), 2) modality (song vs. speech, coded as an orthogonal contrast), 3) test-trial number linear (1 to 12, coded as the linear polynomial), 4) test-trial number quadratic (1 to

12, coded as the quadratic polynomial) and 5) the interaction of boundary condition and modality[2]. The random effects structure of both models was specified to include random intercepts for participant and by-participant random slopes for the effect of experimental condition (internal / straddling). We deliberately chose not to specify the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013), because the use of only two pairs of passages for speech and song did not warrant specification of item-related random effects. Both model 1 and 2 were fit onto box-cox transformed looking times (λ = 0.12 for model 1, λ = -0.02 for model 2 (Csibra, Hernik, Mascaro, Tatone, & Lengyel, 2016)). The R-package "lmerTest" was used to run the models and evaluate significance of the effects (Kuznetsova, Brockhoff, & Christensen, 2016).

Results of mixed-effect model analysis

When only considering trials during which infants listened long enough to reach the critical sequence within the test passage, infants preferred to listen to the passage that contained the phrase-internal sequence in both song and speech (Figure 3). The second linear mixed-effect model (Table 2) run on this Critical Sequence data set (model 1, n = 80, 680 trials) revealed significant main effects of Condition (t = 2.21, ß = 0.05 , p = .03), Modality (t = 2.78, ß = 1.0 , p = .007) and the linear and quadratic polynomial of Test Trial Number (t = -4.39, ß = -0.28 , p < .001; t = 3.42, ß = 0.22 , p < .001). There was no significant interaction between Condition and Modality (t = 0.24, ß = 0.01, p = .81) and thus no evidence that segmentation is easier in song than in speech.

[2] Note that we ran additional analyses on the full dataset with a version of model 1 that included Experiment Version as a fixed effect (4 Versions for each Modality). This model was rank-deficient and had higher AICs and BICs than the same model without this fixed effect. We therefore decided to remove this effect from model 1, and did not include it in model 2. The results for model 1 with and without the factor Experiment Version were qualitatively the same.

Considering all trials from all children (*N* = 83, 996 trials), we did not find evidence for a preference for passages with the phrase-internal or the phrase-straddling sequence nor did we find evidence that looking times differed between song and speech (Figure in the online materials). The linear mixed-effect model 2 (Table 2) only indicated significant main effects of the linear (*t* = -7.34, ß = -1.84, *p* < .001) and quadratic (*t* = 2.19, ß = 0.55, *p* = .03) polynomial of Test Trial Number, indicating that overall looking times decreased over the course of the experiment, but to a lesser degree towards the end of the experiment.

INSERT FIGURE 3 AROUND HERE

INSERT TABLE 2 AROUND HERE

_t_-test analysis

To adhere to more standard analyses of infant looking time data we also report results of *t*-tests within both modalities using the aggregated looking times within the Critical Sequence data set (*n* = 80). Given the number of previous studies that found a preference for the internal sequence, we decided to run one-sided *t*-tests to test the hypothesis that looking times for the internal sequence are longer than for the straddling sequence. Two-sided *t*-tests will be reported for the sake of completeness. Averaged looking times were also box-cox transformed, using λ = 0.2 for song and λ = 0.36 for speech data. Levene's test indicated equal variance among groups in both the song and speech dataset. A Shapiro-Wilk test indicated that the song data deviates from normality even after transformation (*p* = 0.02). Therefore, the results of the *t*-test for song have to be interpreted with caution. Effect sizes Cohen's d_z, and Hedge's g_{av} were calculated for the untransformed looking times in the *t*-test datasets, according to recommendations given by Lakens (2013) and formulas introduced by Cohen (1988) and Hedges and Olkin (1985). A spreadsheet by

Lakens (2013), available under https://osf.io/ixGcd/, was used for the calculation.

_t_-test results

The *t*-tests run on the averaged looking times for both song and speech trials within the Critical Sequence dataset indicated a significant preference for the internal sequence for the song modality only (Table 3). In both modalities, about half of the infants tested showed a preference for the internal sequence, i.e. longer listening times for the internal compared to the straddling passage.

INSERT TABLE 3 AROUND HERE

Discussion

The current study set out to replicate 6-month-old Dutch infants' auditory grouping abilities based on intonational phrase boundaries in ID-speech (Johnson & Seidl, 2008) and assess whether this ability extends to ID-song.

Infants in the current study were tested in a paradigm first developed by Nazzi and colleagues (2000), which has successfully revealed phrase segmentation in an earlier study of Dutch 6-month-old infants (Johnson & Seidl, 2008). We replicated this latter study in the same lab, using the same speaker for the stimuli, and extended it to the ID-song modality. To this end, infants were first familiarized to two critical sequences of the same words in either song or speech (e.g. /koude pizza smaakt niet zo goed/ ("cold pizza does not taste so well")). One sequence was uttered with a well-formed phrase structure, with phrase boundaries at the edges: [koude pizza smaakt niet zo goed] while the other sequence was uttered with an ill-formed phrase structure, straddling a phrase boundary in the middle: koude pizza] [Smaakt niet zo goed. Infants were then presented with two three-sentence test passages, one containing the well-formed word sequence and the other the ill-formed word sequence. In both song and speech, infants listened longer to the passage containing the well-formed sequence. This indicates that infants were able to segment the passages of song and speech into their underlying phrasal constituents and recognized the well-formed familiarized sequence therein. Infants' known ability to recognize the phrase structure of ID-speech thus extends to ID-song.

Contribution

The current study is the first to provide evidence that 6-month-old infants segment native child songs into well-formed phrases. Infants thus capitalize on the acoustic boundary

cues within song melodies to organize a continuous song into structurally relevant constituents and recognize phrases while the song unfolds. The present results significantly extend previous research on infants' musical grouping abilities by using ecologically valid musical stimuli and by requiring infants to group native song melodies into perceptual chunks while the song unfolds (Jusczyk & Krumhansl, 1990; Nazzi et al., 2000; Hawthorne & Gerken, 2013). This study also extends our knowledge on infants' recognition of phonological units in song lyrics from syllables (François et al., 2017; Lebedeva & Kuhl, 2010; Suppanen, Huotilainen, & Ylinen, 2019; Thiessen & Saffran, 2009), rhymes (Hahn et al., 2018) and single words (Snijders et al., 2020) to larger prosodic units, namely phrases. The potential functional relevance of these findings will be discussed below.

The current results also contribute to two more general issues in the field of first language acquisition: the first is the question about shared cognitive mechanisms underlying the processing of music, song and speech; the second pertains the optimal acoustic stimulus for infant language learning. Concerning the first question: infants' mental organization of speech and song into phrases observed in the current study may be grounded in a modality-general processing mechanism (Conway, Pisoni, & Kronenberger, 2009; Schön et al., 2010; Trehub & Hannon, 2006): a conceivable account would be that the salient acoustic structure of instrumental music, ID-song and ID-speech attracts infants' attention to utterance edges (De Diego Balaguer, Martinez-Alvarez, & Pons, 2016; Drake, Jones, & Baruch, 2000; Falk & Kello, 2017; Leong & Goswami, 2015). Alternatively, infants' phrase recognition in ID-song might stem from transfer of a speech-specific or even native language-specific prosodic parsing strategy to the song modality (Morgan & Demuth, 1996). Future research should identify the exact mechanisms underlying phrase segmentation and clarify to what extent

these are bound to a specific developmental stage, input modality or language. Our contribution to this open issue is the finding that at six months, infants' perception of phrase structure is not limited to speech-specific boundary cues (Johnson & Seidl, 2008; Seidl, 2007; Seidl & Cristia, 2008; Wellmann et al., 2012) but encompasses a more generic phrase boundary percept in song melodies, a finding that needs to be incorporated into future accounts of infant speech segmentation.

The second general contribution of the current study concerns the question about the kind of acoustic stimulus from which infants learn best. Infants' astonishing learning success in their first year of life has been attributed to the exaggerated acoustic shape of ID-speech (Kuhl et al., 1997). If this were the case, then infants should learn even better from ID-song, a type of stimulus that is even more exaggerated when compared to ID-speech in terms of pitch, rhythm and tempo (Trehub et al., 1997). In the current study, the pre-test confirmed the naturalness of the song and speech stimuli and infants showed increased attention to the songs versus speech stimuli. Also, the effect sizes of the speech and song modality were in the predicted direction (speech Cohen's d_z = 0.09; song Cohen's d_z = 0.28). Nevertheless, the current study provided no evidence for easier segmentation in ID-song than ID-speech. This is contrary to previous studies which reported a song benefit for infants' linguistic processing (François et al., 2017; Lebedeva & Kuhl, 2010; Thiessen & Saffran, 2009), but is in line with other work where no processing benefit for songs was observed (Snijders et al., 2020; Suppanen et al., 2019). In the following we will discuss possible reasons for the lack of a song advantage in the current study

Understanding the Absence of a Modality Effect

Absence of evidence for easier segmentation from songs compared to speech might reflect the relative acoustic similarity between our song and speech stimuli, resulting from the fact that the stimuli in both modalities were created to be analogous. As a result of this necessary experimental control, the speech stimuli may have been slower while the song stimuli may have displayed less repetition compared to their respective real-life counterparts. Alternatively, the hypothesized processing benefits of ID-song might have been present in the current study but counteracted by the higher familiarity of ID-speech, resulting in overall similar segmentation outcomes from song and speech. Also, it may simply be that more statistical power is needed to provide evidence for an interaction between modalities (speech/song) and phrase segmentation. As the data of the present study are inconclusive regarding the cause of the absence of a modality effect, future studies should elucidate in how far ID-song boosts, hinders, or truly has no impact on infants' segmentation abilities.

Limitations of the Replication

Infants' preference for the passages containing the well-formed sequence in both song and speech was only evident in an analysis that differed from the study we aimed to replicate and extend (Johnson & Seidl, 2008). To understand the first difference between the analyses, one should remember that the test passages consisted of three sentences. The familiarized sequences occurred within the second sentence (see Table 1). Our analysis only included looks after infants had attended long enough to be presented with the first 500 ms of the critical sequence within the test passages (316 of 996 trials and 3 infants excluded). Johnson and Seidl (2008), on the other hand, analyzed data from all test trials. The change in analysis seemed warranted given that our stimuli were substantially slower than those in

the previous study (Johnson & Seidl, 2006). As a second difference, we made use of mixed-effect models on single-trial data in addition to t-tests on data averaged over trials within children. Using this more sophisticated analysis technique might have been necessary because of the relatively small effect sizes in the present study (Cohen's d_z = 0.09 and 0.28 in the aggregated data of speech and song, respectively) compared to the somewhat larger effect observed in the study by Johnson and Seidl (2008, Experiment 1; Cohen's d_z = 0.35).

In an attempt to understand why the effect size in the current study was smaller we can outright disregard a number of factors: language, age, experimental set-up, and even the speaker of the stimuli, and the lab in which the study was conducted were all the same as in the original study. A factor that might have impacted the effect sizes is the tempo of the experimental stimuli. For one, the critical sequences occurred within three seconds from the start of the test passages of the original study (Johnson & Seidl, 2008; range: 1.26-2.99 sec) but only up to 4 sec into the passages of the present study (range: 1.66-4.23 sec, see Table 5 in the online materials). Consequently, infants in the current study had to listen longer before they encountered the critical sequences. Secondly, the comparatively long pauses between the consecutive sentences of the test passages of the present study (range: 400-850 ms in Johnson & Seidl, 2008; range: 923-1541 ms in the current study; Table 4 in the online materials) might have created a less coherent auditory percept of the passages, resulting in overall more challenging listening conditions and hence smaller effect sizes. Despite these differences, the present study nevertheless provides moderate support for infants' processing of prosodic structure in ID-speech (Johnson & Seidl, 2008; Nazzi et al., 2000).

Future research

Previous research has considered ID-song first and foremost as a means of stimulating affiliation and mood regulation (e.g. Cirelli, Trehub, & Trainor, 2018). Consistent with this view, songs are not always included in descriptions of infants' speech input (e.g. Cristia, 2013; Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015). However, other descriptions of infants' linguistic input have been broadened to include ID-song (Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2019; Soderstrom & Wittebolle, 2013). The present data contribute to the mounting evidence that songs can indeed be a source for infants' implicit linguistic learning (François et al., 2017; Hahn et al., 2018; Lebedeva & Kuhl, 2010; Snijders et al., 2020; Suppanen et al., 2019; Thiessen & Saffran, 2009). Consequently, ID-song should be included in descriptions of the linguistically relevant input that infants receive.

In how far could phrase segmentation from ID-song be relevant to infant language acquisition? For one, it could aid infants in identifying smaller linguistic units within the song lyrics, e.g. words occurring at phrase boundaries (see for a similar account for speech e.g. Johnson et al., 2014; Shukla et al., 2011), and help to transfer the song lyrics and melody into working memory by chunking them into manageable units. This, in turn, might help infants to identify the song and its lyrics across different occasions in their daily routines and across different singers, contributing to the formation of context- and singer-independent abstract representations. Phrase segmentation from ID-song might also, indirectly, benefit the processing of (ID)-speech: By attending to melodic phrases in songs infants train to allocate attention to important units in the song input. The same units are also relevant in speech, but presumably less salient and occurring at a much faster time scale. Caregiver singing could thus provide infants with an acoustic playground, a practice field to engage mechanisms that are also at work in the presumably more demanding speech signal.

Future research should investigate the functional relevance of infants' ability to segment songs into phrases. There is ample evidence that prosodic phrase segmentation of speech is a key prerequisite for lexical and morpho-syntactic development (Carvalho et al., 2018). It has even been suggested that impaired recognition of large phrasal boundaries in speech is the key underlying deficit for developmental language disorder (Richards & Goswami, 2019). Consequently, future research should investigate to what extent caregiver singing and other types of rhythmic-melodic input such as rhyming story books (Richards & Goswami, 2019) contribute to infants' perception of phrasal boundaries in speech. Such a relationship between language play and real-life linguistic abilities would speak to recent studies suggesting a link between rhythmic-melodic processing of music and speech on the one hand and grammar development on the other (Gordon, Jacobs, Schuele, & McAuley, 2015; Leong & Goswami, 2015; Politimou et al., 2019). The current study contributes an empirical foundation for such future investigations, by showing that for young infants the major phrasal units in ID-song are at least as accessible as in ID-speech.

Conclusion

Recognizing phrases in continuous speech is a cornerstone of the development of speech perception. This study replicated a previous finding regarding Dutch 6-month-olds' recognition of phrase structure of ID-Speech (Johnson & Seidl, 2008) and extended the results to ID-song. Thus, already within their first half year of life, infants actively process sung input online and memorize well-formed sung phrases. Future research should identify the mechanisms underlying this ability and clarify whether the recognition of the phrasal structure of caregiver singing contributes to linguistic development.

References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database. Linguistic data consortium. University of Pennsylvania, Philadelphia.

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

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), 1–10. https://doi.org/10.1111/desc.12715

Bergeson, T. R., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science*, *13*(1), 72–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11892783

Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer. Retrieved from www.praat.org

Brentari, D., González, C., Seidl, A., & Wilbur, R. (2010). Sensitivity to Visual Prosodic Cues in Signers and Nonsigners. *Language and Speech*, *54*(1), 49–72. https://doi.org/10.1177/0023830910388011

Carvalho, A. De, Dautriche, I., Millotte, S., & Christophe, A. (2018). Early perception of phrasal prosody and its role in syntactic and lexical acquisition. In P. Pilar & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition*. Amsterdam / Philadelphia: John Benhamins Publishing Company (TILAR Series).

Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2019). Effects of Maternal Singing Style on Mother–Infant Arousal and Behavior. *Journal of Cognitive Neuroscience*, *24*(6), 1275–1285. https://doi.org/10.1162/jocn

Cirelli, L. K., Trehub, S. E., & Trainor, L. J. (2018). Rhythm and melody as social signals for infants. *Annals of the New York Academy of Sciences*, 1–7. https://doi.org/10.1111/nyas.13580

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The Importance of Sound for Cognitive Sequencing Abilities: The Auditory Scaffolding Hypothesis. *Current Directions in Psychological Science*, *18*(5), 275–279. https://doi.org/10.1111/j.1467-8721.2009.01651.x

Corbeil, M., Trehub, S. E., & Peretz, I. (2013). Speech vs. singing: infants choose happier sounds. *Frontiers in Psychology*, *4*(June), 372. https://doi.org/10.3389/fpsyg.2013.00372

Corbeil, M., Trehub, S. E., & Peretz, I. (2016). Singing Delays the Onset of Infant Distress. *Infancy*, *21*(3), 373–391. https://doi.org/10.1111/infa.12114

Costa-Giomi, E. (2014). Mode of Presentation Affects Infants' Preferential Attention to Singing and Speech. *Music Perception*, *32*(2), 160–169. Retrieved from http://www.jstor.org/stable/10.1525/mp.2014.32.2.160

Cristia, A. (2013). Input to Language: The Phonetics and Perception of Infant-Directed

Speech. *Linguistics and Language Compass*, *7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical Treatment of Looking-Time Data. *Developmental Psychology*, *52*(4), 521–536.

Custodero, L. A., & Johnson-Green, E. A. (2008). Caregiving in counterpoint: Reciprocal influences in the musical parenting of younger and older infants. *Early Child Development and Care*, *178*(1), 15–39. https://doi.org/10.1080/03004430600601115

Custodero, L. A., Rebello Britto, P., & Brooks-Gunn, J. (2003). Musical lives: A collective portrait of American parents and their young children. *Journal of Applied Developmental Psychology*, *24*(5), 553–572. https://doi.org/10.1016/j.appdev.2003.08.005

de Carvalho, A., Dautriche, I., & Christophe, A. (2017). Phrasal prosody constrains online syntactic analysis in two-year-old children. *Cognition*, *163*, 67–79. https://doi.org/10.1016/j.cognition.2017.02.018

De Diego Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal Attention as a Scaffold for Language Development. *Frontiers in Psychology*, *7*(February), 1–15. https://doi.org/10.3389/fpsyg.2016.00044

Delavenne, A., Gratier, M., & Devouche, E. (2013). Expressive timing in infant-directed singing between 3 and 6 months. *Infant Behavior & Development*, *36*(1), 1–13. https://doi.org/10.1016/j.infbeh.2012.10.004

Deutsch, D., & Feroe, J. (1981). The internal Representation of Pitch Sequences in tonal

Music. *Psychological Review*, *88*(6), 503–522. https://doi.org/10.1037/0033-295X.88.6.503

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition*, *77*(3), 251–88. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11018511

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, *61*(2), 177–199. https://doi.org/10.1016/j.cogpsych.2010.05.001

Falk, S., & Kello, C. T. (2017). Hierarchical organization in the temporal structure of infant-direct speech and song. *Cognition*, *163*, 80–86. https://doi.org/10.1016/j.cognition.2017.02.017

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), 175–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17695343

François, C., Teixidó, M., Takerkart, S., Agut, T., Bosch, L., & Rodriguez-Fornells, A. (2017). Enhanced Neonatal Brain Responses To Sung Streams Predict Vocabulary Outcomes By Age 18 Months. *Scientific Reports*, *7*(1), 12451. https://doi.org/10.1038/s41598-017-12798-2

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249. https://doi.org/10.1016/j.tics.2006.04.002

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to Me: The Social Context of Infant-Directed Speech and Its Effects on Early Language Acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Gordon, R. L., Jacobs, M. S., Schuele, C. M., & McAuley, J. D. (2015). Perspectives on the rhythm-grammar link and its implications for typical and atypical language development. *Annals of the New York Academy of Sciences*, *1337*(1), 16–25. https://doi.org/10.1111/nyas.12683

Hahn, L. E., Benders, T., Snijders, T. M., & Fikkert, P. (2018). Infants ' sensitivity to rhyme in songs. *Infant Behavior and Development*, *52*(July), 130–139. https://doi.org/10.1016/j.infbeh.2018.07.002

Hawthorne, K., & Gerken, L. (2013). Younger versus older infants' use of prosody-like boundaries to locate musical phrases. *J Acoust Soc Am*, *134*(4105). https://doi.org/10.1121/1.4831067 [doi]

Hawthorne, K., & Gerken, L. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, *133*(2), 420–428. https://doi.org/10.1016/j.cognition.2014.07.013

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

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA:

Academic Press.

Heffner, C. C., & Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Frontiers in Psychology*, *6*(1962), 1–14. https://doi.org/10.3389/fpsyg.2015.01962

Hochmann, J. R., Langus, A., & Mehler, J. (2016). An Advantage for Perceptual Edges in Young Infants' Memory for Speech. *Language Learning*, *66*(September), 13–28. https://doi.org/10.1111/lang.12202

Ilari, B. (2005). On musical parenting of young children: musical beliefs and behaviors of mothers and infants. *Early Child Development and Care*, *175*(7–8), 647–660. https://doi.org/10.1080/0300443042000302573

Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy*, *13*(5), 440–455. https://doi.org/10.1080/15250000802329321

Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. *PloS One*, *9*(1), e83546. https://doi.org/10.1371/journal.pone.0083546

Jusczyk, P. W. (2003). *American infants' perception of cues to grammatical units in non-native languages and music: evidence from Polish and Japanese*. *Jusczyk lab final report*. Retrieved from http://hincapie.psych.purdue.edu/Jusczyk

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

Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology. Human Perception and Performance*, *19*(3), 627–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8331317

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science*, *1*(1), 70–73. Retrieved from http://pss.sagepub.com/content/1/1/70.short

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., … Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2016). lmerTest: Tests in Linear Mixed Effects Models. *R Package Version*. Retrieved from https://cran.r-project.org/package=lmerTest

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 1–12. https://doi.org/10.3389/fpsyg.2013.00863

Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language*, *66*(1), 285–306. https://doi.org/10.1016/j.jml.2011.09.004

Lebedeva, G. C., & Kuhl, P. K. (2010). Sing that tune: infants' perception of melody and lyrics

and the facilitation of phonetic recognition in songs. *Infant Behavior & Development*, *33*(4), 419–30. https://doi.org/10.1016/j.infbeh.2010.04.006

Lehrdahl, F., & Jackendoff, R. (1985). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Leong, V., & Goswami, U. (2015). Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PLoS ONE*, *10*(12), 1–37. https://doi.org/10.1371/journal.pone.0144411

Longhi, E. (2009). `Songese': maternal structuring of musical interaction with infants. *Psychology of Music*, *37*(2), 195–213. https://doi.org/10.1177/0305735608097042

Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. G. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition*, *53*(2), 155–180. https://doi.org/10.1016/0010-0277(94)90069-8

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., … Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(November, eaax0868), 1–17. https://doi.org/10.1126/science.aax0868

Meints, K., & Woodford, A. (2008). Lincoln Infant Lab Package 1.0: A new programme package for IPL, preferential listening, habituation, and eye-tracking, 1–44. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Lincoln+Infant+Lab+Package+1+.+0+:+A+new+programme+package+for+IPL+,+Preferential+Listening+,+Habituation#1

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., & Friederici, A. D. (2012). Perception of words and pitch patterns in song and speech. *Frontiers in Psychology*, *3*(March), 76. https://doi.org/10.3389/fpsyg.2012.00076

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

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior & Development*, *27*(4), 455–464. https://doi.org/10.1016/j.infbeh.2004.03.002

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six-Month-Olds ' Detection of Clauses Embedded in Continuous Speech : Effects of Prosodic Well-Formedness. *Infancy*, *1*(1), 123–147.

Nespor, M., & Vogel, I. (2007). *Prosodic phonology: with a new foreword*. Berlin, Germany: Mouton de Gruyter.

Politimou, N., Dalla Bella, S., Farrugia, N., & Franco, F. (2019). Born to speak and sing: Musical predictors of language development in pre-schoolers. *Frontiers in Psychology*, *10*(APR). https://doi.org/10.3389/fpsyg.2019.00948

R Development Core Team. (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, ISBN 3-900051-07-0, http://www.R-project.org/. Vienna, Austria. https://doi.org/10.1007/978-3-540-74686-7

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising? – Linking properties of IDS to statistical learning and attention

at the prosodic level. *Cognition*, *178*(October 2017), 193–206. https://doi.org/10.1016/j.cognition.2018.05.015

Richards, S., & Goswami, U. (2019). Impaired Recognition of Metrical and Syntactic Boundaries in Children with Developmental Language Disorders. *Brain Sciences*, *9*(2), 33. https://doi.org/10.3390/brainsci9020033

Riemann, H. (1912). *Vademecum der Phrasierung* (4th ed.). Berlin: Max Hesses illustrierte Handbücher.

Schön, D., Gordon, R. L., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, *51*(1), 450–61. https://doi.org/10.1016/j.neuroimage.2010.02.023

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., & Cristia, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, *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), 565–73. https://doi.org/10.1111/j.1467-7687.2006.00534.x

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A Prosody Tutorial for Investigators of Auditory Sentence Processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247. https://doi.org/10.1007/BF01708572

Shenfield, T., Trehub, S. E., & Nakata, T. (2003). Maternal Singing Modulates Infant Arousal.

Psychology of Music, *31*(4), 365–375. https://doi.org/10.1177/03057356030314002

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6038–43. https://doi.org/10.1073/pnas.1017617108

Snijders, T. M., Benders, T., & Fikkert, P. (2020). Infants segment words from songs - an EEG study. *Brain Sciences*, *10*(39), 1–25. https://doi.org/10.3390/brainsci10010039

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

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PLoS ONE*, *8*(11). https://doi.org/10.1371/journal.pone.0080646

Suppanen, E., Huotilainen, M., & Ylinen, S. (2019). Rhythmic structure facilitates learning from auditory input in newborn infants. *Infant Behavior and Development*, *57*(July), 101346. https://doi.org/10.1016/j.infbeh.2019.101346

Thiessen, E. D., & Saffran, J. R. (2009). How the melody facilitates the message and vice versa in infant learning and memory. *Annals of the New York Academy of Sciences*, *1169*, 225–33. https://doi.org/10.1111/j.1749-6632.2009.04547.x

Trainor, L. J., & Adams, B. (2000). Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception and Psychophysics*, *62*(2), 333–340.

Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior & Development*, *20*(3), 383–396.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition*, *100*, 73–99. Retrieved from http://www.sciencedirect.com/science/article/pii/S0010027705002222

Trehub, S. E., & Trainor, L. J. (1998). Singing to infants: Lullabies and play songs. In Rovee-Collier, Lipsitt, & Hayne (Eds.), *Advances in Infancy Research. Vol. 12* (pp. 43–77). London, UK: Ablex Publishing.

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Adults identify infant-directed music across cultures. *Infant Behavior & Development*, *16*(2), 193–211. https://doi.org/10.1016/0163-6383(93)80017-3

Tsang, C. D., Falk, S., & Hessel, A. (2016). Infants Prefer Infant-Directed Song Over Speech. *Child Development*, *88*(4), 1207–1215. https://doi.org/10.1111/cdev.12647

Volkova, A., Trehub, S. E., & Schellenberg, E. G. (2006). Infants' memory for musical performances. *Developmental Science*, *9*(6), 583–9. https://doi.org/10.1111/j.1467-7687.2006.00536.x

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7), 905–945. https://doi.org/10.1080/01690961003589492

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*(December), 580. https://doi.org/10.3389/fpsyg.2012.00580

Zink, I., & Lejaegere, M. (2002). *Lijsten voor communicatieve ontwikkeling.*

Leeuven/Leusden: Acco.

Tables

Table *1*

Texts from Passage Pair 1 and 2

	Pair 1 (Melody 1) copied from Johnson and Seidl (2008)	Pair 2 (Melody 2) created in analogy to pair 1 passage 1
passage 1	Tante vraagt zich af wat <u>de jongens eten.</u> **<u>Koude pizza</u> smaakt niet zo goed**. Hun opa vindt dat wel erg lekker. *Aunt wonders what <u>the boys are eating.</u> <u>**Cold pizza**</u> **doesn't taste so good.** Their grandpa really likes that.*	Oma krijgt meteen wat <u>de meisjes maken.</u> **<u>Hete koffie</u> wordt heel snel oud.** Hun vriendje heeft toch nog geen beker. *Grandma soon gets what <u>the girls are making.</u> <u>**Hot coffee**</u> **gets old really quickly.** Their friend has no cup yet.*
passage 2	Het staat in de oven. **De jongens eten <u>koude pizza</u>**. <u>Smaakt niet zo goed</u> in de vroege ochtend. *It's (placed) in the oven. **The boys are eating <u>cold pizza.</u>** <u>(It) doesn't taste so good</u> in the early morning.*	Ze zijn in de keuken. **De meisjes maken <u>hete koffie.</u>** <u>Wordt heel snel oud</u>, als je hem niet op drinkt. *They are in the kitchen. **The girls make <u>hot coffee.</u>** <u>Gets old really quickly,</u> if you don't finish it.*

Note. Phrase-internal sequences in **bold**, straddling sequences <u>underlined</u>. Within one session (song or speech), one infant would hear sequences and passages from one pair of passages. During the familiarization phase of the experiment, an infant would only hear the internal sequence from one passage and the straddling sequence from the other passage within the same pair (crucially, both sequences consist of the same words but differ in phrasal structure). During the test phase infants hear the full passages of each pair. The experimental condition of the test passages (internal or straddling) is determined by the respective sequences heard during the familiarization phase.

Table 2

Parameters of linear mixed-effect model 1 and 2

		Model 1, Critical Sequence data set *n* = 80, 680 trials				Model 2, Full data set *N* = 83, 996 trials			
Predictor	Contrast Coding	β	*SE*	*t*(z)	*p*	β	*SE*	*t*(z)	*p*
Intercept		8.18	0.04	231.73	<0.001	16.25	0.17	94.90	<0.001
Condition	Internal (1), Straddling (-1)	0.05	0.02	2.21	0.03	-0.01	0.07	-0.12	0.90
Modality	Speech (-1), Song (1)	0.10	0.04	2.78	0.007	0.05	0.17	0.29	0.78
TrialLin	Linear polynomial	-0.28	0.06	-4.39	<0.001	-1.84	0.25	-7.34	<0.001
TrialQuad	Quadratic polynomial	0.22	0.06	3.42	<0.001	0.55	0.25	2.19	0.03
Condition * Modality		0.01	0.02	0.24	0.81	0.06	0.07	0.77	0.44

Table 3

Results of the t-tests on transformed aggregated looking times for both modalities

	Song	Speech
n, # n preference Internal	39, 23	41, 25
Mean (SD) internal, in sec	18.40 (7.43)	13.78 (5.69)
Mean (SD) straddling, in sec	15.88 (7.51)	13.37 (6.17)
cor	0.29	0.72
Cohen's d_z, Hedges' g_{av}	0.28, 0.33	0.09, 0.07
t-value, λ (box-cox transformation)	1.86, 0.2	0.76, 0.36
p-Value, Conf Int (one-sided)	0.04 [0.10, ∞]	0.2 [-1.45, ∞]
p-Value, Conf Int (two-sided)	0.07 [-0.01, 2.30]	0.4 [-2.00, 4.54]

Note. λ = value used for the box-cox transformation.

Figures

Figure 1

Melody 1 and 2 as used for the song stimuli

Note. Phrase-internal sequences in **bold**, straddling sequences underlined.

Figure 2

Example stimulus

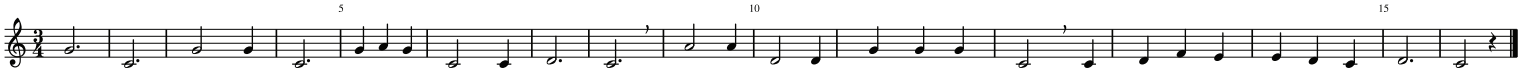
Note. All subfigures denote waveform, sonogram, spectrogram and text-grid annotation and time on the y-axis.

Figure 3

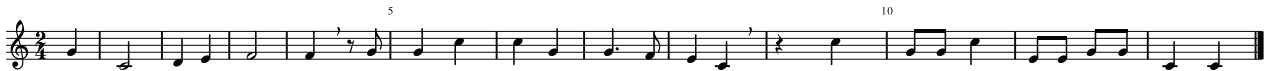
Violin- and Boxplot of box-cox transformed looking times in the Critical Sequence dataset

Note. Center lines represent grand medians, boxes entail first and third quartiles.

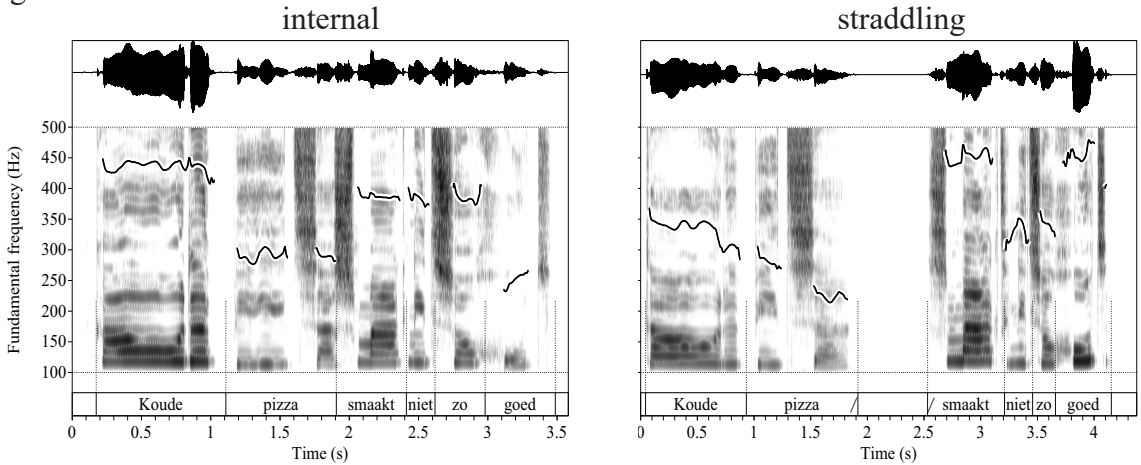
Melody 1



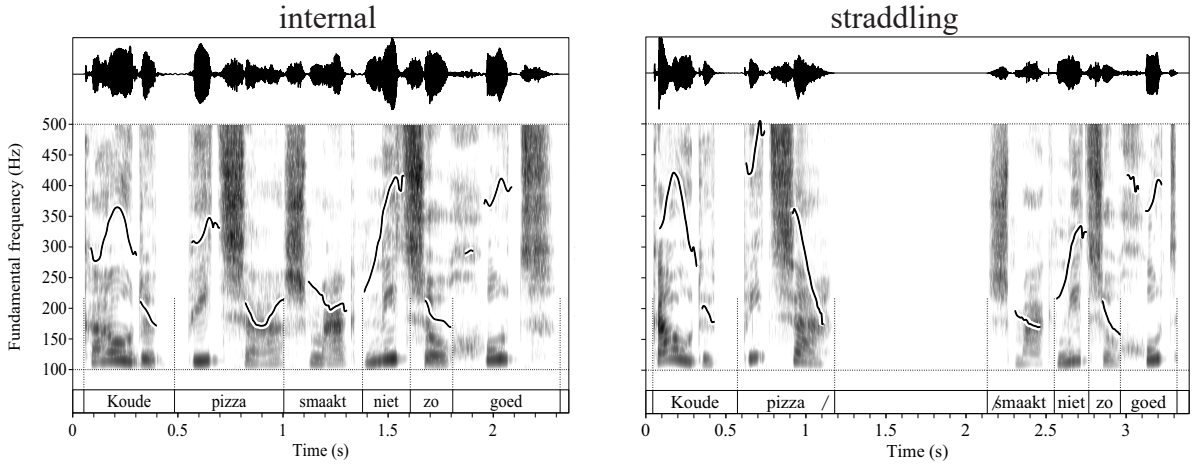
Melody 2

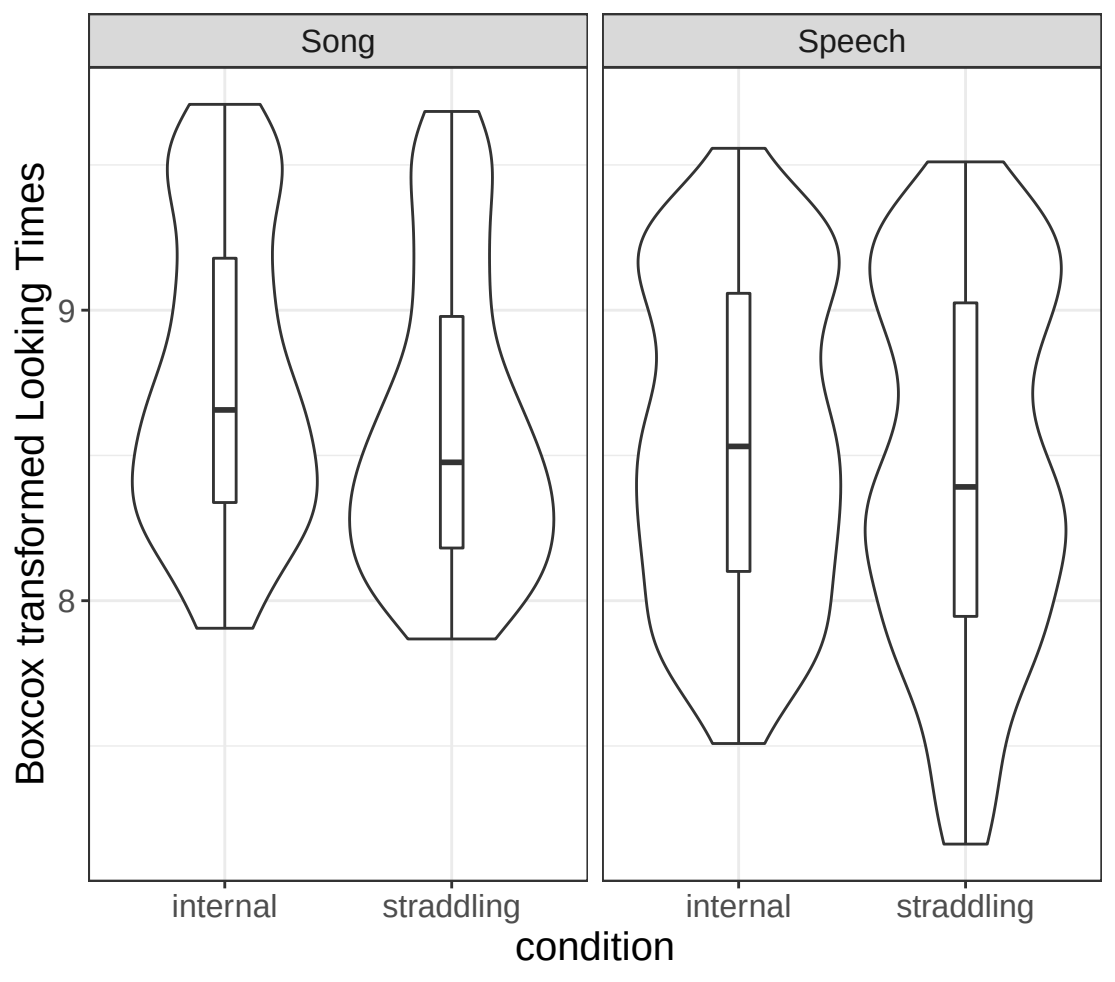


Song


internal
straddling
Fundamental frequency (Hz)
500
450
400
350
300
250
200
150
100
Koude
pizza
smaakt
niet
zo
goed
0
0.5
1
1.5
2
2.5
3
3.5
4
Time (s)

Speech


internal
straddling
Fundamental frequency (Hz)
500
450
400
350
300
250
200
150
100
Koude
pizza
smaakt
niet
zo
goed
0
0.5
1
1.5
2
2.5
3
Time (s)


Song
Speech
Boxcox transformed Looking Times
9
8
internal
straddling
internal
straddling
condition