Proc. of the 17th Speech Science and Technology Conference (SST 2018), Sydney, Australia, 73-76.

F0 peaks are a necessary condition for German infants' perception of stress in metrical segmentation

Katharina Zahner & Bettina Braun

Department of Linguistics, University of Konstanz
{katharina.zahner|bettina.braun}@uni-konstanz.de

Abstract

Infants exposed to stress-timed languages have been demonstrated to use the stressed syllable to localize word beginnings. More recently, intonation has been shown to interfere with stress-based segmentation [1, 2]: For instance, stress-based segmentation was limited to accented words with bell-shaped accents that had the f0 peak aligned with the stressed syllable ("medial-peak" accents); segmentation failed when the f0 peak preceded or followed the stressed syllable in the target word's contour. Here, we test whether metrical segmentation is caused by the f0 peak on the stressed syllable or by the tonal alternation (LHL, bell-shaped contour in [2]). This allows us to probe whether f0 peaks are necessary cues for metrical segmentation. To this end, we replicated Zahner, et al. [2] but used cup-shaped intonation contours on the targets resulting in a tonal alternation in the opposite direction (HLH). Looking times obtained in a head-turn preference experiment showed no evidence of segmentation for the cup-shaped contours. This suggests that an f0 peak is a necessary condition for the stressed syllables to be used in stress-based segmentation, at least for German infants.

Index Terms: stress, pitch accent type, infant, German

1. Introduction

In order to extract units from speech, infants use both general strategies, e.g., transitional probabilities between syllables, and language-specific cues ([3, 4] for overviews). Regarding language-specific cues, prosodic properties of the ambient language shape segmentation behavior. For example, infants from stress-timed languages develop a stress-based strategy and interpret stressed syllables as word onsets, for Dutch [5-8], English [9-13], and German [14-16].

Two recent studies on German showed that intonation affects infants' segmentation behavior [1, 2]. Specifically, using electrophysiological measures, Männel and Friederici [1] demonstrated that word-form recognition is influenced by accentuation, i.e., whether or not a trochee (*Sirup* ['zi:.rʊp] 'syrup') receives a pitch accent. Infants' event-related potentials (ERPs) showed that the recognition of trochees differs with age and is modulated by accentuation: 6-month-olds only recognized trochees that were accentuated in familiarization, surfacing in a positive ERP response 500ms after word onset, i.e., before the end of the trochee. At 9 months, recognition was independent of accentuation and manifested in a (mature) negative response 400ms after word onset; this effect was followed by a late negativity only for accentuated words. At 12 months, infants recognized words independent of accentuation during familiarization (negative response 350ms after word onset). Hence, from that study, we may infer that accentuation generally facilitates segmentation.

Yet, a behavioural study by Zahner, et al. [2] suggests pitch accent type and the resulting consequences of (mis)alignment between the f0 peak and the stressed syllable (rather than accentedness in general) to modulate the segmentation success for German infants. Acoustically, pitch accent types differ in the alignment of the f0 peak in regard to the stressed syllable, making the position of f0 peak an unreliable cue to stress [17], see Fig. 1. In medial-peak accents (H*), used to introduce new information to the discourse [18], the f0 peak and the stressed syllable coincide. In early-peak accents (H+L*), signaling accessible information [19], the f0 peak precedes the stressed syllables, while it follows the stressed syllable in late-peak contours (L*+H / L* H-^H%), commonly employed for sentence-initial topics [20] or in (polar) questions [21].

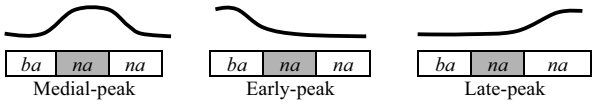

Figure 1: *Different pitch accent types on a trisyllabic word with stress on the second syllable (e.g., banana).*

Specifically, Zahner, et al. [2] familiarized German 9-months-olds with trisyllabic words (WSW stress pattern, e.g., *Lagune* [la.'gu:.nə] 'lagoon') in sentences; recognition of the embedded SW units (e.g., *gune* ['gu:.nə]) was tested. The WSW words were presented in one of 3 naturally occurring intonation conditions (between-subjects): medial-peak condition (peak-stress alignment), early- or late-peak contours (peak-stress misalignment), Fig. 1. Infants recognized the SW units only when f0 peak and the stressed syllable coincided. The authors concluded that a high-pitched accent (bell-shaped with the f0 peak on the stressed syllable, i.e., including a rising-falling movement) is important for stress perception and consequently segmentation.

Here, we test whether an f0 peak on the stressed syllable is a necessary cue for German infants to use a stressed syllable for metrical segmentation or whether it is the tonal alternation (LHL) that rendered the syllable with alternating pitch (the one with the high tonal target) particularly salient. In other words, infants may be particularly sensitive to the stressed syllable when the neighboring syllables differ in f0, e.g., LHL or HLH, but less sensitive when there is little change, LLH or HLL, as in the two misalignment conditions in [2]. This is compatible with an early view on stress perception on the utterance level by Bolinger [22], arguing that it is a "wide departure from a contour" (in any direction) that makes a syllable stand out [22: 112]. It would also be compatible with

a recent proposal on prominence perception according to which unexpected f0 events lead to increased prominence [23, 24]. In the current study, we tested cup-shaped contours, which have the tonal alternation in the opposite direction (HLH). If the change in f0 level is the necessary cue to trigger stress-based segmentation, *gune* is expected to be extracted in the cup-shaped contour. If, on the other hand, the f0 peak is the necessary cue, *gune* is not expected to be extracted. Another test case would be monotonous contours without any f0 movement. Likely, they would be perceived as unaccented, discarding them for use in the current paradigm (see [1]).

2. Experiment

2.1. Methods

2.1.1. Participants

Eighteen full-term infants (more than 37 weeks of gestation) from monolingual German families who finished the familiarization and all 12 test trials were included in the analyses (9 female, average age: 0;9.2 range: 0;8.18-0;9.17). Twelve further infants were tested but not included in the analysis due to crying (5), not attending to blinking lights (5), fussiness (1), and interference of a sibling (1).

2.1.2. Materials

For familiarization, we used the 4 passages from [2], which consisted of 6 sentences each. Each sentence contained the target once in different positions. The 4 target words were: *Kanone* [kʰa.ˈnoː.nə] 'cannon', *Lagune* [la.ˈguː.nə] 'lagoon', *Kasino* [kʰa.ˈsiː.no] 'casino', *Tirade* [tʰi.ˈraː.də] 'tirade'; (1) is an exemplar passage (SW part in trisyllabic carrier in bold; italics denote other accented words in the sentence).

(1) Hier entstand eine La**gune**. Die La**gune** war traumhaft. Die blaue La**gune** zieht Leute an. Eine kleine La**gune** ist schön. Seine La**gune** lag im Süden. Sie fotografierte ihre La**gune**.

'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.'

A female speaker, who was trained in intonational phonology, recorded the passages, realizing the WSW carrier words (e.g., *Lagune*) with a cup-shaped intonation contour (HL*H), Fig. 2. We used natural productions instead of resynthesized stimuli to ensure the same stimulus quality across experiments. Note that the (tritonal) contour is not described in the German intonational system [21], but it is indeed present in German infant-directed speech (henceforth IDS), occurring in 7% of the 426 accentual movements in the KIDS Corpus [25]. We matched the current WSW words closely to those in Zahner, et al. [2], see Tab. 1.

For test, the 4 lists from [2] were used. They consisted of 15 tokens of the SW part of the WSW word, i.e., 15 tokens of [ˈguːnə], 15 tokens of [ˈraːdə], 15 tokens of [ˈnoːnə], and 15 tokens of [ˈsiːno], respectively. The trochees were falling and had an inter-stimulus interval of 800ms, see [2: 1346].

2.1.3. Procedure

The procedure was identical to the one described in Zahner, et al. [2], which employed the classic head turn preference paradigm [26] in a sentence-word order [9]. Infants were tested in a three-sided black booth with blinking lights in the Baby Speech Lab at the University of Konstanz. Sound was infant controlled, such that a new trial was initiated when the infant looked away for more than 2 seconds (s) or at the end of the stimulus. An experimenter with tight-fitting headphones and masking music controlled the experiment from behind the booth. All parents gave written consent and filled in a questionnaire. During familiarization, infants listened to 2 (out of 4) passages until they had accumulated a listening time of at least 45 s to each passage (e.g., *Lagune*- and *Kasino*- passage). For test, all infants listened to the same 4 test lists (*gune*, *sino*, *rade*, *none*), 2 of which were the SW units of the familiarized words (thus familiar) and 2 of which were not (and thus novel to the infants). Familiarization was counterbalanced across participants such that half of the infants were familiarized with the *Lagune*- and *Kasino*-passages and the other half with the *Kanone*- and *Tirade*-passages. The pseudo-randomized order of presentation of stimuli (right, left) was identical to [2].

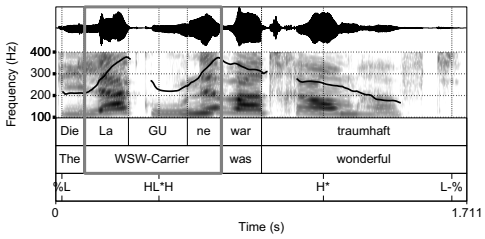

Figure 2: *Example target sentence in the cup-shaped intonation condition.*

Table 1. *Acoustic realization (mean values and standard deviations) of target words in the familiarization. Last column taken from [2: 1347].*

Acoustic variable	Cup-shape		Bell-shape [2]
F0 excursion of movement in st	HL* 8.4 (1.0)	L*H 8.5 (0.9)	LH* 8.8 (1.7)
Duration of first syllable (unstressed) in ms	196 (21)		182 (36)
Duration of second syllable (stressed) in ms	259 (20)		253 (25)
Duration of third syllable (unstressed) in ms	187 (50)		193 (66)
H1*-A3* ratio in middle of first vowel in dB	22.9 (7.4)		21.5 (6.5)
H1*-A3* ratio in middle of second vowel in dB	34.5 (10.7)		31.5 (14.1)
H1*-A3* ratio in middle of third vowel in dB	31.3 (12.4)		23.2 (5.3)

2.2. Results

Looking times were averaged by *familiarity status* (familiar vs. novel) for each infant. Infants looked 9.6s (sd = 3.5s) to familiar and 9.8s (sd = 3.4s) to novel test lists, see Fig. 3 (left). Nine infants out of 18 looked longer to the novel lists. A pairwise t-test indicated that the difference in looking times was not significant ($t(17) = 0.3$, $p > 0.78$). The null hypothesis was 3 times more likely than the alternative hypothesis (Bayes Factor = 0.3 [27]). To compare, infants in the bell-shaped condition in [2] looked 1.4s longer to novel lists, see Fig. 3 (right). To corroborate the difference across intonation

conditions, we pooled the data and tested for an interaction between *familiarity status* and *Experiment* in a repeated measures ANOVA with *familiarity status* as within-subjects factor and *Experiment* as between-subjects factor. The interaction between the two factors approached significance ($F(1,34) = 3.32$, $p = 0.08$). For the sake of completeness, Fig. 4 contrasts the 95% CIs of the looking time differences between the current study (left) and the study in [2] (right).

Hence, in contrast to the bell-shaped condition in [2], infants did not extract the trochaic part-word (SW) from the cup-shaped WSW carrier in this experiment. A mere alternation of tonal targets of opposite pitch height (HL*H) did not lead to a percept of lexical stress and the use in stress-based segmentation for the syllable that had the deviant pitch. The results obtained in the current experiment pattern with the misalignment conditions in Zahner, et al. [2], i.e., the early-peak and late-peak conditions in which the f0 peak is realized before or after the stressed syllable, respectively, speaking in favor of a high-pitched accented syllable that guides the segmentation process, at least in German infants.

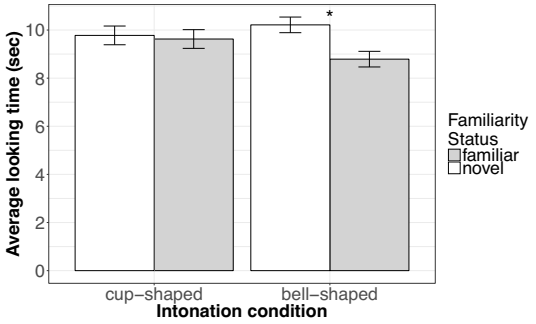

Figure 3: *Average looking times in the cup-shaped condition (left) and bell-shaped condition in [2] (right); whiskers represent ±1 standard error of the mean.*

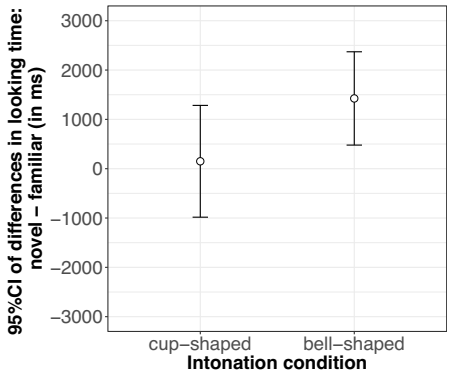

Figure 4: *Average looking time difference in the cup-shaped condition (left) and bell-shaped condition in [2] (right); whiskers represent the 95% confidence interval of the difference in looking time.*

3. General Discussion

The findings in Zahner, et al. [2] and in the current experiment together contribute to our knowledge on what leads to an infant's perception of a syllable as *stressed* [28, for discussion] and its use for segmentation. Zahner, et al. [2] showed that infants only succeed in extracting the trochee out of trisyllabic carrier words if the target is realized with a pitch accent type that renders the stressed syllable high-pitched, i.e., a medial-peak accent (bell-shaped, involving a rising-falling movement). Infants failed to extract the trochee when presented with early-peak or late-peak realizations on the target word. In the current experiment, infants also fail when the target is realized with a flipped medial-peak accent (cup-shaped) that pertains tonal alternation, but differs from the medial peaks in Zahner, et al. [2] in tonal height (LH*L vs. HL*H). These data suggest the f0 peak on a stressed syllable, comprising a rising-falling movement, to be a necessary cue for perceiving a syllable as stressed and thus as a word onset for German infants. The finding extends earlier research on accentedness and segmentation in German [1] by showing that high-pitched accent types are beneficial for segmentation.

Results from artificial language studies on linguistic grouping corroborate our explanation: [29, 30] show that infants tend to group syllable sequences that alternate in height (…H-L-H-L-H…) into trochaic units, taking the high-pitched syllables as the strong element. Currently, we see two explanations why high-pitched syllables (in our case high-pitched stressed syllables) are relevant for infants: first, medial-peak accents are most frequent in German IDS [25] (and also in German adult-directed speech [31]), which is why German infants most often encounter high-pitched stressed syllables, consequently fusing the concept of pitch peak and lexical stress; second, high-pitched stressed syllables might perceptually be more salient for German infants than low-pitched ones (which could either be a result of their acoustic nature or, rather circularly, a phenomenon which is prompted by the high frequency of occurrence itself). Teasing apart these two explanations is a challenge for future research.

Recent studies emphasized the relevance of lively and exaggerated contours in facilitating segmentation in German [32] and English [13]. These IDS productions show larger f0 excursions (phonetic modifications in pitch), along with durational differences. The findings in [2] and in the current paper demonstrate that different pitch accent types, i.e., alignment differences between peak and stress also modulate the segmentation success, with the f0 peak being a necessary cue to stress and segmentation. Future research needs to examine whether the segmentation success in the bell-shaped condition is also affected by the size of the f0 excursion.

We are currently investigating whether f0 peaks on unstressed syllables are sufficient to initiate the use of a syllable as a word onset in German infants, in addition to being a necessary cue. Visual-world eye tracking data from German (and Australian English) shows that this is indeed the case for adults: Adult listeners temporarily activated an initially stressed cohort competitor (e.g., *musical*) when a target, stressed on the second syllable (e.g., *museum*), was realized with an early-peak accent, i.e., an accent type in which the f0 peak preceded the stressed syllable [33, 34].

4. Acknowledgements

We thank Jana Neitsch, Stephanie Gustedt and Johanna Schnell for recording, recruiting and testing. We thank Janet Grijzenhout and Muna Schönhuber for discussion of the data.

5. References

[1] Männel, C. and Friederici, A. D., "Accentuate or repeat? Brain signatures of developmental periods in infant word recognition," *Cortex,* vol. 49, pp. 2788-2798, 2013.

[2] Zahner, K., Schönhuber, M., and Braun, B., "The limits of metrical segmentation: Intonation modulates infants' extraction

of embedded trochees," *Journal of Child Language,* vol. 43, pp. 1338-1364, 2016.

[3] Junge, C., "The proto-lexicon: Segmenting word-like units from the speech stream," in *Early word learning*, Westermann, G. and Mani, N., Eds., Abingdon, New York: Routledge, 2018, pp. 15-29.

[4] Saffran, J. R. and Kirkham, N. Z., "Infant statistical learning," *Annual Reviews,* vol. 69, pp. 181-203, 2018.

[5] Kuijpers, C. T., Coolen, R., Houston, D. M., and Cutler, A., "Using the head-turning technique to explore cross-linguistic performance differences," in *Advances in infancy research*. vol. 12, Rovee-Collier, C., Lipsitt, L., and Hayne, H., Eds., Stamford: Ablex, 1998, pp. 205-220.

[6] Kooijman, V., Hagoort, P., and Cutler, A., "Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech," *Cognitive Brain Research,* vol. 24, pp. 109-116, 2005.

[7] Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., and Cutler, A., "Predictive brain signals of linguistic development," *Frontiers in Psychology,* vol. 4, p. 25, 2013.

[8] Junge, C., Kooijman, V., Hagoort, P., and Cutler, A., "Rapid recognition at 10 months as a predictor of language development," *Developmental Science,* vol. 15, pp. 463-473, 2012.

[9] Jusczyk, P. W., Houston, D. M., and Newsome, M., "The beginnings of word segmentation in English-learning infants," *Cognitive Psychology,* vol. 39, pp. 159-207, 1999.

[10] Cutler, A., Junge, C., Spokes, T., and Kidd, E., "Phonological acquisition: Stress-based segmentation in English," Paper presented at the Laboratory Phonology (LabPhon16), Lisbon, Portugal, 2018.

[11] Houston, D. M., Kuijpers, C., Coolen, R., Jusczyk, P. W., and Cutler, A., "Cross-language word segmentation by 9-month-olds," *Psychonomic Bulletin & Review,* vol. 7, pp. 504-509, 2000.

[12] Mason-Apps, E., Stojanovik, V., Houston-Price, C., and Buckley, S., "Longitudinal predictors of early language in infants with Down Syndrome: A preliminary study," *Research in Developmental Disabilities,* vol. 81, pp. 37-51, 2018.

[13] Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S*., et al.*, "British English infants segment words only with exaggerated infant-directed speech stimuli," *Cognition,* vol. 148, pp. 1-9, 2016.

[14] Höhle, B., "Der Einstieg in die Grammatik: Die Rolle der Phonologie/Syntax- Schnittstelle für Sprachverarbeitung und Spracherwerb [First steps into grammar: The role of phonology/syntax interface for language processing and acquisition]," Habilitationsschrift, Freie Universität Berlin, Berlin, Germany, 2002.

[15] Bartels, S., Darcy, I., and Höhle, B., "Schwa syllables facilitate word segmentation for 9-month-old German-learning infants," in *33rd Annual Boston University Conference on Language Development*, Somerville, M.A., 2009, pp. 73-84.

[16] Altvater-Mackensen, N. and Mani, N., "Word-form familiarity bootstraps infant speech segmentation," *Developmental Science,* vol. 16, pp. 980-990, 2013.

[17] Ladd, D. R., *Intonational phonology*, 2nd ed. Cambridge: Cambridge University Press, 2008.

[18] Kohler, K., "Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics," *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung der Universität Kiel (AIPUK),* vol. 25, pp. 115-185, 1991.

[19] Baumann, S. and Grice, M., "The intonation of accessibility," *Journal of Pragmatics,* vol. 38, pp. 1636-1657, 2006.

[20] Braun, B., "Phonetics and phonology of thematic contrast in German," *Language and Speech,* vol. 49, pp. 451-493, 2006.

[21] Grice, M., Baumann, S., and Benzmüller, R., "German intonation in autosegmental-metrical phonology," in *Prosodic typology. The phonology of intonation and phrasing*, Sun-Ah, J., Ed., Oxford: Oxford University Press, 2005, pp. 55-83.

[22] Bolinger, D., "A theory of pitch accent in English," *Word,* vol. 14, pp. 109-49, 1958.

[23] Kakouros, S., Salminen, N., and Räsänen, O., "Making predictable unpredictable with style – Behavioral and electrophysiological evidence for the critical role of prosodic expectations in the perception of prominence in speech," *Neuropsychologia,* vol. 109, pp. 181-199, 2018.

[24] Kakouros, S. and Räsänen, O., "Perception of sentence stress in speech correlates with the temporal unpredictability of prosodic features," *Cognitive Science,* vol. 40, pp. 1739-1774, 2016.

[25] Zahner, K., Schönhuber, M., Grijzenhout, J., and Braun, B., "Konstanz prosodically annotated infant-directed speech corpus (KIDS corpus)," in *Proceedings of the 8th International Conference on Speech Prosody*, Boston, USA, 2016, pp. 562-566.

[26] Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., and Gerken, L., "The head-turn preference procedure for testing auditory perception," *Infant Behavior and Development,* vol. 18, pp. 111-116, 1995.

[27] Wagenmakers, E.-J., "A practical solution to the pervasive problems of p values," *Psychonomic Bulletin & Review,* vol. 14, pp. 779-804, 2007.

[28] Bhatara, A., Boll-Avetisyan, N., Höhle, B., and Nazzi, T., "Early sensitivity and acquisition of prosodic patterns at the lexical level," in *The development of prosody in first language acquisition*. vol. 23, Esteve-Gibert, N. and Prieto, P., Eds., Amsterdam [u.a.]: John Benjamins, 2018, pp. 37-57.

[29] Bion, R. A. H., Benavides-Varela, S., and Nespor, M., "Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences," *Language and Speech,* vol. 54, pp. 123-140, 2011.

[30] Abboub, N., Boll-Avetisyan, N., Bhatara, A., Höhle, B., and Nazzi, T., "An exploration of rhythmic grouping of speech sequences by French- and German-learning infants," *Frontiers in Human Neuroscience,* vol. 10, p. 292, 2016.

[31] Peters, B., Kohler, K., and Wesener, T., "Melodische Satzakzentmuster in prosodischen Phrasen deutscher Spontansprache - Statistische Verteilung und sprachliche Funktion [Meldodic sentence accent patterns in prosodic phrases of German sponteneous speech - Statistical distribution and linguistic function]," in *Prosodic structures in German spontaneous speech (AIPUK 35a)*, Kohler, K., Kleber, F., and Peters, B., Eds., Kiel: IPDS, 2005, pp. 185-201.

[32] Schreiner, M. S. and Mani, N., "Listen up! Developmental differences in the impact of IDS on speech segmentation," *Cognition,* vol. 160, pp. 98-102, 2017.

[33] Zahner, K., Kember, H., and Braun, B., "Mind the peak: When museum is temporarily understood as musical in Australian English," in *Proceedings of the 18th Annual Conference of the International Speech Communication Association (Interspeech)*, Stockholm, Sweden, 2017, pp. 1223-1227.

[34] Zahner, K., Schönhuber, M., Grijzenhout, J., and Braun, B., "Pitch accent type affects lexical activation in German: Evidence from eye tracking," in *Proceedings of the 16th Speech Science and Technology Conference (SST),* Sydney, Australia, 2016.