Mind the peak:
When *museum* is temporarily understood as *musical* in Australian English

Katharina Zahner[1], *Heather Kember*[2,3], *and Bettina Braun*[1]

[1]Linguistics Department, University of Konstanz, Germany;
[2]The MARCS Institute for Brain, Behaviour & Development, Australia;
[3]The ARC Centre of Excellence for the Dynamics of Language, Australia
Katharina.Zahner@uni-konstanz.de

Abstract

Intonation languages signal pragmatic functions (e.g. information structure) by means of different pitch accent types. Acoustically, pitch accent types differ in the alignment of pitch peaks (and valleys) in regard to stressed syllables, which makes the position of pitch peaks an *unreliable* cue to lexical stress (even though pitch peaks and lexical stress often coincide in intonation languages). We here investigate the effect of *pitch accent type* on lexical activation in English. Results of a visual-world eye-tracking study show that Australian English listeners temporarily activate SWW-words (*musical*) if presented with WSW-words (*museum*) with early-peak accents (H+!H*), compared to medial-peak accents (L+H*). Thus, in addition to signalling pragmatic functions, the alignment of tonal targets immediately affects lexical activation in English.

Index Terms: eye-tracking, lexical activation, lexical stress, pitch accent type, intonation, Australian English

1. Introduction

In West-Germanic languages such as English, German, or Dutch, intonation generally conveys post-lexical information, such as information structure, information status, speech act type, or attitudinal meanings. These pragmatic functions are encoded by means of different pitch accent types, i.e. tonal movements that make a certain word in an utterance particularly prominent (e.g., [1, 2]).

According to Autosegmental-Metrical phonology [1, 3], pitch accents are associated with metrically stressed syllables, which in turn are acoustically cued by a longer duration, higher intensity, more vocal effort, and more peripheral vowel quality, compared to unstressed syllables (see [4] for an overview; and references therein). Acoustically, pitch accent types differ in the alignment of pitch peaks (and valleys) in regard to stressed syllables. One case in point is the marking of information status, i.e. whether a referent is new, given, or accessible [1, 5]: Medial-peak accents (H* or L+H* accents) have been argued to signal new or contrastive information, respectively [6, 7]. Here, the pitch peak coincides with the stressed syllable. Early-peak accents (H+L*, H+!H*) are appropriate when a referent is inferable or accessible in the discourse [6, 8]. For this pitch accent type, the pitch peak precedes the stressed syllable. In late-peak accents (L*+H, which are said to convey a "lack of speaker predication" [6, pp. 296]), the pitch peak follows the stressed syllable.

Essentially, stressed syllables might be high-pitched or low-pitched, depending on the pitch accent type that is selected in a given pragmatic situation. From the point of online speech perception, due to phrase-level intonation, pitch peaks are an *unreliable* cue to the position of lexical stress in intonation languages, as they are mainly indicative of something other than lexical stress.

Note though that pitch peaks and lexical stress often coincide in intonation languages (e.g. [9-11]). In German appointment scheduling dialogues, for instance, medial-peak accents occur on average in 42% of the cases, while early-peak or late-peak contours occur in only 14% and 23%, respectively ([9, p. 353]; Kiel Corpus of Spontaneous Speech). In American English spoken radio news corpora, medial-peak accents (H*/L+H*) strongly predominate (90%), while early-peak accents (H+!H*, 5%) and late-peak accents (L*+H, 1%) are rare [10, p. 118]. In Australian English Map task dialogues, (L)+H*-accents account for 55% of the pitch accents, while L*-accents occur in 29% of the cases [11, p. 185].

The focus of this paper is on the interplay between pitch accent type and lexical stress for online speech processing in Australian English (AusE). Recent studies on German suggest that (phrase-level) pitch accent type affects (word-level) stress perception and lexical access [12, 13], with pitch peaks driving these processes: In an offline stress identification task, [13] showed that German listeners made more errors and had longer reaction times for stress judgements when the pitch peak did not coincide with the stress syllable. In an online eye-tracking study, [12] further showed that German adults temporarily fixated SWW-words (e.g. *Libero,* 'sweeper', underlining indicates lexical stress) when they heard segmentally overlapping WSW-words (e.g. *Libelle,* 'dragonfly') that were realised with an early-peak accent (H+L*, H+!H*), i.e. where the pitch peak was realised on the initial unstressed syllable. In German, lexical activation is hence affected by pitch accent type, with non-intended competitors temporarily competing for lexical access.

Using the visual-world eye-tracking paradigm with four printed words on screen [14, 15], we here investigate whether Australian English listeners also perceive high-pitched but unstressed syllables as lexically stressed and consequently fixate SWW-cohort competitors with the wrong stress pattern (e.g. *musical*) during online word recognition if WSW-targets (e.g. *museum*) are produced with an early-peak pitch accent (H+!H*).

2. Experiment

2.1. Methods

2.1.1. Participants

Forty Australian English participants (∅=25.7 years, SD=7.5 years, 29 female, mostly students at the Western Sydney University) with unimpaired vision participated in the study. All of them received a small payment for participation.

2.1.2. Materials

Sixty-four segmentally overlapping cohort pairs that differed in the position of lexical stress were selected. One of the members was stressed on the first syllable, the other on the second syllable. Thirty-two of the pairs were disyllabic (e.g. SW *carton* [ˈkɑːtən] - WS *cartoon* [kɑːˈtuːn]) and 32 were trisyllabic (e.g. SWW *musical* [ˈmjuːzɪkᵊl] - WSW *museum* [mjuːˈziːəm]). The 64 cohort pairs were segmentally identical until at least the onset consonant of the second syllable. Note that the first syllable in WS(W)-words always contained a full vowel. The cohort pairs were matched for lexical frequency (COBUILD frequency per million in the CELEX database [16]: SW: 9.8, WS: 8.8, SWW: 4.6, WSW: 6.5) and number of characters across groups. For each cohort pair, we selected two semantically and phonologically unrelated distractors with comparable number of characters and syllables, and lexical frequencies to be presented on screen. Thirty-two of the 64 cohort pairs were used for cohort trials, half of which were experimental trials (WS(W) as auditory target), half were distractor trials (SW(W) as auditory target). The remaining 32 cohort pairs were used for filler trials, in which one of the unrelated items served as the auditory target.

All auditory targets were embedded in a semantically non-constraining carrier sentence ("The next word is ..."). A male native speaker of AusE (21 years, from Sydney) recorded the stimuli in a sound-attenuated cabin (44.1kHz, 16Bit) at the University of Konstanz. The sentences for cohort trials (experimental and distractor trials) were produced in two intonation conditions each: with an early-peak (H+!H*) and a medial-peak accent (L+H*) on the target and an accent on "next"; see Figure 1. The two productions of a target word were matched along a number of acoustic parameters across conditions; see Table 1 for acoustic analyses in experimental trials. For fillers, half of the sentences were recorded with an early-peak, half with a medial-peak accent on the target, matching the f0-range of their accental movement with the f0-range of cohort pairs. Auditory targets were cross-spliced into one production of the carrier ("The next words is"); splicing was not noticeable.

2.1.3. Procedure

The procedure was identical to [12], except for the equipment. Participants were tested individually in an experimental booth at the MARCS Institute, using the SR Eyelink 1000 in a tower mount system (sampling rate: 500Hz). They sat in front of an Asus LCD-LED (21.5 inch) monitor and their dominant eye was calibrated (pupil and corneal reflection).

In total, the experiment consisted of 64 trials, 32 cohort trials (16 experimental, 16 distractor trials) and 32 filler trials. In experimental trials, the (W)SW-cohort member was presented as audio (8 WSW, 8 WS); in distractor trials, the SW(W)-member (8 SWW, 8 SW), and in 32 filler trials one of the unrelated items. Distractor and filler trials served a strategic function, making participants click equally often on cohort members and filler items throughout the experiment.

Intonation condition was rotated across trials as follows: In experimental and distractor trials, intonation condition was distributed in a Latin-Square Design, i.e. each subject heard both intonation conditions (early- and medial-peak accent) across the whole experiment, but the same item in only one of the two intonation conditions. Half of the filler trials were presented with an early-peak accent, half with a medial-peak accent. Thus, each subject was presented with the same fillers.

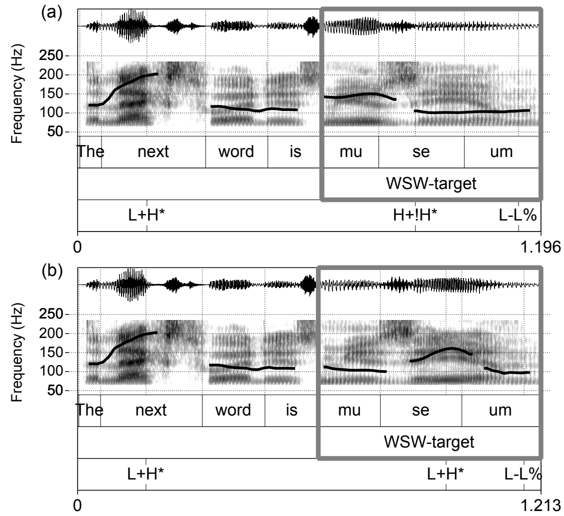

Figure 1: *Example of an experimental trial for the early-peak condition (a) and the medial-peak condition (b).*

Table 1: *Mean values (and standard deviations) of acoustic realisations of WS(W)-targets in the two intonation conditions in experimental trials.*

Acoustic variable	**Early-peak condition**		**Medial-peak condition**	
	WSW	**WS**	**WSW**	**WS**
F0-excursion of accentual movement in st	8.59 (0.60)	8.59 (0.79)	8.60 (0.67)	8.58 (0.77)
Duration of first syllable in ms	138 (34)	154 (26)	136 (36)	153 (23)
Duration of second (stressed) syllable in ms	232 (42)	404 (57)	233 (40)	404 (56)
H1*-A3* ratio ([17]) in middle of vowel 1 in dB	14.1 (3.6)	13.9 (9.8)	8.6 (6.9)	9.1 (6.3)
H1*-A3* ratio in middle of vowel 2 (stressed) dB	8.4 (4.9)	11.4 (8.1)	11.9 (4.6)	14.0 (7.7)

In the experimental lists, we pseudo-randomised the order of the trials such that each experimental half contained the same number of cohort, distractor and filler trials with the constraint of an experimental item being at most the third item of the same intonation condition in a row, among other criteria. 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.

Each trial started with a black cross on white background, centred on screen, which remained there 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 were framed by a rectangular box (6.5cm x 4cm). The position of the items on screen was counterbalanced across conditions, such that the target to click on occurred equally often in the four possible positions for each intonation condition. The carrier phrase started 2000ms after the words occurred on screen, giving participants a preview of the words of 2635ms. Auditory stimuli were presented via headphones (Beyerdynamic DT-770 Pro, 80 OHM) at comfortable loudness. An automatic drift correction occurred after every fifth trial. After half of the trials (32 trials), there was an optional pause. In total, it took participants approximately 15 minutes to complete the experiment.

2.2. Results

Participants correctly clicked on the auditory target in 97.6% of all experimental trials (WS(W) word as auditory target). The average response time in these trials was 670.2ms after target offset. Results of a linear mixed effects regression model (lmer) [18] with *intonation condition* as fixed factor and *participants* and *items* as crossed random factors [19, 20] showed no effect of *intonation condition* (p>0.5). Error rates were also not affected by *intonation condition* (logistic regression model, p>0.4).

Fixation data were extracted in 4ms bins. Only fixations in experimental trials were analysed further. Fixations were automatically labelled as being directed to the target (WS(W), *museum*), the stress competitor (SW(W), *musical*) or to the unrelated distractors if they fell within a square of 200x200 pixels around the respective word. Figure 2 shows the evolution of fixations in experimental trials to the four words on screen in the two intonation conditions, i.e. when the WS(W) target was presented with an early-peak pitch accent (H+!H*; 2a) or a medial-peak pitch accent (L+H*; 2b).

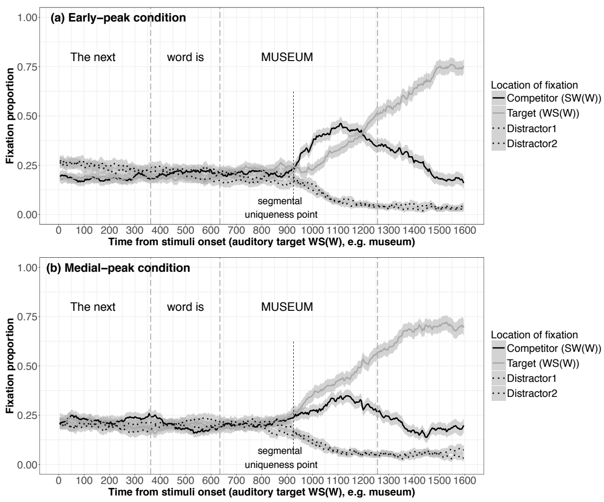

Figure 2: *Evolution of fixations to competitor (SW(S)), target (WS(W) and the two distractors in experimental trials in the early-peak condition (a) and medial-peak condition (b). Grey dashed vertical lines refer to acoustical landmarks.*

Fixation proportions to the (segmentally unrelated) distractors decreased around 260ms after the onset of the auditory target (*museum*) in both intonation conditions (at 895ms after the onset of the target sentence; Figure 2), while fixation proportions to the target (*museum*) and the stress competitor (*musical*) both further increased from the point of distractor divergence onwards. In the early-peak condition (Figure 2a), the competitor (*musical*) was ruled out as the potential word at around 565ms after target onset; in the medial-peak condition (Figure 2b), at around 485ms after target onset. Not only was the stress competitor discarded later in the early-peak condition than in the medial-peak condition, the competitor was fixated more than the target in this condition (from 265ms to 565ms relative to the target onset), while the competitor was never preferred over the target in the medial-peak condition. For ease of comparison, we show the respective differences in fixations to the stress competitor (Figure 3) and the target (Figure 4) across intonation conditions in one graph.

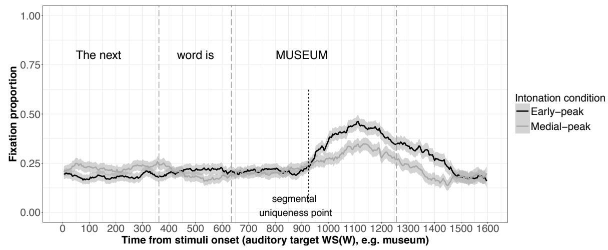

Figure 3: *Fixations to SW(W) stress competitor in experimental trials (WS(W) as auditory target) in the two intonation conditions.*

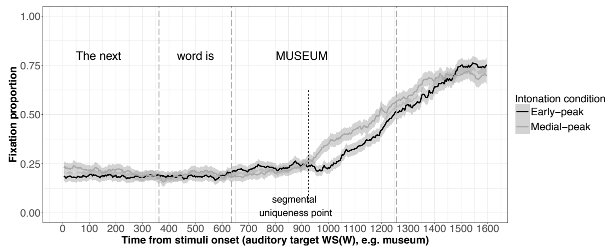

Figure 4: *Fixations to WS(W) target in experimental trials (WS(W) as auditory target) in the two intonation conditions.*

To statistically corroborate the differences of competitor fixations (Figure 3) and target fixations (Figure 4) across intonation conditions, empirical logits of fixations (elogs), the log-ratio of fixations to one of the four words, divided by the fixations directed to the three other objects or somewhere else [21], were calculated. Statistical analyses were run for two windows: Window 1 (895-1185ms) represents the time from target onset (635ms) to the segmental uniqueness point (UP; 925ms), shifted by a processing time of 260ms; window 2 (1185-1515ms) the time from the UP to the end of the target (1255ms), again shifted by 260ms. We entered *intonation condition* as fixed factor in a lmer model; *participants* and *items* were modelled as crossed random factors in order to account for adjustments of intercepts and slopes.

For competitor fixations, there was a main effect of *intonation condition* in both windows. During the processing of the segmentally ambiguous part (window 1), participants fixated the SW(W)-competitor (*musical*) more when the WS(W)-target (*museum*) was presented with an early-peak, H+!H*, (av. elogs: -1.52) than when presented with a medial-peak accent, L+H*, (av. elogs: -2.05; ß=0.5 [0.01;1.04], SE=0.26, t=2.02, p=0.04). This effect was preserved even after segmental disambiguation towards the end of the target (window 2; ß=0.6 [0.13;1.13], SE=0.25, t=2.49, p=0.01); see Figure 3. Target fixations also differed as a function of *intonation condition* during the processing of the segmentally ambiguous part (window 1): There were more fixations to the WS(W)-target (*museum*) in the medial-peak condition (av. elogs: -1.51), compared to the early-peak condition (av. elogs: -2.02), a difference that approached significance (ß=0.5 [0.03; 1.07], SE=0.28, t=1.85, p=0.07); see Figure 4.

Interestingly, the effect of intonation condition on SW(W)-competitor fixations observed for AusE listeners occurred approximately 180ms later than in German [12]. However, there was no interaction between *language* (AusE vs. German) and *intonation condition* (early-peak accent vs. medial-peak accent) in six subsequent 100ms-windows from target onset (all p-values>0.2). Nevertheless, German listeners [12] clicked on the target on average 105ms earlier than AusE listeners (ß=105 [10.7;199.1], SE=47.5, t=2.21, p=0.03).

2.3. Discussion

Our fixation data show that AusE listeners activated the SW(W)-stress competitor more when the WS(W)-target was presented with an early-peak accent (H+!H*) than when presented with a medial-peak accent (L+H*). Conversely, the WS(W)-target received fewer fixations in the early-peak than in the medial-peak condition. Hence, pitch peaks on unstressed syllables (as in early-peak accents) lead to a temporary activation of a competitor with initial stress. For AusE listeners, high-pitched syllables (even when unstressed) seem to be an indicator for lexical stress and consequently pitch accent type influences lexical access.

The slower click latencies for AusE compared to German listeners might be interpreted as indicating that AusE listeners experience higher processing costs for the processing of suprasegmental stress cues. As English encodes stress differences by vowel reduction in many unstressed syllables [22], the use of suprasegmentals might be more costly for English listeners than for listeners that commonly need to rely on these cues to resolve lexical competition (German or Dutch [23]). Alternatively, however, sampling differences may be responsible for the observed differences across languages.

3. General Discussion and Outlook

Taken together, we showed that pitch accent type affects lexical activation in Australian English (as was observed for German [12]). AusE listeners use f0-cues during lexical processing and perceive high-pitched but unstressed syllables temporarily as stressed, which in turn leads to the activation of competitor words that are not intended by the speaker.

This finding is particularly interesting in regard to cross-linguistic differences in the use of suprasegmentals in intonation languages. English listeners have been shown to make little use of suprasegmental cues [24-26]: For instance, stress minimal pairs, such as *forbear – forbear,* primed each other's associates. Hence, listeners treated these pairs as homophones, despite their suprasegmental (stress) difference [25]. It has been argued that the pay-off for using suprasegmentals in English is smaller than in German or Dutch [23], as lexical stress is more frequently cued by vowel reduction (segmental information) in English than in German and Dutch [22]. Yet, our results corroborate the main findings in [27, 28], which showed that English listeners *can* use suprasegmentals, although they do so less efficiently than Dutch or German listeners: stress-mismatching primes (mu – museum) do not inhibit lexical access, while doing so in Dutch or German [27]. [28] also show that when English listeners are encouraged to use suprasegmentals, i.e. when segmental cues are not informative, they primarily rely on pitch (the higher the f0, the more likely a syllable was perceived as stressed; see also [29], as an early study identifying f0 as the primary perceptual cue to stress for English). Thus, English listeners may make less use of suprasegmentals as it is less beneficial in English [23], but pitch peaks seem to be a strong cue for stress perception and hence for lexical activation.

Yet, relying on pitch peaks during online processing is a strategy that is not profitable in all cases (neither for AusE nor for German listeners), as high-pitched syllables are indicative of more than just lexical stress. In fact, post-lexical meaning is conveyed by differences in pitch accent types, which in turn involve differences in the alignment of pitch peaks relative to stressed syllables. If listeners erroneously activate cohort competitors with the wrong stress pattern when words are realised with certain pitch accent types (here: early-peak accents), used to convey pragmatic functions, lexical access is slowed down, resulting in higher processing costs.

The question remains why pitch peaks are used for lexical access in online speech comprehension at all. First, high-pitched syllables are perceived as more salient than low-pitched ones [30-32]; acoustic salience might be interpreted as metrical prominence. Second, high-pitched stressed syllables are more frequent than low-pitched stressed syllables in AusE [11], despite it being an "uptalk" variety [33] (in which L*-accent proportions are assumingly higher than in non-uptalk varieties). At present, both mechanisms are equally likely; they might even both contribute to the effects observed in AusE. In future experiments, we plan to assess the role of input frequency by replicating the experiment in other varieties of English with less high-pitched stressed syllables than AusE, such as Indian English [34]. Alternatively, we plan to use an exposure phase to manipulate the frequency of high-pitched stressed syllables in the immediate input.

Another open question is whether the observed competitor activation caused by alignment differences is unidirectional (i.e. pitch peaks being interpreted as stressed) or whether it is also bidirectional (i.e. peak valleys being simultaneously interpreted as *un*stressed). It is unknown how low-pitched, but metrically stressed syllables are processed. We will thus investigate whether accent types with low-pitched stressed syllables (e.g. L*+H) also lead to stress competitor activation – so that *musical* might temporarily be understood as *museum.*

4. Acknowledgements

This research was supported by a DAAD research grant (Australia – Germany Joint Research Cooperation Scheme) co-awarded to the last author. We thank Christopher Janssen for recording the stimuli. We owe special thanks to Prof. Anne Cutler for hosting us in her lab and for valuable comments and suggestions on the data.

5. References

[1] D. R. Ladd, *Intonational phonology* vol. 119. Cambridge [u.a.]: Cambridge University Press, 2008.

[2] P. Prieto, "Intonational meaning," *Wiley Interdisciplinary Reviews: Cognitive Science,* vol. 6, pp. 371-381, 2015.

[3] J. B. Pierrehumbert, "The phonology and phonetics of English intonation," PhD Thesis, Massachusetts Institute of Technology, Dept. of Linguistics and Philosophy, Bloomington, 1980.

[4] A. Cutler, "Lexical stress," in *The handbook of speech perception*, D. B. Pisoni and R. E. Remez, Eds., Oxford: Blackwell Publishing, 2005.

[5] C. Gussenhoven, *The phonology of tone and intonation*. Cambridge: Cambridge University Press, 2004.

[6] J. Pierrehumbert and J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," in *Intentions in communication,* P. Coehn, J. Morgan, and M. Pollack, Eds., Cambridge: MIT Press, 1990.

[7] D. G. Watson, M. K. Tanenhaus, and C. A. Gunlogson, "Interpreting pitch accents in online comprehension: H* vs. L+H*," *Cognitive Science,* vol. 32, pp. 1232-1244, 2008.

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

[9] B. Peters, K. Kohler, and T. Wesener, "Melodische Satzakzentmuster in prosodischen Phrasen deutscher Spontansprache - Statistische Verteilung und sprachliche Funktion," in *Prosodic Structures in German Spontaneous Speech (AIPUK 35a)*, K. Kohler, F. Kleber, and B. Peters, Eds., Kiel: IPDS, 2005, pp. 185-201.

[10] A. Dainora, "Modeling intonation in English: A probabilistic approach to phonological competence," in *Laboratory Phonology*. vol. 8, L. Goldstein, D. Whalen, and C. Best, Eds., Berlin and New York: Mouton de Gruyter, 2006, pp. 107-132.

[11] J. McGregor and S. Palethorpe, "High rising tunes in Australian English: The communicative function of L* and H* pitch accent onsets," *Australian Journal of Linguistics,* vol. 28, pp. 171-193, 2008.

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

[13] S. Egger and B. Braun, "The role of pitch accents on the identification of metrical prominence," Paper at the International Conference "Prominence in Language", Cologne, Germany, 2015.

[14] J. M. McQueen and M. Viebahn, "Tracking recognition of spoken words by tracking looks to printed words," *The Quarterly Journal of Experimental Psychology,* vol. 60, pp. 661-671, 2007.

[15] E. Reinisch, A. Jesse, and J. M. McQueen, "Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately," *The Quarterly Journal of Experimental Psychology,* vol. 63, pp. 772-783, 2010.

[16] H. R. Baayen, R. Piepenbrock, and L. Gulikers, "The CELEX lexical database [CD-ROM]: Linguistic Data Consortium," Philadelphia, PA: University of Pennsylvania, 1995.

[17] C. Mooshammer, "Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German," *Journal of Acoustical Society of America,* vol. 127, pp. 1047-1058, 2010.

[18] R. H. Baayen, *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press, 2008.

[19] D. J. Barr, R. Levy, C. Scheepers, and H. Tily, "Random-effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language,* vol. 36, pp. 255-278, 2013.

[20] R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language,* vol. 59, pp. 390-412, 2008.

[21] D. J. Barr, T. M. Gann, and R. S. Pierce, "Anticipatory baseline effects and information integration in visual world studies," *Acta Psychologica,* vol. 137, pp. 201-207, 2011.

[22] P. Delattre, "An acoustic and articulatory study of vowel reduction in four languages," *International Review of Applied Linguistics and Language Teaching (IRAL),* vol. 7, pp. 294-325, 1969.

[23] A. Cutler and D. Pasveer, "Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition," in *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden, 2006.

[24] L. H. Small, S. D. Simon, and J. S. Goldberg, "Lexical stress and lexical access: Homographs versus nonhomographs," *Perception & Psychophysics,* vol. 44, pp. 272-280, 1988.

[25] A. Cutler, "Forbear is a homophone: Lexical prosody does not constrain lexical access," *Language and Speech,* vol. 29, pp. 201-220, 1986.

[26] S. Z. Bond and H. L. Small, "Voicing, vowel, and stress mispronunciations in continuous speech," *Perception and Psychophysics,* vol. 34, pp. 470-474, 1983.

[27] N. Cooper, R. Wales, and A. Cutler, "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," *Language and Speech,* vol. 45, pp. 207-228, 2002.

[28] A. Cutler, R. Wales, N. Cooper, and J. Janssen, "Dutch listeners' use of suprasegmental cues to English stress," in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 2007, pp. 1913-1916.

[29] D. B. Fry, "Experiments in the perception of stress," *Language and Speech,* vol. 1, p. 126, 1958.

[30] S. Baumann and C. Röhr, "The perceptual prominence of pitch accent types in German," in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK, 2015.

[31] R. Turnbull, A. J. Royer, K. Ito, and S. R. Speer, "Prominence perception is dependent on phonology, semantics, and awareness of discourse," *Language, Cognition and Neuroscience,* pp. 1-17, 2017.

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

[33] P. Warren, *Uptalk: The phenomenon of rising intonation*. Cambridge: Cambridge University Press, 2016.

[34] L. Pickering and C. Wiltshire, "Pitch accent in Indian-English teaching discourse," *World Englishes,* vol. 19, pp. 173-183, 2000.