Journal of Phonetics 74 (2019) 75–95

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Special Issue: Integrating Phonetics and Phonology, eds. Cangemi & Baumann

Alignment of f0 peak in different pitch accent types affects perception of metrical stress

Katharina Zahner[*], Sophie Kutscheid, Bettina Braun

Department of Linguistics, University of Konstanz, PO186, 78457 Konstanz, Germany

ARTICLE INFO

Article history:
Received 27 March 2018
Received in revised form 16 January 2019
Accepted 23 February 2019
Available online 6 April 2019

Keywords:
Stress judgements
Lexical activation
Eye-tracking
Lexical stress
Pitch accent type
German

ABSTRACT

In intonation languages, pitch accents are associated with stressed syllables, therefore accentuation is a sufficient cue to the position of metrical stress in perception. This paper investigates how stress perception in German is affected by different pitch accent types (with different f0 alignments). Experiment 1 showed more errors in stress identification when f0 peaks and stressed syllables were not aligned – despite phonological association of pitch accent and stressed syllable. Erroneous responses revealed a response bias towards the syllable with the f0 peak. In a visual-world eye-tracking study (Experiment 2), listeners fixated a stress competitor with initial stress more when the spoken target, which had penultimate stress, was realized with an early-peak accent (f0 peak preceding stressed syllable), compared to a condition with the f0 peak on the stressed syllable. Hence, high-pitched unstressed syllables are temporarily interpreted as stressed – a process directly affecting lexical activation. To investigate whether this stress competitor activation is guided by the frequent co-occurrence of high f0 and lexical stress, Experiment 3 increased the frequency of low-pitched stressed syllables in the immediate input. The effect of intonation on competitor fixations disappeared. Our findings are discussed with respect to a frequency-based mechanism and their implications for the nature of f0 processing.

1. Introduction

In stress languages, stress refers to an abstract specification of metrical strength in the lexicon (Ladd, 2008, p. 58). For listeners, there are a number of phonological and phonetic cues to stress: Phonological cues include syllable complexity (stressed syllables are more complex than unstressed syllables, e.g., Berg, 1998, p. 112) and utterance-level accentuation (only stressed syllables can be accented, Ladd, 2008). Phonetic cues to stress reported in the literature are longer syllable duration, higher intensity, a more peripheral vowel quality, more vocal effort, and higher f0 (for German, see Delattre, 1969; Dogil, 1995; Jessen, Marasek, & Claßen, 1995; Mooshammer, 2010; Mooshammer & Geng, 2008; see Gordon & Roettger, 2017 for an overview). The current paper deals with the processing of lexical stress cues and the question of how utterance-level intonation (f0) modulates the perception of lexical stress in German. Different from other studies in this Special Issue (Bishop, Kuo, & Kim, this Special Issue; Cole et al., this Special Issue; Wagner, Cwiek, & Samlowski, this Special Issue), which focused on phrase-level prominence (i.e., a word standing out from the other words in the phrase), we address the perception of word-level prominence, i.e., lexical stress in German. German, like English, Dutch, Italian or Spanish, is a free stress language (Hyman, 1977). Hence, the position of the stressed syllable is not fixed to the same position across all words. Processing stress information correctly reduces the number of lexical competitors and makes lexical processing more efficient (e.g., Cooper, Cutler, & Wales, 2002; Cutler, Wales, Cooper, & Janssen, 2007; Friedrich, Kotz, Friederici, & Gunter, 2004; Jesse, Poellmann, & Kong, 2017; Reinisch, Jesse, & McQueen, 2010).

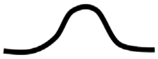
Stress cues differ in how reliably they signal the stressed syllable: While the phonological feature *accentuation* is an unambiguous cue to the position of metrical stress (irrespective of whether it is a high (H*) or a low accent (L*), cf. Dilley & Heffner, 2013; Shattuck-Hufnagel, Dilley, Veilleux, Brugos, & Speer, 2004), the phonetic cue *higher f0* differs depending on pitch accent *type*. Intonation languages typically make use of

[*] Corresponding author.
E-mail addresses: katharina.zahner@uni-konstanz.de (K. Zahner), sophie.kutscheid@uni-konstanz.de (S. Kutscheid), bettina.braun@uni-konstanz.de (B. Braun).

https://doi.org/10.1016/j.wocn.2019.02.004
0095-4470/© 2019 The Authors. Published by Elsevier Ltd.

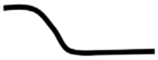
a range of different pitch accent types, mostly for information-structural purposes (Gussenhoven, 2004; Ladd, 2008, for overviews). Across accent types, the alignment of tonal targets (f0 peaks and f0 valleys) with the stressed syllable differs. For instance, in medial-peak accents, typically used to introduce new referents to the discourse (Kohler, 1991; Pierrehumbert & Hirschberg, 1990), the f0 peak is aligned with the stressed syllable (Example 1a, L+H*, capitals indicate lexical stress), while in early-peak accents, signalling semi-active and thus inferable information (Baumann & Grice, 2006; Kohler, 1991), the f0 peak precedes the stressed syllable and is thus aligned with an unstressed syllable (Example 1b, H+L*).

1a) Medial-peak accent, L+H*



German:	Das	ist	eine	Ba	NA	ne.
	This	is	a	ba	NA	na.

1b) Early-peak accent, H+L*



German:	Sie	isst	gern	Obst.	Am liebsten	Ba	NA	nen.
	She	eats	[ADV]	fruit	Preferably	ba	NA	nas.

In so-called early-peak or late-peak accents, in which f0 peaks either precede or follow the stressed syllable they are associated with, listeners cannot rely on the phonetic cue *high f0* for the detection of metrical stress.[1] Despite the theoretically ambivalent role of f0 in signalling metrical stress in intonation languages, studies that systematically manipulated acoustic cues to stress showed that listeners strongly attend to high f0 levels and f0 peaks for the perception of lexical stress (e.g., Fry, 1958; Isačenko & Schädlich, 1966; Kohler, 2008; Niebuhr & Winkler, 2017). As will be outlined in more detail in the Background, these studies typically used stimuli with small, step-wise changes of acoustic cues that are not necessarily attested in natural speech. In this paper, we test whether and how naturally occurring differences in f0-peak alignment relative to the stressed syllable (as present in different pitch accent types) change the perception of metrical strength relations in words by (a) rendering genuinely unstressed syllables perceptually stressed and (b) affecting lexical stress processing and online word recognition.

1.1. Background

Lexical stress plays an important role in online spoken word recognition in stress languages (e.g., Cutler, 2012). Here, we focus on free stress languages that encode stress more by suprasegmental than by segmental cues, e.g., German, Dutch, Italian or Spanish. In a cross-modal priming study, Donselaar, Koster, and Cutler (2005), for instance, showed that Dutch listeners reacted faster to a visually presented target word if a preceding auditory prime had the same stress pattern, e.g., [ˈɔk.to] for [ˈɔk.to.pus] ‘octopus’ compared to an unrelated control prime. On the other hand, listeners were slowed down when there was a mismatch between the stress pattern of the prime and the target, e.g., [ˈɔk.to] for [ɔk.ˈto.bər] ‘october’ (cf. Soto-Faraco, Sebastián-Gallés, & Cutler, 2001; Tagliapietra & Tabossi, 2005). In visual-world eye-tracking studies, listeners’ fixations are also sensitive to suprasegmental stress cues (Reinisch et al., 2010; Sulpizio & McQueen, 2012). Reinisch et al. (2010), for instance, showed that suprasegmental information (duration, f0, spectral tilt, and intensity) immediately constrains lexical competition: Dutch listeners showed more fixations to a target (e.g., [ˈɔk.to.pus]) than to its stress competitor (e.g., [ɔk.ˈto.bər]) before phonemic information disambiguated the pair.

Among the acoustic correlates to stress reported in the literature (see above), the role of f0 is disputed for many languages. In a recent meta-analysis, Roettger and Gordon (2017) analysed 110 studies in 75 languages, 85 of which investigated lab speech. The authors showed that only 22% of these 85 lab studies teased apart phrase-level prominence (pitch accent) and word-level prominence (metrical stress) properly, leaving a high number of studies in which phrase-level accentuation is confounded with word-level stress. Indeed, results from various production studies with linguistically diverse speech materials show that f0 is not a direct acoustic correlate of stress (e.g., Dogil, 1995; Kochanski, Grabe, Coleman, & Rosner, 2005; Sluijter & Van Heuven, 1996a, 1996b; Szalontai, Wagner, Mády, & Windmann, 2016). What is undisputed, however, is the fact that intonational pitch accents only associate with stressed syllables (Ladd, 2008), so that the presence of a pitch accent is an unambiguous cue to the position of metrical stress (Dilley & Heffner, 2013; Shattuck-Hufnagel et al., 2004; Shattuck-Hufnagel, Ostendorf, & Ross, 1994). For instance, Shattuck-Hufnagel et al. (1994) demonstrated in a labelling study with trained annotators that stress shift in multisyllabic words such as *Massachusetts* – in their study a shift of perceived primary stress from the penultimate or final syllable to the initial syllable – occurred if the first syllable was pitch-accented (predominantly with an f0 rise). Dilley and Heffner (2013) also provide evidence for accentuation as a cue to stress. They investigated whether alignment differences in f0 peaks and valleys shift the perception of the accent type category, e.g., from H* on the first syllable to H+L* on the final syllable (see their Experiment 4, p. 41ff.). Specifically, listeners heard words and nonce words with primary stress on the first syllable and secondary stress on the final syllable, and one or two unstressed syllables in-between, e.g., *millionaire* (S_1WS_2), *lemonade* (S_1WS_2), *lannameraine* (S_1WWS_2).[2] *Millionaire* and *lannameraine* were realized with an H* on S_1 and ended in a low boundary tone; *lemonade* had an L* on S_1 and a high final boundary tone. For *millionaire*

[1] We do not claim that alignment differences are phonetic in nature: H+L* and L+H* are clearly distinct phonological accent types (cf. categorical perception experiments in Kohler, 1991). However, regarding the processing of metrical stress, different accent types cause phonetic differences in the alignment of tonal targets with respect to the phonological property *accentuation* (Ladd, 2008).

[2] Note that in the remainder of this paper, we adopt the convention of using S for strong (stressed) and W for weak (unstressed) syllables; indexing indicates primary (S_1) and secondary stress (S_2).

and *lannameraine*, the f0 peak was shifted to the right in various steps so that it ended up on the unstressed syllable; for *lemonade*, the f0 valley was shifted to the right in various steps so that it was ultimately realized on the unstressed syllable preceding the secondary stressed syllable. Listeners had to indicate whether the first or the last syllable carried primary stress. The results showed that the later the peak and the valley respectively, the more frequently listeners judged stress to be on the last syllable of the word. The authors interpreted this shift in perception as a shift from H* associated with S_1 to H+L* associated with S_2 for the *millionaire/lannameraine*-series and from L* associated with S_1 to L+H* associated with S_2 in the *lemonade*-series. These findings indicate that the alignment of f0 peaks and f0 valleys alter the perception of pitch accent type and primary stress of the syllable the accent is associated with. This in turn demonstrates the relevance of the phonological cue *accentuation* for the identification of stress.

Despite the ambivalent role of f0 as a stress correlate, high pitch generally appears to cue metrical strength in perception. For instance, in decision tasks that manipulated different acoustic cues (f0, duration, intensity) in a step-wise fashion, listeners of West Germanic languages strongly attended to pitch information for stress perception, regardless of whether it were rising f0 movements, high f0 levels or f0 peaks (cf. Fry, 1958; Isačenko & Schädlich, 1966; Kohler, 2008; Niebuhr & Winkler, 2017).[3] In a seminal study, Fry (1958) showed that English listeners primarily relied on a high f0 level to decide between stress minimal pairs, e.g., the noun *object* and the verb *object*; even more so than on duration or intensity, which are inherent correlates of metrical stress. He found that the relation between f0 levels on adjacent syllables functions in an all-or-none fashion, with a step-up in f0 leading to an iambic perception and a step-down to the perception of a trochee. Fry (1958) also tested the effect of more complex (but non-linguistic) f0 movements on stress perception and showed that rising movements most distinctly triggered the percept of metrical stress – indirectly via the perception of utterance-level pitch accents. For German, Isačenko and Schädlich (1966) similarly showed that a manipulation of tonal scaling in discrete monotone pitch levels changed the perception of the stressed syllable in a word (from secondary to primary stress and vice versa): *übersetzen* [ˌyː.bɐ.ˈzɛ.ʦn̩] (‘to translate’) was perceived as *übersetzen* [ˌyː.bɐ.ˈzɛ.ʦn̩] (‘to cross over/ferry over’) when the pitch level on *über* [yː.bɐ] was increased in tonal height (Isačenko & Schädlich, 1966, p. 22). Only recently, Niebuhr and Winkler (2017) showed that an increase of 0.5 semitones (st) in height of the f0 peak counterbalanced a 30% increase in syllable duration for the distinction between primary and secondary stress. Kohler (2008) demonstrated that listeners judged the second syllable as more prominent in bisyllabic nonce syllable strings *baba* as soon as an f0 peak (with a preceding 1–2 st rise to the peak and a subsequent fall of 6–7 st) was realized on the second syllable. Conversely, a falling movement of 2–4 st on the second syllable, resulting in a peak-like contour on the first syllable, led to the perception of a trochee.

In sum, these studies indicate that different kinds of high pitch (f0 rises, high f0 levels, and f0 peaks) function as a perceptual cue to metrical strength – even though utterance-level intonation renders f0 a highly unreliable cue to the position of the stressed syllable. However, these studies leave open a number of questions: First, they do not provide evidence on how f0-alignment differences as they occur in naturally produced pitch accents affect the processing of stress. Second, the question of whether and how different pitch accent types affect *online* stress processing, as the utterance unfolds over time, has not been studied before. Third, it is an open issue whether f0 plays an equally important role in words that are not a member of a stress minimal pair. Not surprisingly, previous studies used such minimal pairs (Fry, 1958; Isačenko & Schädlich, 1966; Niebuhr & Winkler, 2017) with two separate lexical entries in the mental lexicon (one for each meaning), or nonce words (Kohler, 2008), which lack lexical stress specifications altogether. However, these stress minimal pairs and nonce words are special and may not reflect well how f0 is used in stress perception: Stress minimal pairs are very rare in German non-compounded words, so one is left with minimal pairs in which the members have very different lexical frequencies or with compounds that differ in primary and secondary stress. The asymmetry in lexical frequency, for instance, may lead to strategic responses on part of the listeners as stress placement has been shown to differ in high- vs. low-frequency words (see e.g., Cutler & Carter, 1987, who report that English high-frequency words show a greater proportion of initial stress than English low-frequency words). Also, stress minimal pairs and nonce words increase the probability of stress on those syllables that can be stressed, which may also evoke strategic responses. Furthermore, it cannot be ruled out that such stress minimal pairs are processed in a special way altogether.

In the current study, we therefore test intonation contours that are modelled on naturally produced pitch accent contrasts and investigate the effect of different pitch accent types on both offline and online perception of metrical stress. We use target words that only occur with one stress pattern, following the tradition of psycholinguistic stress studies (e.g., Cooper et al., 2002; Donselaar et al., 2005; Jesse et al., 2017; Reinisch et al., 2010; Soto-Faraco et al., 2001; Sulpizio & McQueen, 2012; Tagliapietra & Tabossi, 2005). Against the above reviewed background, we hypothesize that f0 peaks that are not aligned with the stressed syllable, i.e., in early- or late-peak contours, hamper stress identification and lexical activation in German because f0 peaks have been shown to cue metrical stress. Specifically, we expect f0 peaks that are realized on unstressed syllables – as leading tones in early-peak accents (H+L*) or as high boundary tones following L*-accents (L* H-^H%, henceforth “late-peak contour”) – to impede stress judgements (Experiments 1a and 1b) and to temporarily activate lexical competitors with a different stress pattern (Experiment 2). If this is indeed the case, Experiment 3 will subsequently examine one of the underlying factors that could make high f0 a cue to stress. Option 1 is that high-pitched syllables stand out perceptually (Cho, 2005; Hsu, Evans, & Lee, 2015; Lieberman, 1967). Option 2 is that

[3] On the level of the phrase, f0 has been demonstrated to be a relevant cue to the perception of phrasal prominence in German (Andreeva, Barry, & Wolska, 2012; Baumann, 2014; Baumann & Röhr, 2015; Baumann & Winter, 2018; Wagner, Cwiek, & Samlowski, 2016, this Special Issue). In this regard, various perception tasks have been used, such as prominence rating tasks with gradient scales (e.g., Baumann & Röhr, 2015) or binary options (e.g., Baumann, 2014; Baumann & Winter, 2018), tasks in which the appropriateness of the communicative intent of a sentence was judged (e.g., Andreeva et al., 2012), or gestural tasks such as drumming (Wagner et al., 2016, this Special Issue).

listeners learned to associate high-pitched syllables with metrical stress because of a frequent occurrence of H*-accents. In Experiment 3, we manipulate the frequency of occurrence of high-pitched stressed syllables (by only providing low-pitched accented syllables). If Option 1 applies, listeners are expected not to be affected by the manipulation; if Option 2 applies, we expect a reduced (or no) stress competitor effect in Experiment 3. This allows us to test whether there is a direct mapping between the acoustic cue f0 to metrical stress or whether the processing of f0 is mediated through phonological categories (accent types). Note that all materials and details on the statistical analysis steps are available at Mendeley https://doi.org/10.17632/2gkpwpg44j.3, cf. Roettger (2019) for a recent discussion of transparency in phonetic research.

2. Experiment 1: Lexical stress judgements

Experiments 1a and 1b examine whether different alignments of the f0 peak relative to the stressed syllable affect the identification of metrical stress. We conducted a decision task with trisyllabic German nouns (weak-strong-weak, WSW) and three possible response options (stress on the first, the second, or the third syllable). Three different f0-alignment contrasts were tested: early-peak accents (H+L* L-%), medial-peak accents (L+H* L-%), and L*-accents followed by a high boundary tone, i.e., late-peak contours (L* H-^H%). For the sake of simplicity, we refer to the medial-peak contour as the "peak-stress-alignment" condition, while we will refer to the early-peak and late-peak contours as "peak-stress-misalignment" conditions. Note that the term "peak-stress-misalignment" solely denotes the fact that the f0 peak is not aligned with the stressed syllable, but does not imply any kind of erroneous alignment. We hypothesize that listeners give fewer correct responses in their stress judgements in the misalignment conditions (early-peak and late-peak contours) compared to the alignment condition (medial-peak contours). We further predict that erroneous stress judgements show a bias towards the syllable with the f0 peak, such that with early-peak contours most errors involve syllable-1-responses and with late-peak contours most errors involve syllable-3-responses.

Experiments 1a and 1b differed in the way the stimuli were PSOLA-resynthesized (Pitch Synchronous Overlap Add, see Boersma & van Heuven, 2001). This is done to rule out that the syllable the f0 peak is aligned with may inadvertently be realized with higher intensity and longer duration in natural speech than syllables without an f0 peak (Kohler, 1991, p. 144; Niebuhr, 2007, pp. 117-150). The main resynthesis procedure was as follows (details are provided in the materials sections below): In Experiment 1a, the medial-peak contours were PSOLA-resynthesized half from naturally produced early-peak contours and half from naturally produced late-peak contours (medial-peak contours were only adapted in scaling). That way, the misalignment conditions had the most natural sound quality, with the smallest degree of manipulation (only the scaling was resynthesized), but possibly the smallest acoustic difference between the unstressed and the stressed syllable. Since syllables on which the f0 peak is realized may additionally be produced with higher intensity and longer duration – cues that also signal stress – the acoustic difference between stressed and unstressed syllables tends to be reduced (Niebuhr, 2007, p. 138). To isolate effects of f0 alignment, Experiment 1b used a different manipulation procedure, in which early-peak contours and late-peak contours were resynthesized from naturally produced medial-peak contours. Here, the medial-peak condition had the most natural sound quality. Conducting the experiment with both kinds of resynthesis procedures (Experiments 1a and 1b) allowed us to exclude potential artefacts that are present in each resynthesis procedure.

2.1. Methods

2.1.1. Participants

We tested 72 monolingual German speakers between 18 and 32 years; 36 of the participants were randomly assigned to Experiment 1a, 36 to Experiment 1b (Experiment 1a: 28 female, 8 male, average age = 23.3 years, SD = 3.1; Experiment 1b: 24 female, 12 male, average age = 22.7 years, SD = 3.5). They had no previous training in intonation and no reported history of hearing problems. Most of the participants grew up in Southern Germany (Baden-Wuerttemberg or Bavaria, 78% in total) and all were students or staff at the University of Konstanz. Participants received either course credit or a small reimbursement. The data of three additional participants of Experiment 1a was excluded because one reported a bilingual language background and two had participated in a related experiment. For Experiment 1b, the data of four additional participants was excluded because two reported a bilingual language background, one a hearing disorder, and one had participated in a related experiment.

2.1.2. Materials

We selected 36 trisyllabic German nouns with a WSW stress pattern (e.g., *Tornado* [tɔʁ.ˈnaː.do] 'tornado') as experimental items. This metrical structure allowed us to realize all three f0-alignment conditions within the word itself. To arrive at an equal distribution of words with stress on the first, second, and third syllable, we used 72 trisyllabic distractor words. Criteria for inclusion of the words were that they were monomorphemic nouns that only contained full vowels, which can be stressed (Féry, 1998; Wiese, 1996). All items were selected from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1993). Since for some items the lexical frequency was not attested in CELEX, we retrieved lexical frequencies from dlexDB (Heister et al., 2011). The frequency of experimental words was on average 143 occurrences per million (o.p.m., SD = 200), the frequency of distractor words was on average 239 o.p.m. (SD = 391).

A female native speaker of Standard German (aged 25), who was trained in intonational phonology, recorded the experimental and distractor items in isolation in a sound-attenuated cabin at the University of Konstanz (at 44.1 kHz, 16 Bit, mono). For Experiment 1a, the experimental items were split into two groups, with an equal number of open vs. closed vowels in the stressed syllable and with matched lexical frequency (average = 139 o.p.m., SD = 170 vs. average = 139 o.p.m.; SD = 191). One group of words was recorded with an early-peak contour (H+L* L-%), the other with a late-peak contour (L* H-^H%). As foreshadowed above, the medial-peak contours were PSOLA-resynthesized from these contours. The naturally recorded early- and late-peak contours were also adapted in terms of scaling of the f0 peak. For Experiment 1b, all 36 experimental items were recorded with a medial-peak contour (L+H* L-%). From these recordings, early-peak contours and late-peak contours were resynthesized. This time, the medial-peak contour was only adapted in f0 peak scaling. For both experiments, distractor words were also recorded with a medial-peak contour and resynthesized in the following way: Half of the SWW distractor words were PSOLA-resynthesized into late-peak contours (L* H-^H%) and half of the WWS distractor words were resynthesized into early-peak contours (H+L* L-%). For the medial-peak contours (L+H* L-%) in both distractor groups (SWW and WWS words), the f0 peak was adjusted in scaling to show the same peak height as the early-peak and the late-peak contours. PSOLA-resynthesis was performed as follows: We extracted the f0 value of the f0 peak within the stressed syllable of each recorded word (experimental and distractor words) and calculated the average f0 value of all f0 peaks (276 Hz). The same was done for low-toned unstressed syllables (177 Hz). Then, all f0 points were removed and f0 points in the centre of the vowels were added with the average f0 values of the peak or the unstressed syllables, respectively. This resulted in an average f0 excursion of 7.7 st. The resynthesized contours were hence closely modelled on naturally produced accent types; the late-peak contour was the mirror image of the early-peak contour (Fig. 1).[4]

To create experimental lists, the 36 experimental items (WSW, e.g., *Tornado*) were split into three groups, which were matched in average lexical frequency (group 1: 155 o.p.m., SD = 189; group 2: 129 o.p.m., SD = 217; group 3: 134 o.p.m., SD = 204). In each experimental list, one group of 12 WSW words was presented with an early-peak contour, the next group of 12 with a medial-peak contour and the final group of 12 with a late-peak contour, resulting in 36 experimental items in each list. Condition was thus distributed in a Latin-Square design, i.e., over the course of the whole experiment, all words were presented in all three intonation conditions and each participant heard all three intonation conditions, but each item in only one of the three conditions. All 72 distractors (36 SWW words, half with medial-peak contours and half with late-peak contours, and 36 WWS words, half with medial-peak contours and half with early-peak contours) were added to the lists. Thus, each list contained an equal number of WSW, SWW, and WWS words. In 48 cases the f0 peak coincided with the stressed syllable (12 experimental items, 36 distractors), while in 30 cases the f0 peak was aligned before the stressed syllable (12 experimental items, 18 distractors) and in 30 cases it was aligned after the stressed syllable (12 experimental items, 18 distractors). From these basic lists, nine experimental lists were created: All 108 items were pseudo-randomized, such that there were at least five items between two words with the same stress pattern and intonation condition, no more than two items with a medial-peak contour in a row, and no more than two words with the same onset consonant in a row. Also, the expected response button (response to syllable 1, 2, or 3) was randomized, such that no button had to be pressed more than twice in a row. Participants were randomly assigned to one of the nine experimental lists (four participants per list).

2.1.3. Procedure

Participants were seated in front of a computer screen (HP Compaq 8000 Elite CMT Business PC) and responses were recorded using a button box with three keys. The stimulus words were presented in isolation via headphones at a comfortable fixed loudness level (Beyerdynamic DT 990 Pro, 250 OHM). Participants were instructed to decide as correctly and fast as possible which syllable of the trisyllabic word they heard was the stressed one (German instruction: *betonte Silbe*). They were tested individually in a quiet room. Prior to the experiment, all participants filled in a language background questionnaire and received written instructions. Along with a general description of the task and the instruction, participants received three exemplar words in written representations in which the stressed syllable was capitalized: one SWW (*Furie* 'fury'), one WSW (*Kanone* 'canon'), and one WWS word (*Diamant* 'diamond'). They were told that for these three practice words, it would be correct to press button 1, 2, and 3, respectively. The experiment was programmed using *Presentation* (Neurobehavioral-Systems, 2001). Each trial started with a fixation cross that appeared in the centre of the screen for 250 ms. Then there was a blank screen of 500 ms after which the stimulus sound was played. The response box was active from the onset of the stimulus. After a participant's response, there was a 500 ms inter-trial interval before the next trial. The 500 ms inter-trial interval and the 750 ms silence at the start of each trial reduced interference from the preceding trial. To familiarize participants with the task and the speaker, the experiment started with six sonorant WSW nonce word trials with different f0-alignment contrasts. The whole experiment lasted approximately ten minutes. Throughout the experiment, there was no feedback and no time-out. Note that all studies in this paper were approved by the IRB of the University of Konstanz (IRB 30/2016).

2.2. Results

We only analysed the responses for experimental trials (WSW words). Responses were coded as correct when participants pressed button 2 and as incorrect otherwise. Fig. 2 shows the average correctness rates for the three intonation conditions in the two experiments. Correctness was statistically analysed using logistic mixed effects models (glmer) with *intonation condition* (medial-peak contour, early-peak contour, late-peak contour) and *Experiment* (1a, 1b, i.e., different manipulation procedures) as fixed factors and *participants* and *items* as crossed random effects (Baayen, Davidson, & Bates, 2008) using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2005) in R (version 3.3.3, R Development Core Team, 2015). Modelling was done in the following way: The initial model included the fixed factors (as main effects and as an interaction term), as well as random intercepts for *participants* and *items* (Bates, Kliegl, Vasishth, & Baayen, 2015; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Starting from this initial model, random slopes were planned to be included if they improved the fit of the model (in terms of LogLikelihood as indicated in model comparisons with the

[4] Note that the final boundary tone of the late-peak contour is perceptually a high-rise (German ToBI: H-^H%), as it often occurs in polar questions (Grice, Baumann, & Benzmüller, 2005) even though it looks like a high-plateau in Fig. 1.

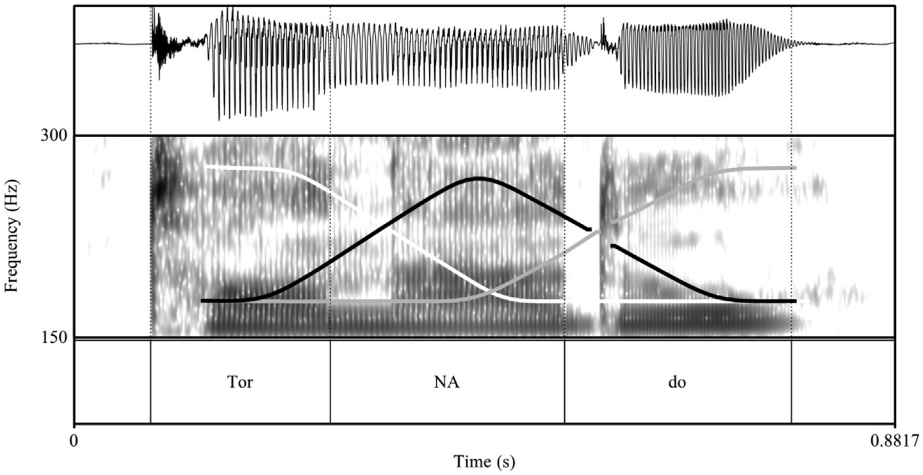


Fig. 1. Sound pressure wave, spectrogram and f0 contours for early- (white), medial- (black) and late-peak (grey) contours (PSOLA-resynthesized) in one experimental trial (e.g., *Tornado*), taken from Experiment 1a (capitals indicate lexical stress).

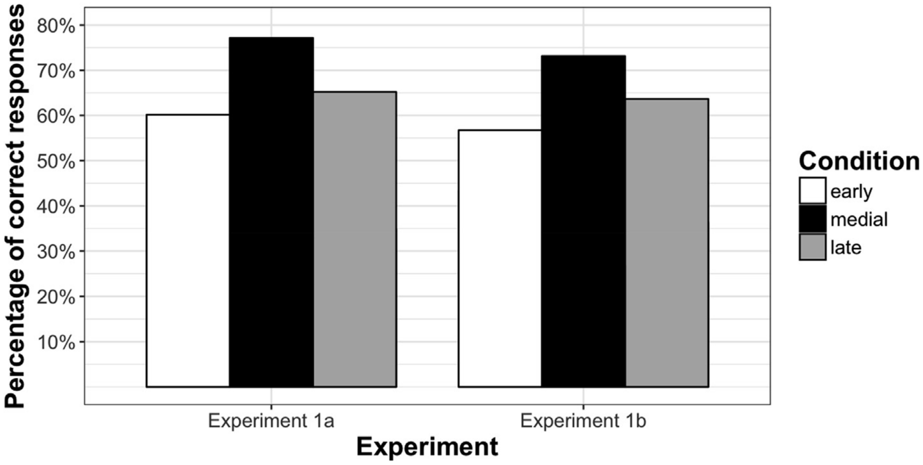


Fig. 2. Average correctness rates in the three intonation conditions (early-peak, medial-peak and late-peak condition) split by manipulation procedure (Experiments 1a vs. 1b).

anova() function). In the subsequent modelling steps, non-significant terms were planned to be consecutively excluded if the removal did not deteriorate the fit of the model (again indicated by the anova() function). As mentioned earlier, details on all analysis steps are provided on Mendeley https://doi.org/10.17632/2gkpwpg44j.3/.

The final model for correctness included *intonation condition* as fixed factor and no random slopes. In a combined analysis for Experiments 1a and 1b, the interaction between *intonation condition* and *Experiment* ($p = 0.77$) and the effect of *Experiment* ($p = 0.72$) was not significant. We therefore report a pooled analysis for the two experiments (the statistical results for the individual experiments are summarized in Appendix A.1). Table 1 shows the model output of the final model (Parts A and B) and its specification (Part C).

As predicted, the final glmer model showed a significant effect of *intonation condition* on correctness: In the medial-peak condition, participants gave on average 75% correct responses. There were significantly fewer correct responses in the early-peak condition (on average 59% correct responses, ß = −1.14, SE = 0.13, $z = -8.9$, $p < 0.0001$) and in the late-peak condition (on average 64% correct responses, ß = −0.74, SE = 0.13, $z = -5.8$, $p < 0.0001$). The difference in error rates between the early-peak and late-peak condition was also significant (ß = 0.39, SE = 0.12, $z = 3.2$, $p = 0.001$), see Table 1 for a summary of the results.

An analysis of the types of errors (i.e., syllable-1- or syllable-3-responses for the WSW word) revealed the expected response bias towards the syllable with the f0 peak (Fig. 3). Overall, participants gave 625 erroneous responses towards syllable 1 and 253 to syllable 3, respectively. For syllable-1-response errors, 49% occurred in the early-peak condition, i.e., the misalignment condition with the f0 peak on the first syllable (25% and 26% in the medial-peak and late-peak condition, respectively; $\chi^2 = 70.1$, df = 2, $p < 0.0001$). For syllable-3-response errors, 57% occurred in the late-peak condition, i.e., the misalignment condition with the f0 peak on the third syllable (20% and 23% in the medial-peak and the early-peak condition, respectively; $\chi^2 = 65.7$, df = 2, $p < 0.0001$). The statistical results for the individual experiments are summarized in Appendix A.2.

2.3. Discussion

The results of the forced-choice stress identification task show fewer correct stress judgements in words in which the

Table 1
Logistic mixed effects model reporting correctness rates in the three intonation conditions (with spelled out variable names): Estimate, Standard Error, *z*- and *p*-Values. Part A reports the model in which the medial-peak condition is represented in the intercept; Part B the model in which the early-peak condition is represented in the intercept. Part C gives the model specification of the final model (original variable names).

Part A.	Estimate	Std. Error	*z*-Value	*p*-Value
Intercept (medial-peak)	1.7517	0.2404	7.288	<0.0001
Intonation condition (early-peak)	−1.1382	0.1285	−8.855	<0.0001
Intonation condition (late-peak)	−0.7440	0.1278	−5.821	<0.0001
Part B.	Estimate	Std. Error	*z*-Value	*p*-Value
Intercept (early-peak)	0.6135	0.2346	2.615	0.0089
Intonation condition (late-peak)	0.3941	0.1219	3.234	0.0012
Intonation condition (medial-peak)	1.1382	0.1285	8.855	<0.0001
Part C. Model specification	glmer(corr ~ cond + (1|subject) + (1|item), data = combined, family = "binomial")			

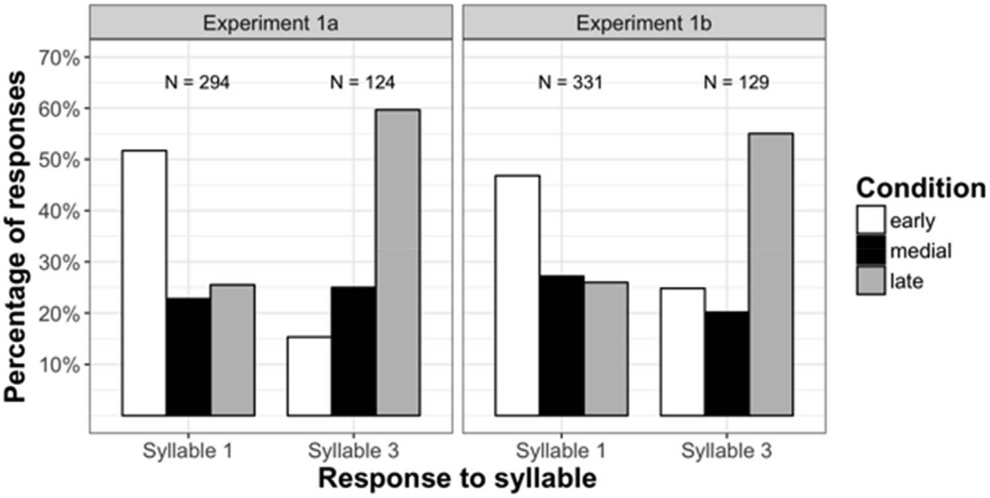


Fig. 3. Percentage of erroneous responses to syllable 1 and syllable 3 in the WSW targets for the three intonation conditions (early-peak, medial-peak and late-peak condition), split by Experiment.

f0 peak is not aligned with the stressed syllable, i.e., in the early-peak or late-peak condition, compared to the medial-peak condition. The erroneous responses are biased towards the syllable with the f0 peak. The error rates were not affected by the manipulation procedure, which leads us to the conclusion that the position of the f0 peak alone was responsible for the differences in error rates. Thus, our findings indicate that (a) participants have difficulties in detecting the stressed syllable in common German nouns when the f0 peak is misaligned with respect to the stressed syllable and (b) that German listeners misperceive a high-pitched but unstressed syllable as the stressed one, both when it precedes and when it follows the stressed syllable. Hence, it appears that peak alignment that is caused by different pitch accent types, i.e., manifestations of utterance-level phonology, influences the perception of lexical stress in German.

Extending future research, our results show that naturally occurring f0-alignment contrasts lead to a misperception of the underlying metrical structure of a word. Our target words were words with only one stress pattern, so we avoided direct lexical competition from an identical word with a different word-prosodic structure (as in the stress minimal pairs or nonce words used in previous studies, cf. Isačenko & Schädlich, 1966; Kohler, 2008; Niebuhr & Winkler, 2017). Thus, we can exclude the possibility that our findings are caused by the presence of direct stress competitors or strategic decisions. We conclude that high f0 is used as a cue to stress even when the perceived metrical strength pattern does not result in an existing German word. It is noticeable that the correctness rates in the alignment condition (medial-peak condition) did not result in a ceiling effect (i.e., 100% correct responses), which one could have expected with native speakers in the condition that may be considered the control condition. Instead, we only find about 75% correct responses in the medial-peak condition. It is conceivable that the rarity of stress minimal pairs in German leads to a reduced sensitivity for locating stress. Also, the lack of schwa syllables in the experimental materials has removed one dominant cue for unstressed syllables (Féry, 1998; Wagner, 2003). In any case, listeners did not just rely on stored (or rule-based) stress representations (for discussion see Levelt, Roelofs, & Meyer, 1999; Protopapas, Panagaki, Andrikopoulou, Gutiérrez Palma, & Arvaniti, 2016; Schwab & Dellwo, 2017). Otherwise, we would have seen few to no errors in stress identification and no effect of *intonation condition*, since listeners could have ignored the acoustic information on the position of the stressed syllable in the signal and relied on the stored representation only. By contrast, the signal clearly affected stress identifications, with a response bias towards the syllable with the f0 peak. Hence, the position of f0 peaks functioned as a cue to stress and guided listeners in their judgements more strongly than stored representations. On the other hand, the position of the f0 peak does not seem to be the main cue to stress for listeners; otherwise we would have seen more than 60% responses to the syllable with the f0 peak (see Fig. 3). What we observe is an interaction between f0 and the other stress cues (duration, intensity, and vocal effort).

In all experimental items in Experiments 1a and 1b, non-tonal phonetic stress cues (duration, intensity, vowel quality) and the phonological association suggest the penultimate syllable to be the stressed one. What differed across conditions was the alignment of the f0 peak with respect to the stressed syllable. These alignment differences represent phonological differences in accent types (and boundary tones), but do not alter the association between the accent and the stressed syllable itself. Listeners reacted to these f0-alignment differences and often judged the syllable that carried the f0 peak as metrically stressed – irrespective of the metrical status of the syllable. Hence, high f0 functions as a strong cue to stress for listeners. An alternative interpretation of this finding is that the stressed syllable stands out more in the medial-peak condition than in the peak-stress-misalignment contours because it has more tonal alternation in its vicinity (cf. Fig. 1). This would call for a replication with more tonal variation in all alignment conditions. The use of isolated words in the current experiment, however, made such complex contours impossible and we will leave this question to future research.

Within the two peak-stress-misalignment contours tested in the current experiment, we find differences in stress judgements. These misalignment contours represent the most extreme alignment positions of the f0 peak on the pre- or post-tonic syllable respectively, with the early- and late-peak contour being acoustically mirrored (Fig. 1). Admittedly, as mentioned earlier, the late-peak contour in the current experiment (L* H-^H%) is a combination of a pitch accent and a boundary tone in which the high boundary tone signals the f0 peak, while in the early-peak contour (H+L* L-%) the f0 peak is realized in form of an H-leading tone. We cannot exclude the possibility that f0 peaks with a different phonological function (H-^H% boundary tone vs. H-leading tone) are processed differently. Across misalignment conditions, our results show more correct responses in the late-peak contours as compared to the early-peak contours, which might reflect the processing consequences of this difference in phonological function. An alternative interpretation for the asymmetry in error rates across the misalignment conditions is that the late-peak condition was interpreted as a question contour. On single word utterances out of context, the question contour may be more frequent than the early-peak contour, which might have led to a processing advantage. In any case, what is important is that both peak-stress-misalignment conditions result in more errors than the alignment condition.

In sum, our findings indicate that f0 peaks on unstressed syllables (as leading tones in early-peak accents or as high boundary tones following L*-accents) lead to more errors in metrical stress perception and seem to change the interpretation of the position of metrical strength of the target words (from WSW to either SWW or WWS). The current offline task required a conscious meta-linguistic judgement of the position of the stressed syllable, which appeared hard for listeners, even though they could use their native lexical knowledge and are expected to be familiar with the concept of stress. Even though we explicitly asked for the position of the stressed syllable (German: *betonte Silbe*), we cannot fully exclude the possibility that the participants judged the acoustically most prominent syllable in the word. This is a drawback of the methodology used in Experiment 1. Experiment 2 used an online word recognition paradigm in which the task taps more closely into lexical activation (Tanenhaus, Magnuson, Dahan, & Chambers, 2000). In Experiment 2, we study whether alignment differences in two rising-falling contours (peak-stress alignment in medial-peak accents, L+H* L-% vs. peak-stress-misalignment in early-peak accents, H+L* L-%) affect the interpretation of the stressed syllable and consequently lexical activation in German.

3. Experiment 2: Visual-world eye-tracking

Experiment 2 uses the visual-world eye-tracking paradigm with four printed words on screen (McQueen & Viebahn, 2007; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to investigate whether an unstressed syllable with an f0 peak is temporarily interpreted as stressed. Similar to Reinisch et al. (2010) and other subsequent eye-tracking studies (cf. Connell et al., 2018; Jesse et al., 2017), the screen showed two written trisyllabic cohort competitors that differed in the position of stress (e.g., WSW, *Libelle* [li.ˈbɛ.lə] 'dragonfly' vs. SWW *Libero* [ˈli.bə.ʁo] 'sweeper'), together with two unrelated distractor words. We predict that the high-pitched unstressed initial syllable in WSW words with an early-peak accent (H+L*) leads to a temporary perception of the first syllable as stressed, thus activating the cohort member with initial stress (SWW word). For the segmentally ambiguous part of the targets (i.e., the target word onset up to the point at which the pairs phonemically diverge), we predict more fixations to the stress competitor in the early-peak condition than in the medial-peak condition.

3.1. Methods

3.1.1. Participants

Forty-eight German native speakers from the same pool of participants (39 female, 9 male, average age = 22.5 years, SD = 3.2 years, 28 right eye-dominant) took part for a small fee. They had normal or corrected-to-normal vision and unimpaired hearing. Most of the participants grew up in Southern Germany (75%). None of them had participated in Experiment 1. Data of four additional participants could not be used due to calibration difficulties.

3.1.2. Materials

Sixty-four trisyllabic cohort pairs were selected: One member of each pair was stressed on the first syllable (SWW, e.g., *Libero*), the other member on the second syllable (WSW, e.g., *Libelle*). The cohort pairs were segmentally identical up to at least the first consonant of the second syllable. Thirty-two of the 64 cohort pairs were used for cohort trials. Sixteen of the cohort trials were experimental trials (WSW word as the auditory target and SWW word as the stress competitor), 16 were distractor trials (SWW as the auditory target and WSW as the stress competitor), see Appendix B for a full list of cohort pairs in cohort trials. The remaining 32 cohort

Table 2
Acoustic realization means (and standard deviations) of WSW targets in two intonation conditions (naturally recorded, before PSOLA-resynthesis).

Acoustic variable	Naturally recorded medial-peak condition *(to be resynthesized to early-peak accents)*	Naturally recorded early-peak condition *(to be resynthesized to medial-peak accents)*
F0 excursion of accentual movement (st)	Rise: 8.36 (0.60)	Fall: 8.43 (0.67)
Duration first syllable (ms)	143 (34)	146 (34)
Duration second syllable (ms)	214 (48)	226 (48)
Intensity middle of first vowel (dB)	71.7 (2.3)	72.4 (3.3)
Intensity middle of second vowel (dB)	71.9 (2.5)	71.9 (1.8)
Mean RMS amplitude first vowel (Pa)	0.075 (0.017)	0.085 (0.026)
Mean RMS amplitude second vowel (Pa)	0.085 (0.019)	0.081 (0.013)
H1*-A3*[a] ratio middle first vowel (dB)	27.0 (10.8)	23.2 (10.7)
H1*-A3* ratio middle second vowel (dB)	30.8 (7.2)	31.8 (6.3)

[a] Following Mooshammer (2010), we used the H1*-A3* ratio as a measure for vocal effort, i.e., the difference between the amplitude of the first harmonic and the third formant (asterisks denote that amplitudes were corrected for formants). In order to perform these acoustic measurements, we adjusted a Praat script downloaded from http://www.seas.ucla.edu/spapl/voicesauce/, last access: 14/03/2018.

pairs were used as filler trials in which one of the unrelated items served as the auditory target.

Cohort members were matched for lexical frequency and number of characters across groups (Dahan & Gaskell, 2007; Lavidor, Ellis, Shillcock, & Bland, 2001; New, Ferrand, Pallier, & Brysbaert, 2006). In cohort trials, the cohort member with initial stress (SWW) had on average 6.5 characters (SD = 1.1) and a lexical frequency of 93 o.p.m. (SD = 99, according to dlexDB (Heister et al., 2011)), the cohort pair with stress on the second syllable had on average 6.9 characters (SD = 1.0) and a lexical frequency of 69 o.p.m. (SD = 119). In filler trials, the SWW cohort member had on average 7.0 characters (SD = 1.3) and a lexical frequency of 201 o.p.m. (SD = 479); the WSW cohort member had on average 7.0 characters (SD = 1.1) and a lexical frequency of 115 o.p.m. (SD = 283). For each cohort pair, we selected two distractors that were semantically and phonologically unrelated to the cohort members and which had a comparable length (on average 7.0 characters, SD = 0.9) and were similar in lexical frequency (on average 94 o.p.m., SD = 79). The trisyllabic distractors were stressed on the first, the second, or the third syllable to increase variability of stress patterns on screen (20 SWW, 20 WSW, and 24 WWS words in total).

The same female speaker as in Experiment 1 recorded the auditory stimuli, i.e., the instructions for the eye-tracking study (*Bitte klicke <TARGET> an*, 'Please click on <*TARGET*>'), in the same cabin and under the same conditions. The instructions for trials referring to one of the cohort members were produced in two intonation conditions each: with a medial-peak accent (L+H*) and an early-peak accent (H+L*) on the auditory target word; all other words in the instruction were unaccented. The final boundary tone was always low (L-%). The sentences were re-recorded until we could select pairs that did not differ other than in relevant acoustic properties (see Table 2 for acoustic analysis of targets used in experimental trials). For fillers, half of the sentences were recorded with a medial-peak, half with an early-peak accent on the target, matching the f0 range of their accentual movement with the f0 range of cohort pairs.

To avoid effects of distal prosodic context (Brown, Salverda, Dilley, & Tanenhaus, 2015), all trials were spliced after the verb (*Bitte klicke* || <*TARGET*> *an*), with || indicating the splicing point. For experimental trials (WSW word as the auditory target), four different versions of the pre-context *Bitte klicke* were used to avoid a mismatch between coarticulatory information in the last syllable of the verb and the onset consonant of the target word (one for WSW words starting in [a], [e], [m], respectively, and one for words starting in any other consonant). For each target, the pre-context was the same across intonation condition. On average, the pre-contexts were 575 ms (SD = 25 ms) long. Overall, the cross-spliced stimuli sounded natural and the splicing was not noticeable (as judged by four members of the linguistics department who were presented with ten randomly selected experimental trials, five in each condition).

After splicing, the stimuli were PSOLA-resynthesized by superimposing the contour of a target word in one intonation condition to the target that was originally recorded in the other intonation condition. Thus, the f0 contours of medial-peak accents (L+H*) were superimposed on target words originally recorded with an early-peak accent (H+L*) and vice versa. Fig. 4 shows the PSOLA-resynthesized version of one item in the two intonation conditions.

PSOLA-resynthesis for experimental and distractor trials was done as follows: We extracted the scaling and proportional alignment of low (L) and high turning points (H) in the target word and transplanted these values on the recording in the other condition.[5] For filler trials, we used the same resynthesis procedure as in Experiment 1 as there were no direct lexical competitors. That is, we calculated the average f0 maximum (314 Hz) and the average f0 minimum in unaccented syllables in the target words (193 Hz) in the original recordings of the filler items in both intonation conditions, resulting in an average f0 range of the accentual movement of 8.4 st.

3.1.3. Procedure

The experiment consisted of 64 trials: 32 cohort trials (16 experimental, 16 distractor trials) and 32 filler trials. Experimental trials are of interest for the current hypothesis. Distractor and filler trials served a strategic function only, protecting against an imbalance in clicking responses. That way, participants had to click equally often on cohort words and non-cohort words throughout the experiment and equally often on words with stress on the first and second syllable. Intonation condition was rotated across trials as follows: For the

[5] Note that the resynthesis procedure for cohort trials was slightly different from the one used in Experiment 1. Piloting showed that the proportional interchanging of intonation contours between the two conditions (as performed for cohort trials in Experiment 2) resulted in the best stimulus quality. Yet, this procedure could only be used with two intonation conditions under consideration, not three (as was the case in Experiment 1).

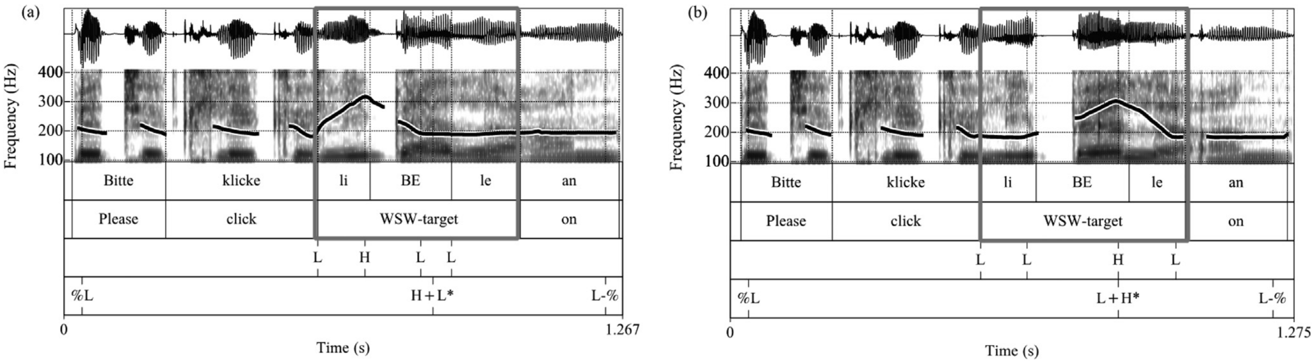


Fig. 4. Sound pressure wave, spectrogram and f0 contours for two exemplar experimental trials (a) early-peak condition, (b) medial-peak condition (both PSOLA-resynthesized). The tiers show the words in German (1) and English (2), the tonal targets that were calculated from the original recordings (3), and the GToBI annotation (4).

experimental and distractor trials, intonation condition was distributed in a Latin-Square design. Half of the filler trials were presented with an early-peak accent, half with a medial-peak accent. Each participant was presented with the same fillers. Eight experimental lists were created, pseudo-randomizing the order of trials, such that each experimental half contained the same number of cohort and distractor trials, with the constraint of the experimental item (WSW) being at most the third item of the same intonation condition in a row. Each list started with seven practice trials: five filler trials, followed by two distractor trials. Participants were randomly assigned to one of eight experimental lists (six participants per list).

Participants were tested individually using the SR Eyelink 1000 Plus in a desktop mount system at a sampling rate of 500 Hz. The distance between participants and a LCD screen (37.5 cm × 30 cm) was approximately 70 cm. Prior to calibration, all participants received written instructions. Their dominant eye was then calibrated in an automatic procedure (pupil and corneal reflection, Eyelink default settings). Every trial of the experiment started with a centred black cross on white background displayed until participants clicked on it. Upon clicking, the four words appeared on screen (Times New Roman Font, size 20). The words were presented in the outer third of the four quadrants of the screen (to avoid peripheral looking) and framed by a rectangular box (6.5 cm × 4 cm), see Fig. 5. The position of the items on screen was counterbalanced across intonation conditions, such that the target that participants had to click on occurred equally often in the four possible positions for each intonation condition. The auditory instruction started 2000 ms after the words occurred on screen, leaving a preview of the words for participants of approximately 2575 ms (2000 ms pause+on average 575 ms pre-context). Participants' task was to click on the word named in the auditory stimulus as fast as possible. Auditory stimuli were presented via headphones at fixed comfortable loudness (Beyerdynamic DT-990 Pro, 250 OHM). Every fifth trial, a drift correction was initiated. After half of the trials, there was an optional pause. The total duration of the experiment was 15 minutes (including a language background questionnaire).

3.2. Analysis and results

Only experimental trials were analysed (WSW words, e.g., *Libelle*). Participants correctly clicked on the auditory target in 97.1% of the cases (13 mistakes in the early-peak condition; 9 in the medial-peak condition). Fixations were extracted in 20 ms-bins. They were automatically coded as being directed to the target (WSW, *Libelle*), the stress competitor (SWW, *Libero*), or to the unrelated distractors if they fell within a square of 200 × 200 pixels around the respective word. Blinks and saccades were not further processed. The grand average of evolution of fixations to the four words in the two intonation conditions is shown in Fig. 6 (using the VWPre package in R, Porretta, Kyröläinen, van Rij, & Järvikivi, 2018). The grey vertical dashed lines indicate the segmental reference points, i.e., word boundaries from left to right, shifted by 200 ms – the time it takes to launch a saccade (Altmann & Kamide, 2004; Fischer, 1992; Matin, Shao, & Boff, 1993). Hence, only after this time fixations can be interpreted as a response to the signal. The segmental uniqueness point (U.P.), i.e., the point in the signal at which acoustic information perceptually distinguishes the cohort pair irrespective of suprasegmental information (e.g., after the release of [b] in *Libelle* [li.ˈbɛ.lə] vs. *Libero* [ˈli.bə.ʁo]) is indicated by a black dashed line (at 1077 ms in Fig. 6). The window of interest for the current study is the time from target word onset until the segmental U.P., both shifted by 200 ms, i.e., from 775 to 1077 ms in Fig. 6. This is the shifted time of segmental overlap in which effects of f0 on stress interpretation are expected to surface.

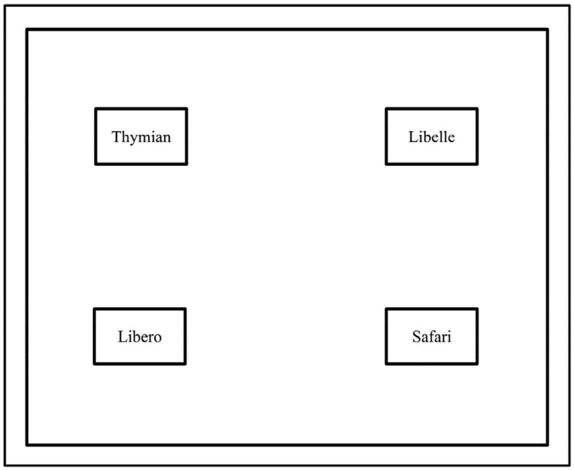


Fig. 5. Exemplar screen for an experimental trial, showing the cohort pair (WSW target *Libelle* (top right) and SWW stress competitor *Libero* (bottom left)), and two unrelated distractor items (SWW *Thymian* 'thyme' and WSW *Safari* 'safari', (top left and bottom right, respectively)). The screen is depicted to scale.

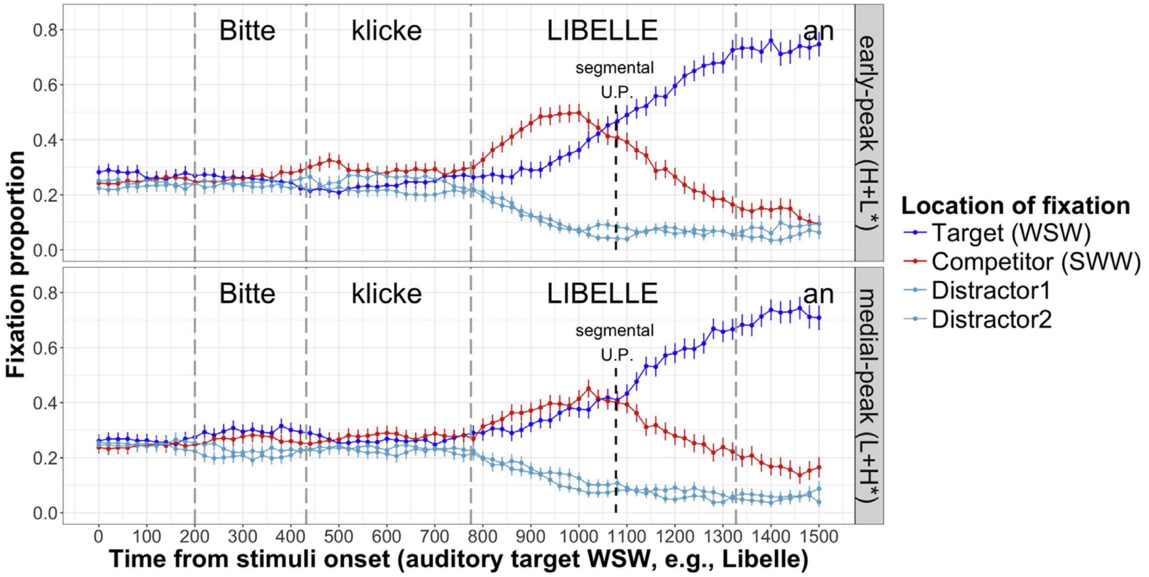


Fig. 6. Evolution of fixations to target (WSW, e.g., *Libelle*, dark blue line), competitor (SWW, e.g., *Libero,* red line) and the two distractors (e.g., *Thymian* 'thyme' (SWW), *Safari* 'safari' (WSW), light blue lines) in experimental trials in the two intonation conditions (early-peak condition upper panel, medial-peak condition lower panel). Acoustical landmarks (grey dashed vertical lines) are shifted by 200 ms.

Fig. 6 shows that as soon as segmental information of the auditory WSW target became available, fixations to the (segmentally unrelated) distractors decreased in both intonation conditions (at 775 ms in Fig. 6), while fixations to the WSW target and the SWW stress competitor both further increased, but with clear differences across intonation conditions: In the early-peak condition, competitor fixations increased more quickly than target fixations; in the medial-peak condition, target and competitor fixations increased with an equal slope. In particular, during the segmentally ambiguous part (775–1077 ms in Fig. 6, i.e., [lib]), the stress competitor was fixated more than the target in the early-peak condition (Fig. 6, upper panel), while target and competitor were fixated almost equally in the medial-peak condition (lower panel). After the segmental U.P. (at 1077 ms in Fig. 6), fixations to the SWW competitor dropped (for both early- and medial-peak condition).

We predicted more competitor fixations in the early-peak than in the medial-peak condition in the time window from 775 to 1077 ms. A visual inspection of the competitor fixations shows that this is indeed the case (Fig. 7).

To statistically corroborate the differences in competitor fixations in the two intonation conditions, we used general additive mixed modelling in R (GAMMs, Baayen, van Rij, de Cat, & Wood, 2018; Baayen, Vasishth, Kliegl, & Bates, 2017; Wieling, 2018; Wood, 2006, 2017). GAMMs represent a state-of-the-art statistical approach to analysing time-varying data with non-linear relationships and autocorrelation (Baayen et al., 2018; Wieling, 2018). The visual representation of GAMMs indicates when in time an effect on a response variable becomes significant. This is an elegant alternative to traditional time-window analyses, which require fixations to be binned in predefined arbitrary analyses windows (Barr, 2008), or to Growth Curve Analysis (Mirman, Dixon, & Magnuson, 2008), which models differences in shape of the curves by fitting polynomials to the time-series data. GAMMs have successfully been used in other eye-tracking studies (e.g., Nixon, van Rij, Mok, Baayen, & Chen, 2016; Porretta, Tucker, & Järvikivi, 2016; van Rij, Hollebrandse, & Hendriks, 2016). Specifically, GAMMs model non-linear dependencies of a response variable and a predictor via smooth functions, which include a pre-specified number of base functions of different shapes, e.g., linear and parabolic functions of different complexity. Fixed effects can be modelled in the same way as in more traditional linear mixed effect regression models (see analyses in Experiment 1, Baayen et al., 2008). For the GAMM analysis, we used the R package *mgcv* (Wood, 2011, 2017); the package *itsadug* (van Rij, Wieling, Baayen, & van Rijn, 2016) was used to plot the model results. Note that GAMM model outputs alone are not sufficient for the interpretation of the results, effects only become obvious through visualization (Wieling, 2018; Wood, 2006, 2017).

In our general additive mixed models, we used the following variables: Competitor fixations were taken as the response variable. They were converted to empirical logits (*elogs*, which is a logit transformed proportion to looks to the competitor, i.e., a ratio of the fixations to the competitor divided by the fixations directed to the three other objects (Barr, 2008)). We included a parametric coefficient for *intonation condition*, along with a random effect for *event* (combining *item* and *subject* as a unique identifier) which allowed for a random intercept (see e.g., Porretta et al., 2016). For *intonation condition*, nonlinear functional relations with the response variable over time were allowed for using the smooth function. In addition, an AR-1 correlation parameter was estimated, using the acf_resid()function implemented in the package *itsadug* in order to account for the autocorrelation in the fixation time series.

Following Porretta et al. (2016), we used a backward step-wise elimination procedure to identify the best model. In all our models, *intonation condition* was kept as a parametric coefficient due to the experimental design of the study. As a first criterion for inclusion of smooth terms, we used the estimated *p*-value of these smooth functions (see Porretta et al.,

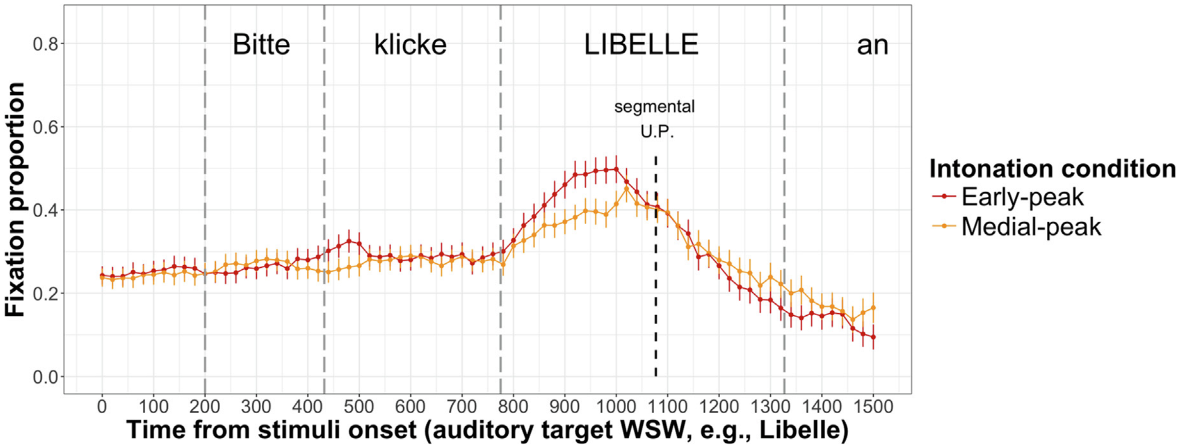


Fig. 7. Competitor fixation across intonation condition (early-peak, red vs. medial-peak condition, orange) for Experiment 2. Acoustical landmarks (grey dashed vertical lines) are shifted by 200 ms.

2016). A smooth term was considered for inclusion in the model only if it showed a significant *p*-value (<0.05). We then compared the model including the smooth term to a simpler version of the model (without the smooth term) using the function CompareML(). This comparison revealed whether the inclusion of this predictor significantly improved the model fit (Maximum Likelihood (ML) scores) or whether this predictor had to be removed. It has been argued that the Akaike information criterion (AIC, Akaike, 1974) is not reliable when a model accounts for autocorrelation in the data (Zuur, Ieno, Walker, Saveliev, & Smith, 2009). Therefore, we compared ML scores as a criterion to determine the better model fit (see Porretta et al., 2016).

Model comparisons following the procedure described above indicated that the model with different time-smooth functions for the different intonation conditions was preferred over the model without a smooth function for the factor *intonation condition* over time. Hence, the final model included *intonation condition* as a parametric coefficient and as a non-linear effect (smooth term) over time, as well as random intercepts for the *event* variable. The final model accounted for 67.7% of the deviance, emphasizing that the model captured important features of the fixations over time, see Table 3 for coefficients of the final model. Part A in Table 3 shows the parametric coefficients (comparable to fixed factors in the output of linear mixed regression models, lmers). The early-peak contour is represented in the intercept (first row), the medial-peak contour (second row) shows the adjustments in coefficients relative to the intercept. The *p*-value 0.1040 for the parametric coefficient indicates that irrespective of time, i.e., for the whole analysis window, the effect of *intonation condition* is not significant. That is, intonation does not have a global effect over the entire analysis window, 775–1077 ms, but only a localized effect in the time window 868–1001 ms, i.e., 44% of the time window, as calculated by the model, see Table 3 for model specification of the final GAMM. This localized effect of *intonation condition* on competitor fixations is captured by the smooth terms and the significant differences are shown in Fig. 8. Recall that the output of GAMM is only meaningful when visualized.

As predicted, Fig. 8 shows that in the time period in which information of the segmentally ambiguous part is processed, there are significantly more fixations to the stress competitor when the target is realized with an early-peak compared to a medial-peak accent. The time window of significant differences (868–1001 ms) lies in the time window in which segmental information is ambiguous between the WSW and SWW word (775–1077 ms).

For the sake of completeness, we also analysed participants' fixations to the target, for which we find the reversed pattern, i.e., fewer fixations to the target in the early-peak condition compared to the medial-peak condition, 840–920 ms (see Zahner, submitted).

3.3. Discussion

The fixation data show that the f0-alignment contrast in naturally existing pitch accent types affects lexical activation: While participants were processing the segmentally ambiguous part of the target word, there were more fixations to the stress competitor (SWW) when the WSW target word was presented with an early-peak accent than when presented with a medial-peak accent (868–1001 ms). This finding shows that H-leading tones, i.e., high-pitched unstressed syllables, temporarily activate competitor words with initial stress in online speech processing.

Importantly, both intonation conditions, the medial-peak accent (L+H*) and the early-peak accent (H+L*), naturally occur in German. We argue that they are both pragmatically appropriate renditions for the carrier phrase used in the current experiment. On the one hand, the objects to be clicked on in each trial represent information-structurally new information, which might favour a medial-peak accent (Kohler, 1991). On the other hand, the repetition of the pre-context creates a notion of accessibility of the objects as a whole (cf. Baumann & Grice, 2006), which makes an early-peak realization also pragmatically appropriate. To support the felicity of the two intonation conditions, we conducted a post-hoc production study: Another ten participants (6 female, 4 male, average age = 25.8 years, SD = 3.9 years) read the experimental stimuli of the eye-tracking experiment aloud (two of the experimental lists with 64 trials each, N = 640 productions). An intonational analysis showed that medial-peak accents were most frequent (71% of the cases); early-peak accents occurred in 19% and late-peak contours in 5%, the remaining cases

Table 3
Final general additive mixed model of Experiment 2 with spelled out variable names. Part A: Estimate, Standard Error, *t*- and *p*-Values for the parametric coefficients. Part B: Estimated degrees of freedom (EDF), reference degrees of freedom (Ref.df), *F*- and *p*-Values for the smooth term. Part C: Model specification of the final model (original variable names).

Part A. Parametric coefficients	Estimate	Std. Error	*t*-Value	*p*-Value
Intercept (early-peak contour)	−0.4035	0.1012	−3.988	<0.0001
Intonation condition (medial-peak)	−0.2285	0.1406	−1.626	0.1040
Part B. Smooth terms	**EDF**	**Ref.df**	***F*-Value**	***p*-Value**
Random effect (intercept) for Event, s(event)	643.277	730.000	8.655	<0.0001
Time, by condition, s(Time):early-peak	5.416	6.718	30.075	<0.0001
Time, by condition, s(Time):medial-peak	1.688	2.133	44.491	<0.0001
Part C. Model specification	bam(IA_2_ELogit ~ cond + s(TIMESTAMP, by = cond) + s(event, bs = 're'), rho = 0.64, AR. start = df_combined[df_combined$experiment == "Exp2",]$start_event, data = df_combined[df_combined $experiment == "Exp2",], discrete = FALSE, nthreads = 4, method = "ML")			

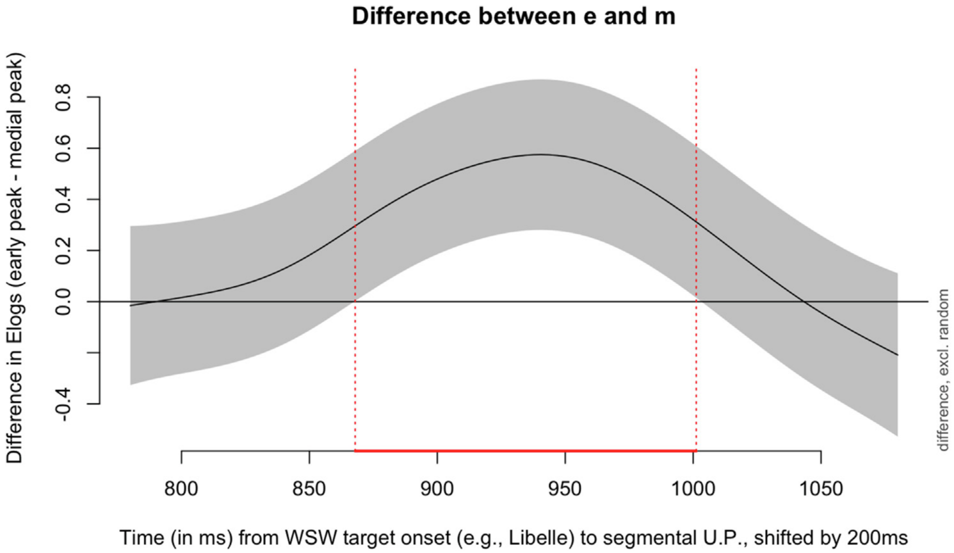


Fig. 8. Difference curve in competitor fixations in early-peak condition minus medial-peak condition in Experiment 2. The grey band indicates the 95% confidence interval (CI) of the mean difference. Values above zero indicate more competitor fixations in the early-peak condition. Conversely, values below zero indicate more competitor fixations in the medial-peak condition. The difference is significant if the 95% CI does not include zero (868–1001 ms). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

were unaccented. This distribution of pitch accents is similar to the frequency counts of these accents in German appointment scheduling dialogues (Peters, Kohler, & Wesener, 2005). Hence, early-peak accents are clearly an option in this setting, albeit not the preferred one.

In Experiment 2, the effect of pitch accent type (i.e., differences in f0 alignment) surfaced immediately, while participants were processing the segmentally ambiguous part, i.e., the first syllable of the target words (which either had higher or lower pitch than the preceding syllable in the pre-context) and the onset consonant of the second syllable. The time window in which the stress competitor effect appeared lasts 133 ms (868–1001 ms), 44% of the time window during which the effect can occur (133 ms of 302 ms). Effects similar in duration to the one we find in our data have been reported in visual-world eye-tracking studies with comparable or even larger acoustic differences (e.g., Jesse et al., 2017; McQueen & Viebahn, 2007).

The competitor activation effect is clearly driven by f0 peak-alignment information. In both intonation conditions the pitch accent was phonologically associated with the stressed syllable of the WSW target; non-tonal stress cues pointed towards the second syllable as the stressed one. The only cue that differed across intonation conditions was the alignment of the f0 peak. Since we used resynthesized materials, we can safely conclude that effects are solely due to the f0 alignment and not caused by other acoustic correlates that cohere with the f0 peak (Niebuhr, 2007). If the natural productions of a medial-peak accent have additional intensity and duration on the stressed syllable and the natural productions of an early-peak accent on the pre-stressed syllable, then the resynthesis reverted these cues. Hence, we assume that with natural early-peak and medial-peak productions the stress competitor effect may be even stronger (cf. Zahner, Kember, & Braun, 2017, for Australian English).

In Experiment 3, we test the role of the distribution frequency of different pitch accent types. As outlined in the Introduction, if the stress competitor effect is driven by the input frequency of high-pitched stressed syllables in natural speech (Option 2), we expect no difference in competitor activation across intonation condition. If, on the other hand, the stress competitor effect is caused by other factors (e.g., inherent salience of high f0, Option 1), we expect to see the same competitor activation differences as in Experiment 2, i.e., more looks to the stress competitor (SWW word) when the WSW target is realized with an early-peak accent compared to a medial-peak accent.

4. Experiment 3: Visual-world eye-tracking with exposure phase

We added a 3-minute exposure phase to the current eye-tracking experiment by which we increased the occurrence frequency of low-pitched stressed syllables in the immediate input. We used a similar design as in accent-adaptation studies (cf. Grohe & Weber, 2016a, 2016b; Reinisch & Weber, 2012; Witteman, Weber, & McQueen, 2014). Studies on adaptation on the lexical level and perceptual-learning studies have shown that a 3-minute exposure is sufficient to affect lexical activation (Grohe & Weber, 2016a, 2016b; Kraljic & Samuel, 2006; Norris, McQueen, & Cutler, 2003). Reinisch and Weber (2012) similarly showed for the suprasegmental level, that listeners quickly adapt to lexical stress placement errors by non-native speakers (by listening to a 2.3-minute story, 28 critical items) and use this information in word recognition.

4.1. Methods

4.1.1. Participants

Another group of 48 German native speakers (32 female, 16 male; average age = 22.2 years, SD = 3.2 years, 34 right-eye dominant) with normal or corrected-to-normal vision and unimpaired hearing participated under the same conditions as in Experiment 2. None of them had participated in Experiments 1 or 2. Again, most of the participants grew up in Southern Germany (73%). Data of five additional participants was excluded due to calibration errors (4) and a bilingual background (1).

4.1.2. Materials

For the exposure phase, we chose 15 alternative-question units and 15 contrastive-topic units, since alternative questions and contrastive topics are commonly realized with low-pitched stressed syllables (Bartels, 1999; Braun, 2006; Büring, 1997; Truckenbrodt, 2011). The alternative-question units consisted of an alternative question and a one-word answer (e.g., ‘Do you prefer Bolognese$_{L*+H}$ or Carbonara$_{H+L*\ L-\%}$? – Carbonara$_{H+L*\ L-\%}$’). The contrastive-topic units consisted of a declarative, a polar question, and another declarative (e.g., ‘The blanket$_{L*+H}$ is made of flannel$_{H+L*\ L-\%.}$ – What is the pillow$_{L*}$ made of$_{H-\hat{}H\%}$? – The pillow$_{L*+H}$ is made of feathers$_{H+L*\ L-\%}$.’). The declaratives followed a theme-rheme structure, in which the theme referred to a topic and the rheme formulated a proposition about the theme (see Braun, 2006, for similar materials). The polar question asked for an alternative theme that formed a contrast to the one given in the first declarative (e.g., ‘blanket’ – ‘pillow’). The second declarative was structurally identical to the first one and provided the answer to the polar question.

The same female speaker as in the previous experiments recorded the exposure stimuli with L*+H, H+L*, and L*-accents. In total, the exposure materials consisted of 120 low-pitched accented syllables. The individual components of each unit (alternative question and a one-word answer for alternative-question units; declarative, polar question, and second declarative for contrastive-topic units) were concatenated with an inter-trial interval of 400 ms to form one unit. The inter-trial interval for the first five trials was 600 ms to acquaint participants with the task. The 30 units were grouped in six blocks à five trials. The order of trials in a block was pseudo-randomized and trials were separated by an inter-trial interval of 1000 ms (first block had a 200 ms longer inter-trial interval to adjust participants to the task). The blocks of stimuli were on average 26.6 sec (SD = 1.7 s) long.

4.1.3. Procedure

During the exposure phase, participants listened to one block of stimuli at a time while fixating a black cross that was centred

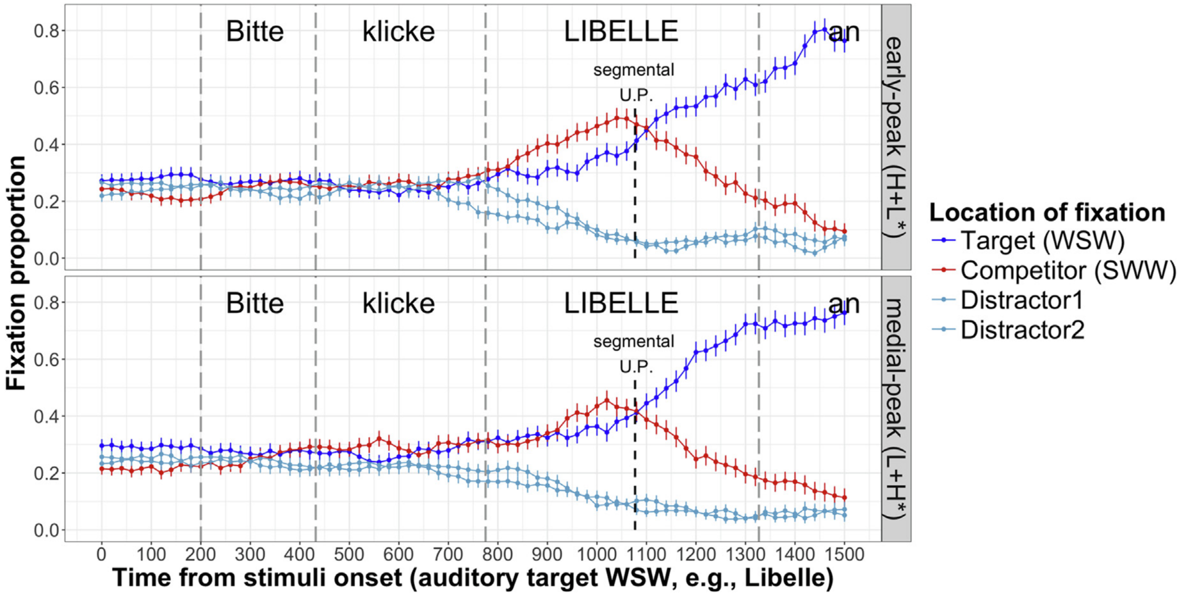


Fig. 9. Evolution of fixations to target (WSW, e.g., *Libelle,* dark blue line), competitor (SWW, e.g., *Libero,* red line) and the two distractors (e.g., *Thymian* ‘thyme’ (SWW), *Safari* ‘safari’ (WSW), light blue lines) in experimental trials in the two intonation conditions (early-peak condition upper panel, medial-peak condition lower panel) for Experiment 3. Acoustical landmarks (grey dashed vertical lines) are shifted by 200 ms.

on a white screen. After every block, participants were asked to rate on a Likert scale from 1 to 5 whether a given adjective applied to the person they were listening to (e.g., *sympathisch?* ‘likeable’). The adjectives were taken from Schweitzer and Lewandowski (2013). In total, there were six ratings, i.e., one rating after each block of stimuli. Participants’ responses were not recorded. The test phase immediately followed the exposure phase with the exact same procedure, stimuli, and lists as in Experiment 2. Calibration took place before the exposure phase to have a smooth transition between the two phases.

4.2. Results

Fig. 9 displays the grand average of the evolution of fixations in experimental trials in Experiment 3 to the four words on screen in the two intonation conditions.

As in Experiment 2, we tested whether competitor fixations differed as a function of intonation condition during the segmentally ambiguous part of the target (again shifted by 200 ms). Data analysis and statistical modelling (inclusion of variables and model comparisons) were the same as in Experiment 2. The final general additive mixed model included *intonation condition* as parametric coefficient and a non-linear effect (smooth term) of *intonation condition* over time, as well as a random intercept for the *event* variable. The model accounted for 71.8% of the deviance, see Table 4 for the coefficients of the final model. The effect of *intonation condition* over time is directly visualized in Fig. 10.

Different from what was observed for Experiment 2, competitor fixations in the two intonation conditions did not significantly differ during the segmentally ambiguous part, i.e., the 95% CI includes zero almost throughout the whole window of interest. Note though that at the end of the analysis window (1077–1080 ms), the model indicated that competitor fixations were more frequent in the early-peak condition; an effect we consider to be negligible due its very short-term appearance (only 3 ms).

In a second analysis step, we combined the data of Experiments 2 and 3 to examine whether the difference of competitor fixation across intonation conditions differs for the two experiments, i.e., whether there is an interaction between *intonation condition* and *experiment* (see analysis steps in Wieling, 2018,

Table 4
Final general additive mixed model of Experiment 3. Part A: Estimate, Standard Error, *t*- and *p*-Values for the parametric coefficients. Part B: Estimated degrees of freedom (EDF), reference degrees of freedom (Ref.df), *F*- and *p*-Values for the smooth term. Part C: Model specification of the final model (original variable names).

Part A. Parametric coefficients	Estimate	Std. Error	*t*-Value	*p*-Value
Intercept (early-peak contour)	−0.4838	0.1048	−4.617	<0.0001
Intonation condition (medial-peak)	−0.0976	0.1466	−0.666	0.5060
Part B. Smooth terms	**EDF**	**Ref.df**	***F*-Value**	***p*-Value**
Random effect (intercept) for Event, s(event)	642.525	711.000	10.73	<0.0001
Time, by condition, s(Time): early-peak	1.005	1.011	159.68	<0.0001
Time, by condition, s(Time): medial-peak	3.206	4.114	26.13	<0.0001
Part C. Model specification	bam(IA_2_ELogit ~ cond + s(TIMESTAMP, by = cond) + s(event, bs = ‘re’), rho = 0.62, AR. start = df_combined[df_combined$experiment == “Exp3”,]$start_event, data = df_combined[df_combined $experiment == “Exp3”,], discrete = FALSE, nthreads = 4, method = “ML”)			

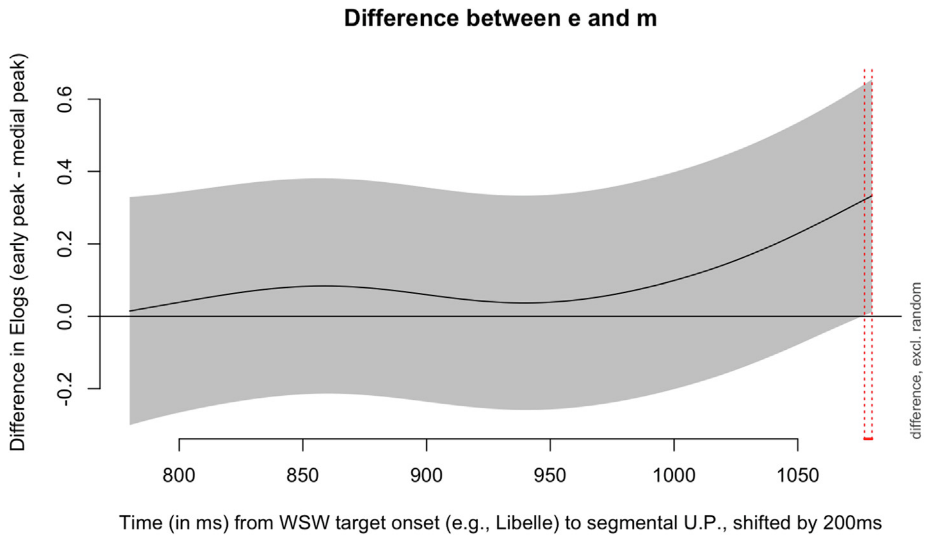


Fig. 10. Difference curve in competitor fixations in early-peak condition minus medial-peak condition in Experiment 3. The grey band indicates the 95% CI of the mean of the difference. Values above zero indicate more competitor fixations in the early-peak condition. Conversely, values below zero indicate more competitor fixations in the medial-peak condition. The difference is significant if the 95% CI does not included zero (1077–1080 ms). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Final general additive mixed model of combined analysis for Experiments 2 and 3. Part A: Estimate, Standard Error, *t*- and *p*-Values for the parametric coefficients. Part B: Estimated degrees of freedom (EDF), reference degrees of freedom (Ref.df), *F*- and *p*-Values for the smooth term. Part C: Model specification of the final model (original variable names).

Part A. Parametric coefficients	Estimate	Std. Error	*t*-Value	*p*-Value
Intercept (Experiment 2, without exposure)	−0.6321	0.1005	−6.288	<0.0001
Experiment: Experiment 3 (with exposure)	0.0468	0.1433	0.326	0.7440
Part B. Smooth terms	EDF	Ref.df	*F*-Value	*p*-Value
Random effect for Event (intercept), s(event)	1285.671	1441.000	9.524	<0.0001
s(Time): Exp2	2.383	3.043	32.945	<0.0001
s(Time): Exp3	4.029	5.136	19.763	<0.0001
s(Time): IsEarly	2.001	2.001	2.826	0.0593
s(Time): IsExp2Early	6.307	7.580	8.937	<0.0001
Part C. Model specification	bam(IA_2_ELogit ~ experiment + s(TIMESTAMP, by = experiment) + s(TIMESTAMP, by = IsEarly) + s (TIMESTAMP, by = IsExp2Early) + s(event, bs = 're'), data = df_combined, rho = 0.63, AR. start = df_combined$start_event)			

p. 106ff). We first created a new variable (*Expcond*), which is the interaction of *intonation condition* and *Experiment*, resulting in four levels (Exp2-Early-peak; Exp2-Medial-peak; Exp3-Early-peak, Exp3-Medial-peak). This interaction variable was used instead of the two-dimensional *intonation condition* variable in the model. Similar to the model fitting for individual experiments, for the model of the combined data, we included a parametric coefficient for the interacting variable (*Expcond* with four different levels), along with a random intercept for *event*. Further, for the interacting factor, nonlinear functional relations with the response variable over time were included using the smooth function. Again, an AR-1 correlation parameter was estimated to account for the autocorrelation in the fixation data. Then, we checked whether the model with the interacting factor (*Expcond*) is better than the model with a smooth term for *intonation condition* only. This was indeed the case; the model including the interacting factor had a significantly lower ML score than the model with *intonation condition* (22407.07 compared to 22432.13, $p < 0.0001$). Hence, it was necessary to distinguish the competitor activation between experiments. To formally assess whether the difference in competitor fixations in the early- vs. the medial-peak condition is different between the two experiments (Experiment 1 vs. Experiment 2), the model was re-specified by implementing binary difference smooths, following the description in Wieling (2018, p. 109ff). Specifically, along with *experiment* as a parametric factor, we included a binary difference smooth distinguishing between the early- and medial-peak conditions irrespective of experiment (*IsEarly*) as well as a binary difference smooth that distinguished early-peak accents in Experiment 2 from all other conditions (*IsExp2Early*), see Table 5 for coefficients of the final model. To illustrate, s(Time, by = IsExp2Early) represents the difference between the early-peak-vs.-medial-peak contrast in Experiment 2 vs. that in Experiment 3, while s(Time, by = IsEarly) represents the difference between the early-peak-vs.-medial-peak contrast for Experiment 2. The difference of the difference in competitor fixations across experiments is significant, s(Time, by = IsExp2Early), see Wieling (2018). This difference is directly displayed in Fig. 11, which shows that the difference in competitor fixations for the early-peak-vs.-medial-peak contrast is larger in Experiment 2 than in Experiment 3. For a short time window (~920–950 ms), zero is not included in the 95% CI of the difference curve of the difference, suggesting a significant interaction.

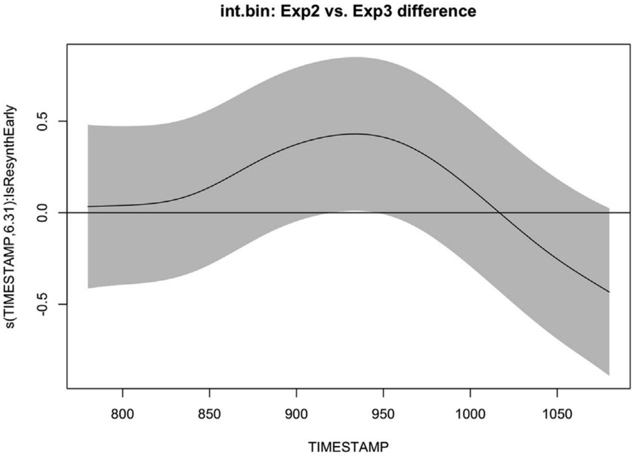


Fig. 11. Difference curve of the difference in competitor fixations in early-peak vs. medial-peak condition across the two experiments (difference between fixations in early-peak vs. medial-peak condition for Experiment 2 minus difference between fixations in early-peak vs. medial-peak condition for Experiment 3). The grey band indicates the 95% CI of the mean of the difference (across experiments) of the difference (between intonation conditions).

4.3. Discussion

In Experiment 3, we used a 3-minute exposure phase prior to the eye-tracking study. The design followed a procedure that has been successfully employed in adaptation studies, i.e., to regional and foreign accents on the segmental level (e.g., Evans & Iverson, 2004; Grohe & Weber, 2016a, 2016b; Witteman, Bardhan, Weber, & McQueen, 2015), to different speaking styles (e.g., Poellmann, Mitterer, & McQueen, 2014), and to suprasegmental variation in non-native speech (Reinisch & Weber, 2012). In each of these studies, an exposure phase of only a few minutes was sufficient to result in processing differences; the number of critical tokens was even higher in our study than in these comparable studies. We thus assume that our exposure phase changed the relative frequency of high-pitched stressed syllables in favour of low-pitched stressed syllables, at least for the speaker participants heard in the experiment (cf. Xie & Myers, 2017, on speaker-specific adaptation). The absence of the stress competitor effect in the early-peak condition compared to the medial-peak condition in Experiment 3 shows that the frequency of occurrence of high-pitched stressed syllables affects stress processing. At the same time, it corroborates the plasticity of speech perception and listeners' ability to generalize accentual realizations to different sentence types. Note that the sentence types differed between exposure and test phase, such that listeners had to generalize the accentual realizations across sentence types.

The stress competitor effect observed in Experiment 2 was not replicated in Experiment 3: There was no effect of *intonation condition* on competitor fixations in Experiment 3 throughout the greatest part of the window of interest (shifted target word onset to shifted segmental U.P.). The short-lived and tiny difference in competitor fixations at the very end of the analysis window in Experiment 3 is probably due to strategic effects (note that we used an average U.P. across all trials, so some trials have an earlier U.P. than others). In a direct statistical comparison between experiments, the effect of *intonation condition* significantly differed in the two experiments, although the effect size of the difference was small. Thus, our finding suggests that it is not directly the acoustic salience of high f0 that causes the competitor activation effect (Option 1) but the frequent exposure to high-pitched stressed syllables (Option 2). Otherwise we would have expected the same stress competitor effect as in Experiment 2. Since we had the same number of participants and stimuli in both experiments, the statistical power was the same in both experiments. Further implications for stress processing and current models of spoken word recognition are addressed in the General Discussion.

5. General discussion

Our results show that f0 peaks on unstressed syllables are (temporarily) interpreted as stressed, leading to more stress identification errors and more fixations to stress competitors. That way, we extend previous studies by providing evidence for the f0 peak as a stress cue in materials other than stress minimal pairs or nonce words, and for online processing in particular. Importantly, the activation of stress competitors can be avoided (or at least considerably reduced) by an increased exposure to low-pitched stressed syllables. Our data speak in favour of a phonological basis of the association between high pitch and metrical stress. In other words, it is not (or not only) the acoustic cue high f0 that is relevant for the perception of metrical stress but the learned association between high f0 and stress that creates the expectation that high f0 signals a stressed syllable, cf. Bishop et al. (this Special Issue) for a similar idea of mediated processing of prominence through phonological structure. Extrapolating from our frequency manipulation in the immediate input, we would expect the effect of pitch accent type on stress processing to be strongest in languages and/or varieties with a majority of medial-peak accents in the input, see below for further discussion.

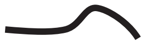
It is an open question which parts of the f0 peak led to our findings. The Introduction reported that various types of high f0 (peak, rise, high plateau) lead to the perception of stress. Experiment 2 only tested rising-falling contours, i.e., the f0 peak was always preceded by a noticeable f0 rise. This is the most frequent realization of H*-accents (Peters et al., 2005; Zahner, Schönhuber, Grijzenhout, & Braun, 2016). It is unclear whether high-pitched unstressed syllables with a high plateau instead of an f0 peak may lead to the same stress competitor effects as reported in Experiment 2. The rising component of a peak contour could be the perceptually most relevant part for the impression of stress, as already indicated by Fry (1958); at the same time, the f0 peak itself might trigger the percept of metrical prominence. Preliminary evidence for the fact that high f0 *per se* is relevant comes from Experiment 1. Here, the words were presented in isolation and had a high plateau on the first syllable in the early-peak condition. The increased number of errors in this misalignment condition suggests that neither a rising part nor a clearly defined f0 peak are strictly necessary to induce a percept of stress but that it is the higher pitch (compared to the preceding or following syllable) that is relevant.

The fact that peak-stress-misalignment affects stress perception straightforwardly explains a number of recent findings. Schwab and Dellwo (2017), for instance, showed that intonational variability hampers stress identification in both native and non-native listeners. They investigated the detection of a stress deviant in Spanish trisyllabic words (stress minimal triplets). The words in a trial were spoken by the same vs. different speakers and with the same vs. different intonation contours (falling vs. rising). Their results showed that all German non-native and Spanish native listeners were less accurate in identifying a stress deviant when intonation varied across test words. The authors argue that listeners' performance was impeded by intonational variability, since f0 could not be relied on in the same way in questions than in declaratives. From our perspective, this processing difficulty is very likely the result of misinterpretations of the stressed syllables due to alignment differences. In the study by Dilley and Heffner (2013), reviewed in the Introduction, there were more last syllable stress responses in the rising continua (which rendered the last syllable high-pitched) compared to the falling continua (Dilley & Heffner, 2013, p. 46, see their Fig. 9). This is what we would predict if this syllable can be stressed. Furthermore, Friedrich, Alter, and Kotz (2001) showed that German listeners were slower and made more errors in reacting to a SW target in an identification study (e.g., *Amboss* [ˈam.bɔs] 'anvil'), when the pitch contour in the target was taken from an accented WS word (e.g., *Abtei* [ap.ˈtaɪ] 'abbey') than when the pitch contour was taken from a SW word. Given that the contours seem to be realized with H*-accents (see Figures in Friedrich et al., 2001) processing appears to be hampered if the stressed syllable is not high-pitched.

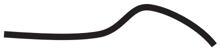
Experiment 3 showed that the frequency of occurrence of high-pitched stressed syllables is partly responsible for why participants treated high-pitched syllables as stressed. It is expected that languages differ in the frequency of occurrence of pitch accents, so our model predicts different sensitivity to f0 in stress perception across languages. This is difficult to test since (a) there are no comparable parallel spoken corpora from which to estimate the frequency distributions and (b) psycholinguistic studies on the processing of stress have not manipulated f0. However, there are a small number of studies that conducted post-hoc correlation analyses between acoustic cues to stress and the observed lexical-decision responses or fixations (Cutler et al., 2007; Reinisch et al., 2010). The results revealed differences in the use of stress cues across languages: English listeners' behaviour correlated only with f0, the cue that showed the largest effect size for the difference between the stressed and the presented unstressed syllables (Cutler et al., 1914, 2007, their Table 1). Dutch listeners' behaviour, on the other hand, correlated with duration and intensity in the stimuli, but not with f0 in Reinisch et al. (2010), despite a larger effect size for f0 than for duration and intensity. These cross-linguistic differences have largely been explained in terms of language-specific cues to stress. Dutch listeners strongly rely on suprasegmental cues, English listeners on vowel quality (or f0, as the most salient cue, if vowel quality is not informative, cf. Cutler, 2012, p. 235). The success of our frequency manipulation provides an alternative explanation for these cross-linguistic differences: Compared to many English varieties with an abundance of H*-accents (Dainora, 2006; Fletcher & Stirling, 2014; Grabe, 2004), high-pitched stressed syllables are less frequent in Dutch. Although there are no quantitative studies for Dutch pitch accents (Miriam Ernestus, p.c), descriptions on Dutch intonation claim that the hat pattern is the most frequent contour in declaratives (Cohen & t'Hart, 1968). The hat pattern consists of a range of low-pitched syllables: a rise and a downstepped fall on the nuclear syllables, a nuclear contour that is acoustically and perceptually close to the early-peak contour. This default pattern leads to a high proportion of low-pitched stressed syllables overall. To corroborate the claim that low-pitched syllables are frequent in Dutch, we analysed the 88 declarative filler sentences from the experiments reported in Braun, Dainora, and Ernestus (2011). In nuclear position, the downstepped fall occurred in more than 90% of the utterances. In prenuclear accents, the late-peak accent was the most frequent accent type (38%), followed in frequency by !H* (24%) and H* (15%). The high number of low-pitched stressed syllables would explain why Dutch listeners do not rely on H* as a stress cue as strongly as English listeners do. We argue that the processing of f0 as a stress cue depends on the language/variety. From a learner perspective, such cross-linguistic differences in the distribution of pitch accents may pose challenges. For instance, given that American English listeners encounter an overwhelmingly high number of high-pitched stressed syllables (90%, Dainora, 2006), we predict inference in L2-processing when learning languages with a high proportion of low-pitched stressed syllables, as in Dutch, Swiss German (Fitzpatrick-Cole, 1999; Leemann, 2012), Indian English (Pickering & Wiltshire, 2000), or Glaswegian English (Smith & Rathcke, this Special Issue).

Importantly, our finding that pitch accent type affects lexical activation in a language in which f0 is not contrastive, poses questions for models of spoken word recognition (e.g., McClelland & Elman, 1986; Norris & McQueen, 2008; Norris, 1994), which currently do not account for suprasegmental information. Experimental findings have shown that listeners use a variety of cues to stress as soon as they become available in the signal. These cues can be segmental, suprasegmental, and purely intonational (see our findings). Therefore, we will have to think about ways to include non-segmental information into current models of spoken word recognition (e.g., similar to Shuai and Malins (2017) for lexical tone in Mandarin Chinese, although pitch plays a different role in tone languages than in intonation languages).

Beyond the lexical level, our findings finally bear further important implications for speech comprehension in general. They suggest that early-peak accents may not just result in (temporary) lexical misinterpretations, but may further lead to a misinterpretation of information structure. For instance, when an early-peak accent results in an f0 peak on a preceding word, this word may be interpreted as accented (H*), see Dilley and Heffner (2013, p. 59) for a similar suggestion. An example of such a potential misinterpretation of information structure is shown in (2) and (3), where the f0 peak on the preposition or the adjective respectively may be temporarily interpreted as accented, which may signal a contrast (e.g., the contrast of *flying from Paris* in (2) or *a small stork* in (3); underlining represents accented words).



(2) He flew to Paris



(3) They saw a big stork.

Taken together, our findings show that high-pitched unstressed syllables (in the form of H-leading tones or high boundary tones) lead to the perception of lexical stress. This directly influences stress judgments and lexical activation in German and also offers explanations for previously observed cross-linguistic differences in stress processing.

6. Conclusion

To conclude, we provide perceptual evidence that the f0 peak is a cue that prompts a percept of stress – regardless of whether the syllable on which it is realized is the metrically stressed one or not. Our results indicate that the perception of metrical strength can be shifted if the f0 peak and the metrically stressed syllable are not aligned, even though there is no direct stress minimal pair. Hence, the phonetic correlates of different pitch accent types can lead to a metrical re-interpretation of the phonological structure of a word, at least temporarily. We argue that this re-interpretation is not driven by the acoustic salience of high pitch, but by a frequent exposure to high-pitched stressed syllables in the input (H*). This suggests that the processing of f0 as a stress cue is mediated by phonological categories, i.e., pitch accent types.

Acknowledgements

We thank Clara Huttenlauch, Angela James, Helena Schlipf, Johanna Schnell, Stefan Volk, and Maximilian Wiesner for testing. We are grateful to Clara Huttenlauch for help with stimuli preparation, to Nathalie Czeke for help with the accompanying R-Markdown, to Ann-Kathrin Grohe for discussion on the design of the exposure-test paradigm (Experiment 3), to Antje Schweitzer for comments on the materials for the rating task in Experiment 3, and finally to Vincent Porretta for help with fixation plots and GAMM. We further appreciate discussion of parts of the data at the *International Conference on Prominence in Language* (June 2015, Cologne), *16th Speech Science and Technology Conference* (December 2016, Sydney), *Phonetics and Phonology in Europe* (June 2017, Cologne). We owe special thanks to Anne Cutler, Carsten Eulitz, Janet Grijzenhout, and Muna Schönhuber for valuable comments on the results of the individual experiments. Finally, we thank the three anonymous reviewers and the two guest editors of this Special Issue, Stefan Baumann and Francesco Cangemi, for their insightful feedback on earlier versions of the manuscript.

Funding: This work was supported by a research grant from the University of Konstanz (*Into-stress,* AFF15/10).

Appendix A

A.1. Experiment 1: Results for correctness rates, split by Experiment (1a vs. 1b)

Experiment 1a:

Model specification: glmer_corr = glmer(corr ~ cond + (1|subject) + (1|item), data = data_1a, family = “binomial”)

Part A. Medial-condition in Intercept	Estimate	Std. Error	*z*-Value	*p*-Value
Intercept	1.7394	0.2923	5.952	<0.0001
Condearly	−1.0825	0.1759	−6.155	<0.0001
Condlate	−0.7759	0.1757	−4.416	<0.0001

Part B. Early-condition in Intercept	Estimate	Std. Error	*z*-Value	*p*-Value
Intercept	0.6569	0.2824	2.326	0.0200
Condlate	0.3066	0.1651	1.857	0.0633
Condmedial	1.0825	0.1759	6.155	<0.0001

Experiment 1b:

Model specification: glmer_corr = glmer(corr ~ cond + (1|subject) + (1|item), data = data_1b, family = “binomial”)

Part A. Medial-condition in Intercept	Estimate	Std. Error	*z*-Value	*p*-Value
Intercept	1.7858	0.3864	4.621	<0.0001
Condearly	−1.2233	0.1903	−6.428	<0.0001
Condlate	−0.7168	0.1876	−3.820	<0.0001

Part B. Early-condition in Intercept	Estimate	Std. Error	*z*-Value	*p*-Value
Intercept	0.5624	0.3789	1.484	0.1377
Condlate	0.5066	0.1817	2.788	0.0053
Condmedial	1.2234	0.1903	6.428	<0.0001

A.2. Experiment 1: Results for analysis of erroneous responses, split by Experiment (1a vs. 1b)

Experiment 1a:

- Syllable-1-erros: 52% in early-peak condition, 23% in medial-, 25% in late-peak condition ($\chi^2 = 50.0$, df = 2, $p < 0.0001$).
- Syllable-3-errors: 60% in late-peak condition, 25% in medial-, 15% in late-peak condition ($\chi^2 = 40.6$, df = 2, $p < 0.0001$).

Experiment 1b:

- Syllable-1-erros: 47% in early-peak condition, 27% in medial-, 26% in late-peak condition ($\chi^2 = 27.2$, df = 2, $p < 0.0001$).
- Syllable-3-errors: 55% in late-peak condition, 20% in medial-, 25% in late-peak condition ($\chi^2 = 27.8$, df = 2, $p < 0.0001$).

Appendix B

Cohort pairs used in cohort trials for Experiments 2 and 3: WSW – SWW and (English translations)

Alaska – Alibi (‘Alaska’ – ‘alibi’); *Albaner – Albatros* (‘Albanian’ – ‘albatross’); *Anapher – Ananas* (‘anaphora’ – ‘pineapple’); *Anode – Anorak* (‘anode’ – ‘anorak’); *Aroma – Arie* (‘aroma’ – ‘aria’); *Embargo – Embryo* (‘embargo’ – ‘embryo’); *Enklave – Enkelin* (‘enclave’ – ‘granddaughter’); *Eskorte – Eskimo* (‘escort’ – ‘Inuit’); *Exotik – Exodus* (‘exoticism’ – ‘exodus’); *Facette – Faserung* (‘facet’ – ‘fibrillation’); *Furore – Furie* (‘sensation’ – ‘fury’); *Genetik – Genesis* (‘genetics’ – ‘genesis’); *Kabine – Kabeljau* (‘cabin’ – ‘cod’); *Kamille – Kamerun* (‘camomile’ – ‘Cameroon’); *Kanister – Kanapee* (‘canister – ‘couch’); *Kanone – Kanada* (‘cannon’ – ‘Canada’); *Karotte – Karitas* (‘carot’ – ‘caritas’); *Kaverne – Kaviar* (‘cavern’ – ‘caviar’); *Kolumne – Kolibri* (‘column’ – ‘colibri’); *Libelle – Libero* (‘dragonfly’ – ‘sweeper’); *Manege – Manitu* (‘arena’ – ‘Manitou’); *Marille – Marathon* (‘apricot’ – ‘marathon’); *Markise – Magier* (‘awning’ – ‘magician’); *Marotte – Marrakesch* (‘quirk’ – ‘Marrakesh’); *Medaille – Medikus* (‘medal’ – ‘medico’); *Monokel – Monitor* (‘monocle’ – ‘monitor’); *Panade – Panama* (‘breading’ – ‘Panama’); *Posaune – Positiv* (‘trombone’ – ‘positive’); *Prämisse – Prämie* (‘premise’ – ‘premium’); *Radieschen – Radius* (‘radish’ – ‘radius’); *Spirale – Spiritus* (‘helix’ – ‘spiritus’); *Statistik – Statue* (‘statistics’ – ‘statue’)

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2019.02.004.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

Altmann, G., & Kamide, J. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347–386). New York: Psychology Press.

Andreeva, B., Barry, W. J., & Wolska, M. (2012). Language differences in the perceptual weight of prominence-lending properties. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association (Interspeech)*, Portland, OR, USA (pp. 2426–2429).

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1993). *The CELEX lexical database [CD-ROM]: Linguistic Data Consortium*. Philadelphia, PA: University of Pennsylvania.

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. N. (2018). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed effects regression models in linguistics* (pp. 49–69). Berlin: Springer.

Baayen, R. H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language, 94*, 206–234.

Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457–474.

Bartels, C. (1999). *The intonation of English statements and questions: A compositional interpretation*. New York, London: Routledge.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967, Retrieved from https://arxiv.org/abs/1506.04967.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2005). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Baumann, S. (2014). The importance of tonal cues for untrained listeners in judging prominence. In *Proceedings of the 10th International Seminar on Speech Production (ISSP)* (pp. 21–24). Cologne, Germany.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics, 38*, 1636–1657.

Baumann, S., & Röhr, C. (2015). The perceptual prominence of pitch accent types in German. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS XVIII)*. Glasgow, UK.

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics, 70*, 20–38.

Berg, T. (1998). *Linguistic structure and change: An explanation from language processing*. Oxford [u.a.]: Clarendon Press.

Bishop, J., Kuo, G., & Kim, B. (this Special Issue). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*.

Boersma, P., & van Heuven, V. J. (2001). Speak and unSpeak with Praat. *Glot International, 5*, 341–347.

Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech, 49*, 451–493.

Braun, B., Dainora, A., & Ernestus, M. (2011). An unfamiliar intonation contour slows down online speech comprehension. *Language and Cognitive Processes, 26*, 350–375.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 306–323.

Büring, D. (1997). *The meaning of topic and focus: The 59th Street Bridge Accent*. London: Routledge.

Cho, T. H. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English. *Journal of the Acoustical Society of America, 117*, 3867–3878.

Cohen, A., & t'Hart, J. (1968). On the anatomy of intonation. *Lingua, 19*, 177–192.

Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & Napoleão de Souza, R. (this Special Issue). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*.

Connell, K., Hüls, S., Martinez-Carzia, M., Qin, Z., Shin, S., Yan, H., & Tremblay, A. (2018). English learners' use of segmental and suprasegmental cues to stress in lexical access: An eye-tracking study. *Language Learning, 68*, 635–668.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45*, 207–228.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, Mass. [u.a.]: MIT Press.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2*, 133–142.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)* (pp. 1913–1916). Saarbrücken, Germany.

Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language, 57*, 483–501.

Dainora, A. (2006). Modeling intonation in English: A probabilistic approach to phonological competence. In L. Goldstein, D. Whalen, & C. Best (Eds.), *Laboratory phonology VIII* (pp. 107–132). Berlin and New York: Mouton de Gruyter.

Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. *International Review of Applied Linguistics and Language Teaching (IRAL), 7*, 294–325.

Dilley, L. C., & Heffner, C. C. (2013). The role of f0 alignment in distinguishing intonation categories: Evidence from American English. *Journal of Speech Sciences, 3*, 3–67.

Dogil, G. (1995). Phonetic correlates of word stress. *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (Univ Stuttgart), 2*, 1–60.

Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology, 58*, 251–274.

Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America, 115*, 352–361.

Féry, C. (1998). German word stress in optimality theory. *Journal of Comparative Germanic Linguistics, 2*, 101–142.

Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia, and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 31–45). New York: Springer.

Fitzpatrick-Cole, J. (1999). The alpine intonation of Bern Swiss German. In *Proceedings of the 14th International Congress of the Phonetic Sciences (ICPhS)* (pp. 941–944). San Francisco, USA.

Fletcher, J., & Stirling, L. (2014). Prosody and discourse in the Australian Map Task Corpus. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 562–575). Oxford: Oxford University Press.

Friedrich, C. K., Alter, K., & Kotz, S. A. (2001). An electrophysiological response to different pitch contours in words. *Cognitive Neuroscience and Neurophysiology, 12*, 3189–3191.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. C. (2004). ERPs reflect lexical identification in word fragment priming. *Journal of Cognitive Neuroscience, 16*, 541–552.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1*, 126–152.

Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistic Vanguard, 3*.

Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles & J. Peters (Eds.), *Regional variation in intonation* (pp. 9–31). Tübingen: Niemeyer.

Grice, M., Baumann, S., & Benzmüller, R. (2005). German intonation in autosegmental-metrical phonology. In J. Sun-Ah (Ed.), *Prosodic typology. The phonology of intonation and phrasing* (pp. 55–83). Oxford: Oxford University Press.

Grohe, A.-K., & Weber, A. (2016a). Learning to comprehend foreign-accented speech by means of production and listening training. *Language Learning, 66*, 187–209.

Grohe, A.-K., & Weber, A. (2016b). The benefit of salience: Salient accented, but not unaccented words reveal accent adaptation effects. *Frontiers in Psychology, 7*, 864.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Heister, J., Würzner, K. R., Bubenzer, J., Pohl, E., Henneforth, T., Geyken, A., & Kiegl, R. (2011). dlexDB: Eine lexikalische Datenbank für die psychologische Forschung [A lexical database for research in psychology]. *Psychologische Rundschau, 62*, 10–20.

Hsu, C.-H., Evans, J. P., & Lee, C.-Y. (2015). Brain responses to spoken f0 changes: Is H special? *Journal of Phonetics, 51*, 82–92.

Hyman, L. M. (1977). On the nature of linguistic stress. In L. M. Hyman (Ed.), *Studies in stress and accent* (pp. 37–82). Los Angeles: University of Southern California USC Linguistics Department.

Isačenko, A. V., & Schädlich, H.-J. (1966). Untersuchungen über die deutsche Satzintonation [Investigations on German sentence intonation]. In *Studia grammatica VII* (pp. 7–67).

Jesse, A., Poellmann, K., & Kong, Y. Y. (2017). English listeners use suprasegmental cues to lexical stress early during spoken-word recognition. *Journal of Speech, Language, and Hearing Research, 60*, 190–198.

Jessen, M., Marasek, K., & Claßen, K. (1995). Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German. *Proceedings of the 13th International Congress of the Phonetic Sciences (ICPhS XIII)*. Stockholm, Sweden.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America, 118*, 1038–1054.

Kohler, K. (1991). 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), 25*, 115–185.

Kohler, K. (2008). The perception of prominence patterns. *Phonetica, 65*, 257–269.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*, 262–268.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.

Lavidor, M., Ellis, A. W., Shillcock, R., & Bland, T. (2001). Evaluating a split processing model of visual word recognition: Effects of word length. *Cognitive Brain Research, 12*, 265–272.

Leemann, A. (2012). *Swiss German intonation patterns*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

Lieberman, P. (1967). *Intonation, perception, and language* (Vol. 38) Cambridge, MA: M. I.T. Press.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53*, 372–380.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.

McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology, 60*, 661–671.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*, 475–494.

Mooshammer, C. (2010). Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *Journal of Acoustical Society of America, 127*, 1047–1058.

Mooshammer, C., & Geng, C. (2008). Acoustic and articulatory manifestations of vowel reduction in German. *Journal of the International Phonetic Association, 38*, 117–136.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13*, 45–52.

Niebuhr, O. (2007). *Perzeption und kognitive Verarbeitung der Sprechmelodie. Theoretische Grundlagen und empirische Untersuchungen [Perception and cognitive processing of intonation. Theory and empirical investigations]*. New York: Mouten de Gruyter.

Niebuhr, O., & Winkler, J. (2017). The relative cueing power of f0 and duration in German prominence perception. In *Proceedings of the 18th Annual Conference of the International Speech Communication Association (Interspeech)* (pp. 611–615). Stockholm, Sweden.

Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: Eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language, 90*, 103–125.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*, 357–395.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.

Peters, B., Kohler, K., & Wesener, T. (2005). 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 K. Kohler, F. Kleber, & B. Peters (Eds.), *Prosodic structures in German spontaneous speech (AIPUK 35a)* (pp. 185–201). Kiel: IPDS.

Pickering, L., & Wiltshire, C. (2000). Pitch accent in Indian-English teaching discourse. *World Englishes, 19*, 173–183.

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Coehn, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge: MIT Press.

Poellmann, K., Mitterer, H. A., & McQueen, J. M. (2014). Use what you can: Storage, abstraction processes, and perceptual adjustments help listeners recognize reduced forms. *Frontiers in Psychology, 5*, 437.

Porretta, V., Kyröläinen, A., van Rij, J., & Järvikivi, J. (2018). VWPre: Tools for preprocessing visual world data, R package version 1.1.0.

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics, 58*, 1–21.

Protopapas, A., Panagaki, E., Andrikopoulou, A., Gutiérrez Palma, N., & Arvaniti, A. (2016). Priming stress patterns in word recognition. *Journal of Experimental Psychology Human Perception and Performance, 42*, 1739–1760.

R Development Core Team (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology, 63*, 772–783.

Reinisch, E., & Weber, A. (2012). Adapting to suprasegmental lexical stress errors in foreign-accented speech. *The Journal of the Acoustical Society of America, 132*, 1165–1176.

Roettger, T. B. (2019). Researcher degrees of freedom in phonetic research. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 10*, 1–27.

Roettger, T. B., & Gordon, M. (2017). Methodological issues in the study of word stress correlates. *Linguistic Vanguard, 3*.

Schwab, S., & Dellwo, V. (2017). Intonation and talker variability in the discrimination of Spanish lexical stress contrasts by Spanish, German and French listeners. *The Journal of the Acoustical Society of America, 142*, 2419–2429.

Schweitzer, A., & Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech)* (pp. 525–529). Lyon, France.

Shattuck-Hufnagel, S., Dilley, L. C., Veilleux, N., Brugos, A., & Speer, R. (2004). F0 peaks and valleys aligned with non-prominent syllables can influence perceived prominence in adjacent syllables. In *Proceedings of the 2nd International Conference on Speech Prosody* (pp. 705–708). Nara, Japan.

Shattuck-Hufnagel, S., Ostendorf, M., & Ross, K. (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics, 22*, 357–388.

Shuai, L., & Malins, J. G. (2017). Encoding lexical tones in jTRACE: A simulation of monosyllabic spoken word recognition in Mandarin Chinese. *Behavior Research Methods, 49*, 230–241.

Sluijter, A. M. C., & Van Heuven, V. J. (1996a). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP)* (pp. 630–633). Philadelphia.

Sluijter, A. M. C., & Van Heuven, V. J. (1996b). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*, 2471–2485.

Smith, R., & Rathcke, T. (this Special Issue). Dialectal phonology constrains the phonetics of prominence. *Journal of Phonetics*.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45*, 412–432.

Sulpizio, S., & McQueen, J. M. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language, 66*, 177–193.

Szalontai, Á., Wagner, P., Mády, K., & Windmann, A. (2016). Teasing apart lexical stress and sentence accent in Hungarian and German. In *Proceedings of the 12th Conference on Phonetics and Phonology in German-speaking countries (P&P)* (pp. 215–218). Munich, Germany.

Tagliapietra, L., & Tabossi, P. (2005). Lexical stress effects in Italian spoken word recognition. In *Proceedings of the XXVII Conference of the Cognitive Science Society*, Stresa, Italy (pp. 2140–2144).

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*, 557–580.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Truckenbrodt, H. (2011). On rises and falls in interrogatives. In *Proceedings of the Conference on Interfaces Discourse & Prosody (IDP) 2009* (pp. 33–46). Paris, France.

van Rij, J., Hollebrandse, B., & Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. In A. Holler & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution* (pp. 267–293). Berlin: Walter de Gruyter.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2016). itsadug: Interpreting time series and autocorrelated data using GAMMs, R package version 2.2.

Wagner, P. (2003). Improving automatic prediction of German lexical stress. In *Proceedings of the 15th International Conference of the Phonetic Sciences (ICPhS)* (pp. 2059–2062). Barcelona, Spain.

Wagner, P., Cwiek, A., & Samlowski, B. (2016). Beat it! Gesture-based prominence annotation as a window to individual prosody processing strategies. In *Proceedings of the 12th Conference on Phonetics and Phonology in German-speaking countries (P&P)* (pp. 211–214). München, Germany.

Wagner, P., Cwiek, A., & Samlowski, B. (this Special Issue). Exploiting the speech-gesture link to capture fine-grained prosodic prominence impressions and listening strategies. *Journal of Phonetics*.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics, 70*, 86–116.

Wiese, R. (1996). *The phonology of German*. Oxford: Clarendon Press.

Witteman, M. J., Bardhan, N. P., Weber, A. C., & McQueen, J. M. (2015). Automaticity and stability of adaptation to a foreign-accented speaker. *Language and Speech, 58*, 168–189.

Witteman, M. J., Weber, A., & McQueen, J. M. (2014). Tolerance for inconsistency in foreign-accented speech. *Psychonomic Bulletin & Review, 21*, 512–519.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton [u. a.]: CRC Press.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*, 3–36.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Boca Raton [u.a.]: CRC Press.

Xie, X., & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language, 97*, 30–46.

Zahner, K., Kember, H., & Braun, B. (2017). 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)* (pp. 1223–1227). Stockholm, Sweden.

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

Zahner, K. (submitted). Pitch accent type affects stress perception in German: Evidence from infant and adult speech processing. PhD thesis, University of Konstanz, Konstanz.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York, NY: Springer.