ABSTRACT
We tested German nine month olds’ reliance on pitch and metrical stress for segmentation. In a headturn preference paradigm, infants were familiarized with trisyllabic words (weak–strong–weak (WSW) stress pattern) in sentence contexts. The words were presented in one of three naturally occurring intonation conditions: one in which high pitch was aligned with the stressed syllable and two misalignment conditions (with high pitch preceding vs. following the stressed syllable). Infants were tested on the SW unit of the WSW carriers. Experiment 1 showed recognition only when the stressed syllable was high pitched. Intonation of test items (similar vs. dissimilar to familiarization) had no influence (Experiment 2). Thus, German nine month olds perceive stressed syllables as word onsets only when high pitched, although they already generalize over different pitch contours. Different mechanisms underlying this pattern of results are discussed.

INTRODUCTION
In fluent speech, the prosodic realization of words differs depending on a number of factors, such as speaking rate, emotional attitude of the speaker, the word’s position in the phrase, sentence type, illocution, etc. For *We thank Sophie Egger and Jana Schlegel for recording, acoustic analyses, recruitment of participants, and testing. We are very grateful to Janet Grijzenhout, head of the Baby Speech Laboratory at the University of Konstanz, for making available the lab’s database and facilities as well as insightful comments. We also acknowledge support from an APF research grant from the University of Konstanz awarded to Bettina Braun (FP 15/10). Further, we thank René Kager and Anne Cutler for discussion on the experiments and data, and in particular Elizabeth Johnson for sharing her invaluable HPP expertise and for very helpful comments on an earlier version of this paper. Finally, we owe special thanks to two anonymous reviewers and the editor for their suggestions and remarks. Address for correspondence: Katharina Zahner, University of Konstanz Linguistics, PO Box 186, 78457 Konstanz, Germany. e-mail: katharina.zahner@uni-konstanz.de

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-0-314456
instance, the word ‘mommy’ is produced with falling pitch in declaratives (“Look! There is mommy.”), but with rising pitch in most polar questions (“Is that mommy?”). In both renditions, the first syllable of the word is stressed (trochaic, strong–weak pattern, henceforth SW), but it is high pitched in the first production and low pitched in the second. We know from previous research that speech segmentation is not trivial, since the acoustic speech signal lacks reliable and unambiguous acoustic cues to word boundaries (Cutler, 2012; Lehiste, 1960). We also know that the rhythmic structure of the ambient language influences segmentation strategies, and that for infants exposed to stress timed languages, such as English or German, stressed syllables provide a strong cue to word onsets (see below). What we do not know is how infants are able to extract recurring SW units despite the prosodic variability induced by utterance level intonation. This paper takes a first look at the role of intonation in speech segmentation.

It has been argued that one of the earliest cues that infants use for speech segmentation are transitional probabilities between syllables (Saffran, Aslin & Newport, 1996; Thiessen & Erickson, 2013; but see Johnson, 2012, or Johnson & Tyler, 2010, all for infants from an American English (AE) language environment). From seven and a half months onwards, infants raised in stress timed language environments also rely on metrical stress (e.g. Bartels, Darcy & Höhle, 2009, for German nine month olds; Jusczyk, Houston & Newsome, 1999, for American English seven and a half month olds; Kuipers, Coolen, Houston & Cutler, 1998, for Dutch nine month olds), and later on make use of language specific phonotactic constraints (Mattys, Jusczyk, Luce & Morgan, 1999, for AE nine month olds), co articulatory phonetic cues (Johnson & Jusczyk, 2001, for AE eight month olds) or position specific allophonic variants (Jusczyk, Hohne & Bauman, 1999, for AE ten and a half month olds). When confronted with input containing conflicting cues, American English infants at the age of five and seven and a half months rely on transitional probabilities between syllables more than on stress cues for segmentation (Thiessen & Erickson, 2013; Thiessen & Saffran, 2003). From eight months onwards, however, stress cues outweigh statistical cues (Johnson & Jusczyk, 2001; Johnson & Seidl, 2009; Thiessen & Saffran, 2003; all for AE), even when stress is signaled by no other cue than the energy distribution in the spectrum (Thiessen & Saffran, 2004, for AE). Moreover, at the age of nine months, stress is given more weight than phonotactic information when the two types of information are in conflict (Mattys et al., 1999, for AE). Thus, before infants learn to integrate several types of information, stress appears to be the most powerful cue for determining potential word boundaries.

Infants who grow up with stress timed languages, such as German, English, or Dutch, soon develop a stress based segmentation strategy and interpret stressed syllables as word onsets (e.g. Bartels et al., 2009;
Jusczyk, Houston & Newsome, 1999; Kuipers et al., 1998). Until ten and a half months of age, they fail to extract iambic (weak–strong, henceforth WS) patterns from fluent speech (Jusczyk, Houston & Newsome, 1999). However, what leads to an infant’s perception of a syllable as STRESSED when segmenting fluent speech has not been exhaustively studied to date (but see Bion, Benavides Varela & Nespor, 2011, for Italian infants’ metrical grouping preferences in a different paradigm). In this paper, we particularly investigate the role that utterance level $f_0$ (the acoustic correlate of speech pitch) plays in segmentation for German nine month old infants. Stressed syllables are salient in the speech stream (see Cutler, 2005, for an overview on lexical stress and its acoustic and perceptual correlates): in German, they are longer (Jessen, Marasek & Claßen, 1995; Schneider & Möbius, 2007) and louder than unstressed ones (Dogil, 1995; Jessen et al., 1995), they are produced with increased vocal effort (Mooshammer, 2010), and often have more peripheral vowel qualities (Delattre, 1969). When stressed syllables additionally receive phrase level prominence, thus functioning as pitch accents, they are produced with a pitch movement. In various perception studies, adult listeners have been shown to exploit these acoustic realizations of stress when identifying prominence (e.g. for English listeners, Fry, 1958; for German listeners, Kohler, 2012, and references therein). According to Kohler (2012), lexical stress is best looked at from a dynamic perspective which involves prosodic frames that are determined by $f_0$, energy, and segmental timing profiles across utterances. Of the acoustic properties of stressed syllables, $f_0$ is special, since it is induced by sentence level intonation and hence does not uniquely specify stress. Phonologically speaking, stress is a property at the word level and $f_0$ is a property of the phrase. Therefore, the $f_0$ movement associated with a stressed syllable may vary in its alignment and, depending on the pitch accent type, the stressed syllable of any given word may be rising, high, falling, or low. The choice of pitch accent and the realization of the stressed syllable in turn are governed by a number of factors, such as the position of the word in the phrase (Silverman & Pierrehumbert, 1990), the sentence type (Grice, Baumann & Benzmüller, 2005), the information structure of the utterance (Kohler, 1991), and the information status of a particular referent (Baumann & Grice, 2006). In Standard German neutral declaratives, especially in those where referents are newly introduced into the discourse (Baumann & Hadelich, 2003; Kohler, 1991), the pitch peak is usually aligned with the stressed syllable (henceforth medial peak accent, as the peak typically lies within the boundaries of the stressed syllable; see Kohler, 1991). An example realization is shown in (1) on the word Auto ‘car’ (bold face marks the stressed syllable). In autosegmental metrical phonology (see Ladd, 1996, for an overview), this accent would be labeled as L + H* or
H*. (L stands for a low and H for a high tone. The asterisk indicates the association of a given tone – here High – with the stressed syllable. Boundary tones are not indicated). Note that in the literature, the term ‘alignment’ is used when referring to the actual positioning of fo peaks and valleys in regard to the text, while the term ‘association’ is reserved for the abstract link between pitch accents and stressed syllables (e.g. Ladd, 1996, p. 55).

In other accent types, the pitch peak is misaligned with the stressed syllable. For instance, sentence topics and many other phrase initial accents (especially in southern varieties of German) are realized with a low toned stressed syllable that is followed by a rise, leading to a late peak accent with the peak on the following unstressed syllable (L* + H; Braun, 2006; Truckenbrodt, 2007). A stylized pitch contour is given in example (2a). An acoustically similar pattern is found in questions with a final rise in which the stressed syllable before the rise is often low toned and is followed by a pitch peak in the following syllable. A stylized pitch contour is shown in example (2b).

Information that is regarded as semi active and hence inferable for the listener is signaled by a pitch fall whose peak is realized before the stressed syllable, resulting in a so called early peak accent (H + L*; Baumann & Grice, 2006; Kohler, 1991). A stylized pitch contour is given in example (3).

Thus, due to the influence of utterance level intonation, duration and intensity seem to be more reliable cues to stress than fo movements. On the other hand, fo movements may be perceptually more salient than changes in duration and intensity, especially in infants’ early development. Infants have been reported
to be highly sensitive to pitch differences from a very young age (e.g. Fernald &
For instance, four month olds show a strong listening preference for
infant directed speech (IDS) over adult directed speech (ADS), mainly
because of exaggerated \( f_0 \) patterns in IDS and not so much because of
differences between ADS and IDS in terms of duration or amplitude
(Fernald & Kuhl, 1987). Studies within the framework of the
iambic trochaic law (Hay & Diehl, 2007; Hayes, 1995) revealed that infants
older than seven and a half months exploit pitch information, but not
duration, when grouping units in an artificial language task (Bion et al.,
2011). In that study, infants have been shown to pair sequences of nonsense
syllables that alternate in pitch as SW patterns (with high pitched syllables
forming the strong, and low pitched syllables the weak element), while they
are not able to use alternating patterns in duration for grouping. Referring to
Gussenhoven (2004), who claims that some kind of melodic contour (in the
form of tone, lexical pitch, or intonation) is present in all types of languages,
Frota et al. (2014) suggest that this early sensitivity to pitch information
might be caused by its perceptual salience and by its frequent use for
linguistic contrasts across languages. In sum, the role of \( f_0 \) is ambivalent: its
cross linguistic availability and acoustic salience render \( f_0 \) a highly valuable
cue to segmentation, whereas the variable alignment of \( f_0 \) peaks with
metrical stress seems to diminish its power as a segmentation cue.

In this study, we examined the role of \( f_0 \) for speech segmentation in German
nine month olds. Using the head turn preference paradigm (Fernald, 1985;
Kemler Nelson, Jusczyk, Mandel, Myers, Turk & Gerken, 1995, for a
review on this method) with a familiarization phase consisting of spoken
passages and a consecutive test phase that employed words in isolation (see
Jusczyk, Houston, & Newsome, 1999, Experiment 2), we investigated if
the position of the \( f_0 \) maximum (aligned or misaligned with the metricaly
stressed syllable) affects German infants’ segmentation behavior.

In the familiarization phase, rare trisyllabic WSW carrier words (e.g. Lagune
[la.ˈɡu.-national] ‘lagoon’) were embedded in short passages and presented in three
different intonation conditions, manipulated between subjects (Experiment
1): one alignment condition in which the pitch peak was aligned with the
stressed syllable (medial peak accent, as in example (1) above), and two
misalignment conditions in which the pitch peak and metrical prominence
fell apart, as in examples (2) with a late peak, or (3) with an early peak. We
used two misalignment conditions in order to gain a more complete
understanding of the role of \( f_0 \) for infant speech segmentation. Note that the
WSW stress pattern is the most frequent pattern in German trisyllabic
monomorphemic words and accounts for 51% of the cases (Féry, 1998).

In the test phase, infants were tested on whether they extracted trochaic
sequences (e.g. [ˈɡuː.na]), i.e. the SW unit, from the trisyllabic WSW
carrier words (e.g. [la.’gu:nə]). Closely following earlier head turn preference segmentation studies (e.g. Jusczyk & Aslin, 1995; Jusczyk, Houston & Newsome, 1999; for more detailed information see below), in Experiment 1, we chose falling pitch contours for the SW part words used in the test phase. However, this leads to a relatively close acoustic similarity of the intonation contours of the familiarization and test items in the medial peak condition (a falling pitch accent on the relevant SW structures), but not in the other two conditions. In order to rule out any potential effects of acoustic similarity and differences in task difficulty, Experiment 2 replicated the medial peak condition with test stimuli that had a rising intonation contour. That way, we created a comparably strong mismatch between familiarization and test stimuli as in the two misalignment conditions in Experiment 1.

We used trisyllabic WSW carrier words in our experiments because this allowed us to locate the entire pitch accent on the target word in all three intonation conditions (note that with disyllabic SW carrier words, as used in many other segmentation studies (e.g. Bartels et al., 2009; Jusczyk, Houston & Newsome, 1999; Kuijpers et al., 1998), the early pitch peak would have been placed on the word preceding the target word). Beyond that, WSW words provide a strong test case for investigating the role of stressed syllables as word onset cues, since these trisyllabic carrier words provide conflicting segmentation cues: on the one hand, metrical stress hints at a word boundary after the first unstressed syllable (W#SW, where ‘#’ signals a word boundary), while, on the other hand, linguistic, phonetic, and statistical regularities (such as the presence of function words, co-articulatory information, and transitional probabilities) in fact indicate a word boundary before the first unstressed syllable of the WSW carrier words (#WSW). We do not know exactly to what extent German infants at this age exploit all these cues hinting at a boundary before the trisyllabic carrier (i.e. whether there is no support for, or whether there is support against, the boundary before the SW unit). Yet, what is clear is that in the current setting infants can rely solely on stress to extract the respective SW unit.

EXPERIMENT 1

METHODS

Participants

Eighty full term infants (at least 37 weeks of gestation) from monolingual German speaking homes took part in the experiment. Infants had not been exposed to languages other than German. They were randomly assigned to one of the three intonation conditions. Those fifty four infants (aged between 0:8.19 and 0:9.16; 25 female, 29 male) who finished the
familiarization phase and all twelve test trials were included in the analysis. Twenty six infants were excluded from the analysis due to fussiness (10), crying (13), or not attending to the blinking lights (3). The exclusion criteria were the following: all infants that were tested received a fussiness score ranging from 1 (for very patient infants who behaved very well during the experiment) to 4 (for very fussy and restless infants who moved a lot or turned around on their caregiver’s lap). Those infants who received a score higher than 3 were excluded from the analysis. One third of the infants (9 female, 9 male, average age 0;9·1, sd = 0;9·8) were tested in the medial peak condition, one third (7 female, 11 male, average age 0;9·1, sd = 0;9·7) in the early peak condition, and one third (9 female, 9 male, average age 0;9·1, sd = 0;9·8) in the late peak condition. Parents were reimbursed for public transport or parking fees and received a small present for the child.

Stimuli

Familiarization stimuli. Four trisyllabic words with low lexical frequency (less than 0·1 occurrences per million in the CELEX word form dictionary; Baayen, Piepenbrock & Gulikers, 1995) that are not expected to be familiar to nine month old infants served as carrier words. All of them consisted of CV syllables with stress on the penultimate: Kanone [kʰa.ˈnoː.na] ‘cannon’; Lagune [la.ˈɡuː.na] ‘lagoon’; Kasino [kʰa.ˈsiː.no] ‘casino’; Tirade [tʰi.ˈraː.da] ‘tirade’. (Note that the unstressed initial syllables are not reduced to [ə] in German but retain their full vowel quality.) For each of the four WSW carrier words, we constructed six sentences, such that the carrier word appeared in different lexical contexts and different sentence positions (twice in sentence final position, four times early in the sentence following an article or prenominal adjective). The words preceding and following the target words differed across sentences. The four passages are listed in ‘Appendix A’.

A twenty six year old female native speaker of Standard German from the southwest of Germany (Baden Wuerttemberg) who was trained in intonational phonology recorded the twenty four target sentences in the three intonation conditions. In the medial peak condition, the stressed syllable was high toned (H* or L + H*), followed by a pitch fall (see Figure 1). In the early peak condition, the stressed syllable was low toned and the preceding unstressed syllable was high toned (H + L*), i.e. the pitch fall occurred earlier (see Figure 2). In the late peak condition, the stressed syllable was also low toned, but the following unstressed syllable was high toned (L* + H; see Figure 3).

Since word stress is signaled by a variety of acoustic cues that are distributed over the stressed and neighboring unstressed syllables
(Kohler, 2012; Niebuhr, 2007), we used naturally produced auditory stimuli in order to make all potential cues available. An additional advantage of naturally produced stimuli is that they are easier to process than (re) synthesized speech (Nix, Mehta, Dye & Cutler, 1993). Care was taken that the distribution of other pitch accents in each sentence was the same across...
intonation conditions. The speaker read the sentences in a natural and lively way. Stimuli were recorded in a sound attenuated cabin (44.1 kHz, 16 Bit) and analyzed using Praat (Boersma & Weenink, 2014). To achieve equally salient F0 movements across intonation conditions, the sentences were recorded several times and the best matching sentences were chosen, so that eventually the average F0 excursion of the fall in the medial peak and early peak condition and the rise in the late peak condition were matched. The average F0 excursion was 8.8 st (sd = 1.7 st) in the medial peak condition, 8.5 st (sd = 1.4 st) in early peak condition, and 8.4 st (sd = 1.7 st) in the late peak condition. Further acoustic analyses confirmed that the target sentences in the three intonation conditions were very similar with regard to a number of acoustic variables, as displayed in Table 1. The existing differences are typical of these kinds of pitch accents (e.g. Niebuhr, 2007). The average duration of the passages was 16.4 s (sd = 0.4 s) in the medial peak condition, 16.7 s (sd = 0.4 s) in the early peak condition, and 16.0 s (sd = 0.3 s) in the late peak condition.

Test stimuli. Each infant was tested on the same set of four test ‘words’. These four test words consisted of the SW unit of the WSW carrier words: [ˈɡuːna], taken from Lagune, [ˈɽaːdə], taken from Tirade, [ˈnoːnə], taken from Kanone, and [ˈsiːno], taken from Kasino. To increase comparability with earlier head turn preference segmentation studies (e.g. Jusczyk & Aslin, 1995; Jusczyk, Houston & Newsome, 1999), the test stimuli were elicited in the same way as in those studies, i.e. produced “as if naming the object for an

Fig. 3. Example sound pressure wave, spectrogram, and pitch track of a target sentence in the late-peak condition.
infant” (Jusczyk & Aslin, 1995, p. 6; Jusczyk, Houston & Newsome, 1999, p. 167), and with varied pitch range. Closely following Jusczyk, Houston, and Newsome (1999), who report a decrease in fo between the first and the second syllable in their SW units (see their acoustic analyses of Experiments 1–3, for instance), we chose falling pitch contours for the SW part words used in the test phase. The same speaker as for the familiarization stimuli recorded each of these disyllabic trochees approximately thirty times with a pitch fall and slightly different durations and fo excursions to increase the phonetic variability of this contour. For the experiment, we chose fifteen tokens of each disyllable, such that the average fo excursion of the pitch fall and the average duration of the test word did not differ across test words. The average fo excursion of the pitch fall was 10.1 st (sd = 1.8 st), ranging from 6.4 st to 13.1 st. The average duration of the test words was 710 ms (sd = 88 ms), ranging from 540 ms to 884 ms. The fifteen tokens of each test word were concatenated with an inter stimulus interval (ISI) of 800 ms. The lists were on average 22.71 s long (sd = 0.4 s). Further acoustic measures for the individual test lists are summarized in ‘Appendix B’.

**Procedure**
Parents first filled in a questionnaire regarding their infant’s language background and infant data. Infants were then seated on their parent’s lap,
facing a three sided black experiment booth in the Baby Speech Lab at the University of Konstanz. Each trial of the head turn preference experiment started with a green blinking light at the center of the experiment booth. As soon as the infant oriented towards the center, the green center light was switched off and a red light to the right or left of the child started blinking. When infants turned their heads towards the sidelight, the auditory stimuli started playing. The sound played as long as infants oriented towards this side. If infants looked away for more than 2 s, the next trial started. In the familiarization phase, the two passages were presented semi randomly from the left or the right side with at most two trials from the same side until children had listened to each of the two paragraphs for at least 45 s. Then, the test started automatically. In the test phase, infants listened to lists of the SW part words. They were also presented in a semi random order from the left or the right side, with no more than two trials from either side in a row. Looking times were coded online by an experimenter who monitored infants via a video camera and controlled the experiment by pressing buttons. The experimenter as well as parents wore headphones with masking music so they could not hear the auditory stimuli the infants were exposed to. The experimental session lasted approximately 6 minutes.

In each of the three intonation conditions, half of the infants were assigned to the Kanone and Tirade familiarization trials, the other half to the Kasino and Lagune familiarization trials. In the test phase, all infants listened to test lists consisting of fifteen repetitions of the four isolated SW units, two of which were part words of the WSW carrier words presented in the familiarization phase (e.g. none and rade), and two of which were part words of the WSW carrier words used with the other subsample of infants, thus novel syllable sequences (e.g. sino and gune). In total, there were three blocks of four trials, with three pseudo randomized repetitions of the four test lists. Across infants, we counterbalanced the sides from which the test lists were presented (right vs. left loudspeakers) as well as the list beginnings (such that all four part words occurred once at a list beginning).

RESULTS

Looking times in seconds were averaged by FAMILIARITY STATUS (novel vs. familiar) for each infant. The average looking times were 10.2 s (sd = 3.1 s) to novel and 8.8 s (sd = 2.7 s) to familiar lists in the medial peak condition. Thirteen out of eighteen infants oriented longer to the novel lists. In the early peak condition, infants looked on average 8.3 s (sd = 2.3 s) to novel and 8.5 s (sd = 2.2 s) to familiar lists, with ten out of eighteen infants orienting longer to the novel lists. In the late peak condition, average
looking times were 8.5 s (sd = 2.5 s) to novel and 8.4 s (sd = 2.6 s) to familiar items and eight out of eighteen infants oriented longer to the novel lists. The mean looking times to novel and familiar items are illustrated in Figure 4. Note that prior reliability studies in our lab have shown that the inter coder reliability between online and offline coding is very high. A trained person re coded the looking behavior of four randomly chosen videotapes recorded in the head turn preference paradigm (corresponding to 5% of the data of Experiments 1 and 2). The looking time data for online and offline coding were very strongly correlated (r = .99, n = 48 trials), suggesting that the online coding was reliable.

For statistical analysis, we first calculated the average looking time difference and the 95% confidence interval for all three conditions by subtracting the looking time to familiar test lists from the looking time to novel test lists for each infant (see Figure 3). The results show a robust difference in looking times only in the peak stress alignment condition (medial peak), but not in the two misalignment conditions (early peak and late peak). The overlap between the confidence interval of the medial peak condition and those of the other two intonation conditions is small enough to suggest a robust difference between the alignment condition and the two misalignment conditions (Cumming & Finch, 2005).

Results of a repeated measures ANOVA with INTONATION CONDITION as a between subject factor and FAMILIARITY STATUS as a within subject factor showed a statistically significant interaction between the two factors (F (2,51) = 3.53, p = .04). Post hoc pairwise t tests for the three intonation

Fig. 4. Average looking time in the three intonation conditions split by familiarity status (Experiment 1). Whiskers represent ±1 standard error of the mean.
conditions separately showed a statistically significant difference between looking times to novel and familiar test lists only in the medial peak condition \( t(17) = 3.2, p = .01 \), but not in the two misalignment conditions (both \( p \) values >0.7). A data analysis according to a Bayesian approach (e.g. Lee & Wagenmakers, 2013) shows that in the medial peak condition the alternative hypothesis is nine times more likely than the null hypothesis \( r = 0.71, bf = 8.70 \), while the null hypothesis is approximately four times more likely than the alternative hypothesis in the early peak condition \( r = 0.71, bf = 0.26 \) and in the late peak condition \( r = 0.71, bf = 0.25 \).

**DISCUSSION**

In the medial peak condition, in which high pitch was aligned with the stressed syllable, infants looked significantly longer to the novel than to the familiar test lists. The magnitude of this looking time difference (1.4 s) is comparable to other segmentation studies using this paradigm (e.g. Bartels et al., 2009; Jusczyk, Houston & Newsome, 1999). In the two misalignment conditions, there was no looking time difference to novel and familiar test lists. These findings show that infants extracted the embedded SW part words from fluent speech only when the stressed syllable was high pitched (peak stress alignment condition), but not when the pitch peak and stressed syllable were misaligned. In other words, only high pitched stressed syllables are taken as word beginnings, while
low pitched stressed syllables did not serve as good word onset cues for German nine month olds.

Note that the intonation contour of the test items in the peak stress alignment condition was rather similar to the contours of the target sequences in the familiarization phase (high pitch on stressed syllable, followed by a low toned post tonic syllable in both phases of the experiment). This was different in the two peak stress misalignment conditions. This asymmetry allows for an alternative interpretation of the data, which has to be excluded before drawing stronger conclusions: infants might have benefitted from the similarity between familiarization and test stimuli in the medial peak condition and suffered from the intonational change in the other two conditions. In other words, it is conceivable that the task was easier in the alignment condition (allowing for a more direct match) than in the two misalignment conditions (which necessitate abstracting away from intonational information). Previous findings on infants’ early representations speak against such a direct matching account, however: infants older than nine months of age have been shown to abstract over certain prosodic variations that are not lexically contrastive, such as speaker identity (male vs. female; Houston & Jusczyk, 2000; van Heugten & Johnson, 2012; the latter find evidence already for infants older than seven and a half months), emotions (neutral vs. happy; Singh, Morgan & White, 2004), and pitch levels (high vs. low; Singh, White & Morgan, 2008). It is therefore very likely that the change in intonation from familiarization to test phase does not hinder infants’ recognition of the SW units.

Nonetheless, in order to corroborate the results of Experiment 1, we conducted a follow up experiment in which infants were familiarized with the stimuli of the medial peak condition from Experiment 1, but now the test stimuli were presented with a rising intonation contour (instead of a fall). If the looking time difference in the medial peak condition of Experiment 1 stems from the similarity in intonation contours between familiarization and test alone, infants should not be able to segment the SW part words under these modified conditions. If infants instead rely on the peak stress alignment as a crucial segmentation cue, then they are expected to show a similar novelty effect as in the medial peak condition of Experiment 1.

EXPERIMENT 2

METHODS

Participants

Twenty nine full term infants (at least 37 weeks of gestation) took part in Experiment 2 under the same conditions as in Experiment 1. They had
not been exposed to a language other than German. Eighteen infants (aged between 0;8·18 and 0;9·21; 6 female, 12 male) who finished the familiarization phase and all twelve test trials were included in the analysis (average age 0;9·1, sd = 0;0·9). They had the same age as the infants in the medial peak condition in Experiment 1 (average age 0;9·1, sd = 0;0·8). Eleven infants had to be excluded from the analysis due to fussiness (3), crying (2), not attending to the blinking lights (3), falling asleep (2), or due to an unusually short overall average looking time (>2 sd below the average looking time (1)).

Stimuli
The familiarization stimuli were those used in the medial peak condition in Experiment 1. The four SW part words for the test phase were the same as in Experiment 1, but this time they were recorded with a rising pitch contour, resulting in a low pitched stressed syllable followed by a high pitched second syllable. The average f0 excursion of the pitch rise was 12·4 st (sd = 1·6 st), ranging from 9·1 st to 15·1 st. The average duration of the test words was 679 ms (sd = 57 ms), ranging from 546 ms to 814 ms. As before, the fifteen selected tokens were concatenated with an ISI of 800 ms, resulting in test lists with an average duration of 21·4 s (sd = 0·3 s). Further acoustic measures for the individual test lists are provided in ‘Appendix C’.

Procedure
The procedure was the same as in Experiment 1.

Results
Participants looked on average 11·1 s (sd = 2·1 s) to novel test lists and 9·6 s (sd = 2·2 s) to familiar ones. Fifteen out of eighteen infants looked longer to novel than to familiar items. Infants’ average looking times to novel and familiar lists are shown in the right hand bars in Figure 6 (for ease of comparison, the results of the medial peak condition in Experiment 1 are displayed again on the left hand side of the figure).

The average looking time difference and the 95% confidence interval are shown on the right hand side of Figure 7 (for ease of comparison, the results of the medial peak condition in Experiment 1 are displayed again on the left hand side of the figure). The results show that the difference in looking time in Experiment 2 is similar to that of the medial peak condition in Experiment 1.

A pairwise t test for the medial peak condition with rising test intonation also showed a statistically significant difference between looking times to novel and familiar test lists (t(17) = 2·9, p = 0·01), as observed in the
Fig. 6. Average looking time in the medial-peak condition of Experiments 1 and 2 split by familiarity status. Whiskers represent ±1 standard error of the mean.

Fig. 7. Means of difference in looking time to novel and familiar items in the medial-peak conditions of Experiments 1 and 2. Whiskers represent the 95% confidence interval of the difference in looking time.
medial peak condition in Experiment 1. A Bayesian factor analysis revealed that the alternative hypothesis is five times more likely than the null hypothesis ($t = 0.71$, $bf = 5.33$).

For ease of comparison, the data of both medial peak conditions (from Experiment 1 and Experiment 2) were pooled. The results of a repeated measures ANOVA with test intonation as a between subject factor and familiarity status as a within subject factor showed only a main effect of familiarity status ($F(1,34) = 18.3, p = .0001$), but no main effect of test intonation ($p = .27$) or interaction between the two factors ($p = .95$).

**Discussion**

As in the peak stress alignment condition in Experiment 1, participants showed significantly longer looking times to novel than to familiar test lists, despite the fact that the intonation of the familiarization and test stimuli was different. We can therefore exclude the alternative explanation that infants in the medial peak condition of Experiment 1 extracted the trochaic test items only because of the intonational similarity between familiarization and test stimuli. The looking time difference to novel and familiar test lists of Experiment 2 instead corroborates our earlier interpretation that infants segment embedded SW units only when the stressed syllable is high pitched. The present data hence suggest that high pitch is an essential cue for perceiving a syllable as stressed and thus as a likely word onset for German nine month olds. We will return to this claim in the ‘General discussion’.

Two further aspects of Experiment 2 are noteworthy. First, our results extend earlier findings on infants’ abilities to generalize over certain prosodic aspects in the stimuli, such as speaker identity, pitch level, and emotions (Bortfeld & Morgan, 2010; Houston & Jusczyk, 2000; Singh, 2008; Singh et al., 2004). Our results show that nine month olds are also able to generalize over different intonational realizations (from falling in familiarization to rising in test). Infants seem to have formed representations of the units extracted from fluent speech that do not include pitch information, i.e. the correct kind of representations for speakers of an intonation language. Second, the results of Experiment 2 demonstrate that infants extract SW units from WSW carrier words, without the support of linguistic, phonetic, and statistical information. In fact, these cues all hint at a different word boundary, the boundary of the trisyllabic WSW carrier word (#WSW). Possibly, at a slightly older age, when infants are able to extract iambic (WS) patterns (Jusczyk, Houston & Newsome, 1999), the extraction of the SW units would become more difficult or impossible (comparable to the difficulty to activate embedded words, such as *date* from *sedate*; see Norris, Cutler, McQueen &
Butterfield, 2006). It is an open question whether infants in our study also extracted the whole WSW carrier word. On the one hand, the transitional probabilities and the frequent occurrence of schwa syllables before the WSW carrier words may have made it possible; on the other hand, the unstressed word onset in the WSW carriers may have prevented them from entertaining this kind of segmentation.

GENERAL DISCUSSION

The present study investigated German nine month olds’ ability to segment SW disyllables from fluent speech in three different intonation conditions (pitch peak realized before, on, or after the stressed syllable). In the medial peak condition, in which the pitch peak was aligned with metrical stress, infants extracted the SW part words from trisyllabic WSW carrier words, but they did so neither in the early peak nor in the late peak condition, the two intonation conditions in which the pitch peak was misaligned with the metrically stressed syllable (early peak and late peak condition). Thus, only high pitched stressed syllables were perceived as stressed and consequently taken as word onsets in our study. Experiment 2 replicated the results of the medial peak condition with test stimuli whose intonation differed from those of the familiarization stimuli. These data demonstrate that infants generalized over intonational realizations (from a falling contour in familiarization to a rising contour during test).

We start our discussion with infants’ ability to generalize, which ties in with previous studies showing that infants’ representations become more abstract towards the end of the first year of life (Bortfeld & Morgan, 2010; Houston & Jusczyk, 2000; Singh et al., 2004; Singh et al., 2008). Our data extend these previous findings by showing that infants do not only abstract over different voices and pitch levels (Houston & Jusczyk, 2000; Singh et al., 2008), but also over different pitch contours (falling vs. rising). It seems that nine month old infants are aware of the fact that pitch is not lexically contrastive in German, and they consequently do not store pitch together with the segmental form of the extracted units.

Considering infants’ ability to abstract over prosodic patterns, and their strong reliance on high pitched stressed syllables, it becomes clear that pitch plays different roles in the SEGMENTATION and RECOGNITION process, respectively. From the current data it appears that, in a first step, high $f_0$ seems to be essential in the perception of stress, and consequently needs to be present in order to extract units in fluent speech. In a second step, when the task demands the recognition of the previously embedded SW units, $f_0$ is no longer considered relevant in the comparison of stored forms to the input. Once infants have established a (be it only temporary) representation of the extracted sound sequence, they seem to generalize
over lexically non contrastive pitch contours. A study by Vihman, Nakai, DePaolis, and Hallé (2004) similarly reports that prosodic cues play a minor role in word recognition at this age: in a head turn preference experiment with English learning eleven month olds, they found that recognition of (untrained) familiar words was only delayed but not inhibited through the misplacement of stress, whereas segmental mismatches, particularly the mispronunciation of the initial consonant, hindered word recognition. While infants have learned to neglect information on pitch in recognition processes, high pitch seems to be important in German nine month olds’ stress perception, thus playing an important role in segmentation processes.

The question that arises from these findings is why high pitch should be crucial in perceiving stressed syllables for young infants exposed to an intonation language. Currently, we see three possible explanations. First, and least likely, the effect may not be due to the fact that the stressed syllable is high, but rather to the fact that it differs from the adjacent unstressed syllables in its height. Infants may be particularly sensitive to the stressed syllable when the neighboring syllables differ in pitch (e.g. LHL or HLH), but less sensitive when there is little change (LLH or HLL, as in the two misalignment conditions). Our medial peak stimuli, which prompted infants to extract the embedded units, optimally employ such an alternation (LHL). In future studies, we plan to use flipped pitch contours (inverting high pitch into low and vice versa) to test whether HLH patterns lead to the same results.

Second, high pitch might be considered relevant in infants’ stress perception because the alignment between high pitch and metrical stress is (a) very salient (high pitched syllables are more salient than low pitched syllables) and (b) very frequent (there are more pitch accents with high pitched stressed syllables than with low pitched stressed syllables in German). Let us briefly elaborate: on the one hand, high pitched stressed syllables are judged as more salient/prominent than low pitched stressed syllables (see Baumann & Röhr, 2015, for German adults). Further, high pitch has been shown to be a salient cue for infants in their linguistic grouping of synthetically manipulated stimuli: Bion et al. (2011) report that infants grouped sequences alternating in pitch as HL sequences, thus exploiting high pitch as a word onset cue when no other cues are available. On the other hand, high pitched stressed syllables are more frequent than low pitched stressed syllables, both in ADS (Peters, Kohler & Wesener, 2005) and in IDS (Zahner, Schönhuber, Grijzenhout & Braun, Konstanz prosodically annotated infant directed speech corpus (KIDS corpus), unpublished observations). For instance, a recent corpus study by Zahner et al. (unpublished observations) analyzed utterances directed towards infants younger than one year by sixteen different mothers (utterances
from eight mothers were retrieved from the CHILDES database, MacWhinney, 2000; utterances from another eight mothers stemmed from own recordings in the Baby Speech Lab at the University of Konstanz; in total, 524 intonational phrases, 832 pitch accents). The results of a fine grained analysis that considered the tonal patterns in the vicinity of stressed syllables found that, with 34% of all accents, medial peak accents are most frequent in IDS, whereas both early and late peak patterns are considerably less frequent (12% and 14%, respectively; a summary of this particular analysis is provided in ‘Appendix D’). These frequency counts are particularly relevant since they hold for those metrical patterns that matched the stimuli used in the current segmentation study, i.e. an accented syllable that was preceded and followed by one or more unstressed syllables. Hence, the medial peak condition seems to be most conducive to segmentation, prompting infants to extract SW units from the WSW carriers, as this is the most frequent pattern infants encounter in natural situations. The equally rare occurrence of the two misalignment conditions fits in well with our segmentation findings, namely infants’ similar behavior in the two misalignment conditions (no part word extraction). If stressed syllables are often high pitched (as shown in Zahner et al., unpublished observations), infants may treat different stress cues as equally relevant for signaling metrical prominence. When one cue is missing, the syllable might no longer be perceived as stressed, in analogy to findings showing that German eight month olds are able to distinguish different prosodic phrasings only when the prosodic phrase boundary is signaled by phrase final lengthening and a pitch movement (Wellmann, Holzgrefe, Truckenbrodt, Wartenburger & Höhle, 2012). Yet the roles of salience and frequency of high pitched stressed syllables are intertwined in German: high pitch enhances the acoustic salience of stressed syllables and at the same time this pitch pattern is the most frequent one. Possibly, other languages with different distributions of pitch accent types may enable us to dissociate the two factors.

Third and finally, the effect of high pitch may be powerful enough to serve as a sufficient cue in the perception of metrical stress, such that, due to its strong acoustic salience, high pitch on its own is a stronger segmentation cue for infants than metrical stress (at least in German or Dutch, where unstressed syllables are spectrally not as strongly reduced as in English; see Cutler, 2012; Delattre, 1969). If it is solely the salience of high pitch that is relevant (and not its alignment with a stressed syllable), we would expect German nine month olds to extract the last two syllables of a WWS carrier word produced with a LHL pattern as a SW sequence (e.g. rodi [ˈroːdi] from Parodie ‘parody’ [paˌroːˈdiː]). In future studies, we plan to investigate this possibility.
The influence of other word onset cues could only be minimized in our experimental paradigm by testing infants ‘mis segmentation’ (extraction of embedded nonsense words). Even though from an adult point of view the extraction of SW units out of WSW carriers has to be considered a ‘failure’, from the infants’ perspective, however, these ‘failures’ will be extremely rare, since the typical IDS input is mostly of a different nature. (Note that only 4% of the accented words are WSW words (see Zahner et al., unpublished observations), and there are only few everyday words with this stress pattern, most of them being associated with food, animals, or clothing, e.g. Banane ‘banana’, Karotte ‘carrot’, Kartoffel ‘potato’, Giraffe ‘giraffe’, Kaninchen ‘bunny’, Pullover ‘jumper’, Sandale ‘sandal’; the rarity of this word prosodic structure in German IDS transfers to German children’s early production attempts where WSW words are equally rare, compared to other languages with different distributions; see Lleó, 2002). On the other hand, trochaic words (SW) preceded by a weak syllable (e.g. an article), such as die Mama ‘the mommy’, der Papa ‘the daddy’, die Katze ‘the cat’, are very frequent (46% of the accented words followed this prosodic structure in Zahner et al., unpublished observations). In our study, in which infants were familiarized with WSW sequences in single carrier words (e.g. Lagune), infants may have used the mechanism that proved successful for the frequently occurring trochaic words in their input. This mechanism may have led to the extraction of the embedded trochaic nonword sequence (gune). What becomes clear from the frequency distributions of the word prosodic structure in IDS is that the extraction of the SW part of a WSW carrier does not harm or complicate first language acquisition: on the contrary, it is beneficial in most cases where infants are confronted with this pattern in the real world. In fact, it helps them successfully extract, e.g. Katze from a sequence die Katze. Thus, relying on the metrical segmentation strategy seems to be a useful first heuristic until infants have learned to integrate other segmentation cues.

In sum, the present study replicates the findings of previous research that indicate that infants exposed to stress timed languages are able to extract SW units from fluent speech. In addition, our results show that German nine month olds can segment SW syllable sequences from fluent speech even if they are embedded in WSW carrier words and thus provide misleading linguistic, phonetic, and statistical cues to word onsets. This finding further strengthens the crucial role of stressed syllables for segmentation in Germanic languages, and extends earlier studies on the metrical segmentation strategy in German, English, and Dutch (Bartels et al., 2009; Jusczyk, Houston & Newsome, 1999; Kuijpers et al., 1998). Importantly, this study is the first to manipulate utterance level intonation in a segmentation study. While some previous studies have used
exaggerated and lively productions which are characterized by longer durations and larger f0 excursions overall (which allegedly lead to different patterns of results than less exaggerated, more adult directed stimuli; e.g. Keren Portnoy et al., 2015; Thiessen, Hill & Saffran, 2005), our manipulation involved phonological intonation contrasts. Specifically, the durational structure and f0 excursions were similar across conditions, but the alignment of the pitch peak with respect to the stressed syllable was varied. Our data clearly demonstrate that these alignment differences, which lead to different phonological pitch accent types, are relevant: German nine month olds’ perception of stress is clearly modulated by utterance level intonation, such that only high pitched stressed syllables are perceived as stressed and therefore become likely word onsets.

REFERENCES


1359


APPENDIX A
Familiarization passages. **Bold** face refers to the trisyllabic carrier words; *italics* indicate other accentuated words in the sentence; translations are provided below each passage.

‘**Lagune**’ passage
Hier entstand eine **Lagune**. Die **Lagune** war **traumhaft**. Die blaue **Lagune** zieht **Leute** an. Eine kleine **Lagune** ist **schön**. Seine **Lagune** lag im **Süden**. Sie fotografierte ihre **Lagune**.

‘Here originated a lagoon. ‘The lagoon was wonderful. The blue lagoon attracts people. A small lagoon is nice. His lagoon was situated in the South. She took a photo of her lagoon.’

‘**Kasino**’ passage
Die **Stadt** plante ein **Kasino**. Es **sollte** ein großes **Kasino** werden. Seine **Frau** wollte ins **Kasino**. Das **Kasino** war noch ganz **neu**. Ein kleines **Kasino** ist nicht **schön**. Das neue **Kasino** wurde sehr **beliebt**.

‘The town was planning a casino. It should become a big casino. His wife wanted to go to the casino. The casino was still very new. A small casino is not nice. The new casino became very popular.’

‘**Kanone**’ passage
Der **Mann** wollte eine **Kanone**. Er **hatte** eine schwarze **Kanone** gesehen. Die neue **Kanone** war **teuer**. Die **Kanone** sollte lang **halten**. Der **Nachbar** verkaufte eine alte **Kanone**. Mit einer **Kanone** fühlt man sich **sicher**.

‘The man wanted a cannon. He had seen a black cannon. The new cannon was expensive. The cannon should last long. The neighbor sold an old cannon. With a cannon one feels safe.’

‘**Tirade**’ passage
Das **Mädchen** machte eine **Tirade**. Die **Tirade** wollte nicht **enden**. Es war eine große **Tirade**. Mit der **Tirade** war sie **vertraut**. Er **vertrug** ihre **Tirade** schlecht. Diese **Tirade** sollte bald **aufhören**.

‘The girl released a tirade. The tirade did not want to end. It was a big tirade. She was familiar with the tirade. He could not take her tirade. This tirade should stop soon.’
APPENDIX B

Mean values (and standard deviations) of the individual test lists in Experiment 1.

<table>
<thead>
<tr>
<th></th>
<th>gune-list</th>
<th>sino-list</th>
<th>none-list</th>
<th>rade-list</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fo-excursion of pitch fall in st</td>
<td>10.07 (1.72)</td>
<td>10.33 (1.85)</td>
<td>9.70 (2.01)</td>
<td>10.12 (1.81)</td>
</tr>
<tr>
<td>Duration of test word in ms</td>
<td>693 (95)</td>
<td>719 (44)</td>
<td>749 (104)</td>
<td>680 (86)</td>
</tr>
<tr>
<td>Duration of first syllable (stressed) in ms</td>
<td>331 (48)</td>
<td>361 (31)</td>
<td>406 (61)</td>
<td>358 (62)</td>
</tr>
<tr>
<td>Duration of second syllable (unstressed) in ms</td>
<td>361 (48)</td>
<td>358 (33)</td>
<td>342 (48)</td>
<td>321 (31)</td>
</tr>
</tbody>
</table>

APPENDIX C

Mean values (and standard deviations) of the individual test lists in Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>gune-list</th>
<th>sino-list</th>
<th>none-list</th>
<th>rade-list</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fo-excursion of pitch rise in st</td>
<td>12.18 (1.32)</td>
<td>12.11 (1.64)</td>
<td>12.12 (1.46)</td>
<td>12.54 (1.67)</td>
</tr>
<tr>
<td>Duration of test word in ms</td>
<td>664 (60)</td>
<td>735 (71)</td>
<td>680 (61)</td>
<td>666 (33)</td>
</tr>
<tr>
<td>Duration of first syllable (stressed) in ms</td>
<td>362 (49)</td>
<td>385 (90)</td>
<td>392 (83)</td>
<td>380 (53)</td>
</tr>
<tr>
<td>Duration of second syllable (unstressed) in ms</td>
<td>302 (32)</td>
<td>349 (79)</td>
<td>288 (22)</td>
<td>306 (20)</td>
</tr>
</tbody>
</table>
APPENDIX D

Distribution of fo movement around the accentual syllable (marked by ✪). The condition displayed matches the experimental stimuli used in Experiment 1 and Experiment 2.

<table>
<thead>
<tr>
<th>Fo-movement around accentual tone</th>
<th>Proportion (out of 426 accentual patterns)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>34%</td>
</tr>
<tr>
<td></td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>6%</td>
</tr>
</tbody>
</table>