JSLHR

Research Article

English Listeners Use Suprasegmental Cues to Lexical Stress Early During Spoken-Word Recognition

Alexandra Jesse,[a] Katja Poellmann,[b] and Ying-Yee Kong[b]

Purpose: We used an eye-tracking technique to investigate whether English listeners use suprasegmental information about lexical stress to speed up the recognition of spoken words in English.
Method: In a visual world paradigm, 24 young English listeners followed spoken instructions to choose 1 of 4 printed referents on a computer screen (e.g., "Click on the word *admiral*"). Displays contained a critical pair of words (e.g., *ˈadmiral*–*ˌadmiˈration*) that were segmentally identical for their first 2 syllables but differed suprasegmentally in their 1st syllable: One word began with primary lexical stress, and the other began with secondary lexical stress. All words had phrase-level prominence. Listeners' relative proportion of eye fixations on these words indicated their ability to differentiate them over time.
Results: Before critical word pairs became segmentally distinguishable in their 3rd syllables, participants fixated target words more than their stress competitors, but only if targets had initial primary lexical stress. The degree to which stress competitors were fixated was independent of their stress pattern.
Conclusions: Suprasegmental information about lexical stress modulates the time course of spoken-word recognition. Specifically, suprasegmental information on the primary-stressed syllable of words with phrase-level prominence helps in distinguishing the word from phonological competitors with secondary lexical stress.

In variable-stress languages, such as English, Dutch, and Spanish, the relative emphasis of syllables within a word is not governed by a rule but rather is a property of the word. This lexically defined stress can thus be used to recognize words (e.g., Cooper, Cutler, & Wales, 2002; Friedrich, Kotz, Friederici, & Gunter, 2004; Jesse & McQueen, 2013; Mattys, 2000; Reinisch, Jesse, & McQueen, 2010; Soto-Faraco, Sebastián-Gallés, & Cutler, 2001; Sulpizio & McQueen, 2012; van Donselaar, Koster, & Cutler, 2005). Lexical stress can be expressed through segmental information, primarily through a vowel change in unstressed syllables, and through systematic suprasegmental changes in fundamental frequency, duration, amplitude, and spectral tilt (Beckman, 1986; Campbell & Beckman, 1997; Fry, 1955, 1958; Lieberman, 1960; Sluijter & van Heuven, 1996). The relative implementation of lexical stress through segmental and suprasegmental cues in production differs across languages, and along with it, listeners of these languages weigh these cues differently in perception (e.g., Cooper et al., 2002; Zhang & Francis, 2010).

Early studies on stress perception suggested that English listeners may consider only segmental cues to lexical stress (e.g., Bond, 1981; Cutler, 1986; Cutler & Clifton, 1984; Fear, Cutler, & Butterfield, 1995). In contrast, recent work has shown that although English listeners rely primarily on segmental cues to lexical stress (e.g., Braun, Lemhöfer, & Mani, 2011; Cutler & Clifton, 1984; Fear et al., 1995), they also use suprasegmental cues to make fine-grained distinctions between different degrees of lexical stress (e.g., Mattys, 2000). Distinguishing various degrees of lexical stress is useful because secondary-stressed syllables are ubiquitous in English; Mattys (2000) estimated that 41% of all English words contain a syllable with secondary lexical stress. English listeners can use suprasegmental cues to distinguish primary-stressed syllables from secondary-stressed ones (e.g., Fry, 1958; Mattys, 2000; Mattys & Samuel, 2000) as well as from unreduced, unstressed syllables (Cooper et al., 2002). However, English listeners' ability to use this suprasegmental information is limited compared with that of Dutch listeners. Dutch listeners are more accurate in determining the stress pattern of word fragments than English

[a]Department of Psychological and Brain Sciences, University of Massachusetts, Amherst
[b]Department of Communication Sciences and Disorders, Northeastern University, Boston, MA
Correspondence to Alexandra Jesse: ajesse@psych.umass.edu

Editor: Nancy Tye-Murray
Associate Editor: Mitchell Sommers
Received September 28, 2015
Revision received April 25, 2016
Accepted June 13, 2016
DOI: 10.1044/2016_JSLHR-H-15-0340

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

listeners, even when both groups of listeners are presented with English materials (Cooper et al., 2002). This cross-linguistic difference in performance goes hand in hand with cross-linguistic differences in the lexical statistics between the two languages in that stress is more informative for word recognition in Dutch than in English (Cutler & Pasveer, 2006; van Heuven & Hagman, 1988). For example, Cutler and Pasveer (2006) reported that lexical stress information reduces the number of embedded words in Dutch more than in English.

In most previous research, lexical stress was associated with phrase-level prominence.[1] Critical words were accentuated as they were recorded in isolation or in a focal position within a sentence. In these cases, pitch accents are aligned with the primary-stressed syllable of the accentuated word, consequently increasing the syllable's perceptual salience (e.g., Bolinger, 1958). Unstressed syllables do not receive pitch accents. Secondary stressed syllables can be associated with a pitch accent when they precede a syllable with primary lexical stress in a word with prominence (e.g., Beckman, 1996; Bolinger, 1958; Hayes, 1995; Plag, Kunter, & Schramm, 2011). These prenuclear accents placed on the secondary-stressed syllables differ acoustically from nuclear pitch accents placed on primary-stressed syllables (for an overview of accent and stress in English, see Gussenhoven, 2011), allowing listeners to make a perceptual distinction. Given that accentuated words were used in previous studies, the discussion on the use of suprasegmental cues to lexical stress has thus focused on this specific case where the perception of lexical stress could have been enhanced, if not solely driven, by prominence.

The aforementioned studies suggested that English listeners can use suprasegmental information to detect the stress pattern of accentuated words; however, this information may not be used immediately in spoken-word recognition in English. Using cross-modal fragment priming, Cooper et al. (2002) tapped into the ongoing spoken-word recognition process. In one experiment, English listeners heard bisyllabic fragment primes that had primary lexical stress (e.g., ˈ*admi* taken from ˈ*admiral*) or secondary lexical stress on their first syllable (ˌ*admi* taken from ˌ*admi*ˈ*ration*) and were always unstressed and unreduced in their second syllable. In a second experiment, English listeners heard monosyllabic fragment primes that had primary lexical stress (e.g., ˈ*mu* taken from ˈ*music*) or were unstressed with unreduced vowels (*mu* taken from *mu*ˈ*seum*). Again, all fragments came from words with phrase-level prominence. Two results from these experiments are notable. First, both monosyllabic and bisyllabic fragment primes sped up subsequent lexical decisions to printed words if the primes matched the beginnings of these targets both segmentally and suprasegmentally (e.g., hearing ˈ*admi* before seeing ˈ*admiral*; hearing ˈ*mu* before seeing ˈ*music*) compared with when the fragments matched targets only segmentally and not suprasegmentally (e.g., hearing ˈ*admi* before seeing ˌ*admi*ˈ*ration*; hearing ˈ*mu* before seeing *mu*ˈ*seum*). Second, bisyllabic suprasegmentally mismatching (segmentally matching) primes produced no priming effect compared with a control condition with unrelated primes, suggesting that segmental facilitation was canceled out by suprasegmental inhibition. However, this was not the case for monosyllabic suprasegmentally mismatching (segmentally matching) primes where facilitatory priming was still observed. These results contrast those found for Dutch, where Dutch listeners efficiently extracted suprasegmental information from only one syllable of an accentuated word (van Donselaar et al., 2005). Cooper et al. (2002) concluded that English listeners rely primarily on segmental information for lexical activation and that two syllables are needed for listeners to use suprasegmental information about lexical stress for the recognition of spoken words. The second syllable could help because it may provide context for the interpretation of the suprasegmental information of the first syllable or because the presentation of a second syllable gives listeners more time to process the information from the first syllable before the subsequent target presentation. Either way, these results suggest that English listeners' use of suprasegmental cues to lexical stress during spoken-word recognition may be delayed rather than immediate.

The findings from Cooper et al. (2002) cannot pinpoint when these cues are used during online word recognition. They cannot answer the crucial question of whether English listeners can exploit these suprasegmental cues for word recognition in time before segmental cues disambiguate the word. That is, when hearing not just fragments but rather entire spoken words, can English listeners use suprasegmental information from hearing ˈ*ad* to resolve lexical competition and speed up word recognition before the spoken word is segmentally disambiguated as ˈ*admiral*? It is possible that, in the Cooper et al. study, suprasegmental cues were not fully processed by the listeners until the target word was presented. In this case, suprasegmental cues would not contribute much to the recognition of the spoken full words because segmental information could have disambiguated the word by then. But if English listeners can indeed use suprasegmental cues to lexical stress early enough to resolve lexical competition and facilitate target recognition, this would alter the time course of spoken-word recognition.

In the present study, we investigated this question by tracking the time course of spoken-word recognition with a visual world paradigm. When trying to recognize a spoken word, listeners consider words stored in their mental lexicon as viable candidates to the degree that they temporally match the incoming information (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Luce & Pisoni, 1998; Salverda, Dahan, & McQueen, 2003; Zwitserlood, 1989). The visual world paradigm allows tracking this spoken-word recognition process over time because, at any given time, listeners' spontaneous fixations on a referent shown on a computer screen (as a picture or a printed word) are linked to the

[1]The term *phrase-level prominence* is used in this article to refer to the relative emphasis of words within a phrase. These words are realized by a larger pitch movement, longer duration, and increased amplitude compared with nonemphasized words.

momentary degree of overall support for this word as a lexical candidate (Allopenna et al., 1998). The visual world paradigm with pictures or with printed words can capture the time course of phonological (segmental) competition among lexical candidates (e.g., Huettig & McQueen, 2007; McQueen & Viebahn, 2007; Poellmann, Mitterer, & McQueen, 2014; Reinisch, Jesse, & McQueen, 2011a; Salverda & Tanenhaus, 2010) and is sensitive to how prosodic information alters online spoken-word recognition in various languages (e.g., Ito & Speer, 2008; Reinisch et al., 2010; Salverda et al., 2003, 2007). Most relevant to the present study, in a visual world paradigm with printed words, Dutch listeners used suprasegmental information associated with the primary-stressed syllable of accentuated words to speed up spoken-word recognition and to resolve lexical competition before the target words became segmentally unique (Reinisch et al., 2010). Given the previously mentioned differences between Dutch and English in regard to lexical statistics and the efficiency of processing suprasegmental information, it is yet to be determined whether English listeners, just like Dutch listeners, can use suprasegmental information immediately for spoken-word recognition.

To address this question, we followed the design of Reinisch et al. (2010). As in Cooper et al. (2002), critical words in our study were stress pairs that overlapped segmentally in at least their first two syllables but had either primary or secondary lexical stress on the first syllable (e.g., *ˈadmiral–ˌadmiˈration*). We tested whether, before segmental information distinguishes these words, listeners would fixate more on the printed target word (e.g., *ˈadmiral*) that matched the auditory input both segmentally and suprasegmentally than on a segmentally matching but suprasegmentally mismatching competitor word (e.g., *ˌadmiˈration*). We also tested whether words would compete more for recognition if they had primary rather than secondary lexical stress on their initial syllable, as previously observed for Dutch (Reinisch et al., 2010). Demonstrating that English listeners can use suprasegmental cues to lexical stress as soon as they become available would significantly advance our understanding of the importance of suprasegmental prosodic cues for spoken-word recognition.

Method

Participants

Twenty-four undergraduate students (mean age = 20.5 years) from the University of Massachusetts Amherst participated for class credit. All participants were monolingual native speakers of American English who had no language, hearing, or attention deficits. All had normal or corrected-to-normal vision.

Materials

Twenty-four sets of four words were created. Each set consisted of two critical and two noncritical words, which always were shown together as printed words on a display during the experiment. One of the words was also presented auditorily as the target word at the end of the sentence "Click on the word ____." Critical words were three- or four-syllable English word pairs, with the exception of one five-syllable word. In all of these word pairs, one member had primary lexical stress (e.g., *ˈadmiral*) and the other member had secondary lexical stress on the first syllable (e.g., *ˌadmiˈration*). It is important to note that these critical words overlapped segmentally for at least their first two syllables (supported by statistical analyses showing that the vowels in these syllables do not significantly differ, respectively, in their first formant and second formant frequencies as a function of stress; all $p > .05$). Thus, during the first two syllables, these words differed only in their suprasegmental properties. Noncritical words in these sets were three- to five-syllable English word pairs that overlapped segmentally and suprasegmentally in their first two syllables (e.g., *converter–convergence*). One third of these noncritical pairs had primary lexical stress on the first syllable, one third had primary lexical stress on the second syllable, and the remaining third had primary lexical stress on the third syllable. By using phonologically overlapping word pairs as noncritical words, we were able to keep target words unpredictable as we repeated these sets over the course of the experiment.

Within a set, all four words were morphologically and semantically unrelated. To assess overall phonological competition the same way as in Reinisch et al. (2010), critical and noncritical word pairs within a set were also phonologically unrelated to each other. They were matched in their spoken word frequency (Davies, 2008): critical words, $M = 520$; noncritical words, $M = 463$; $t(47) = 0.31$, $p = .76$. Critical words were also matched in frequency across stress type: primary-stress items, $M = 521$; secondary-stress items, $M = 518$; $t(23) = 0.02$, $p = .99$. Eight more sets were created for filler trials, and six more sets were created for practice trials. In practice sets, words within both pairs overlapped segmentally and suprasegmentally in their first two syllables, and primary lexical stress occurred equally often on the first, second, or third syllable.

A female native speaker of American English was recorded at 44.1 kHz in a sound-attenuated booth producing all words at the end of the sentence "Click on the word ____." Similar to Reinisch et al. (2010), the critical words therefore received pitch accents (see the Appendix for pitch excursion information on the first vowels). Stimuli were equalized in their overall root-mean-squared intensity. The duration of the carrier sentences was 744 ms on average and was similar across stress type for critical items, $t(23) = 0.05$, $p = .96$. We compared the acoustics of the vowel in the first syllable of all critical words by stress type (see Table 1). One stress pair (*uniform–universal*) was removed from all acoustic analyses (but not from the experiment) because its difference in spectral tilt exceeded the mean of all stress pairs by more than 2.5 *SD*s. Pitch excursion was larger in vowels when the syllable carried primary rather than secondary lexical stress. Vowels unexpectedly had lower mean pitch and root-mean-squared intensity for first syllables with primary lexical stress than with secondary lexical

Table 1. Mean values of acoustic measures on the vowel in the first syllable of words with primary and secondary lexical stress on that syllable and significance levels of their difference.

Measurement	Primary stress vowel	Secondary stress vowel	$t(22)$	p	Cohen's d
Duration (ms)	97	93	1.70	*ns*	0.13
Mean pitch (Hz)	163	178	–6.48	< .0001	1.50
Pitch excursion (semitones/s)[a]	18	12	3.86	< .0010	0.79
Spectral tilt (dB)[b]	–7.55	–7.60	0.11	*ns*	0.01
Root-mean-square (dB)	70.61	72.00	–3.47	< .0100	0.65

[a]Calculated as the range of fundamental frequency over the vowel duration in the first syllable (de Pijper, 1983). [b]Defined as the difference in amplitude of the band containing the fundamental frequency (0–300 Hz) and the band containing the formants.

stress. Spectral tilt and duration did not differ significantly by stress type.

Procedure

Participants were tested individually in a sound-attenuated booth and were seated 60 cm away from a 1,024 × 768 (17-in. diagonal) computer screen (60-Hz refresh rate). During an initial familiarization phase, all words of the experiments were presented one at a time in a random order for participants to read out loud. Presentation was self-paced, and no feedback was given.

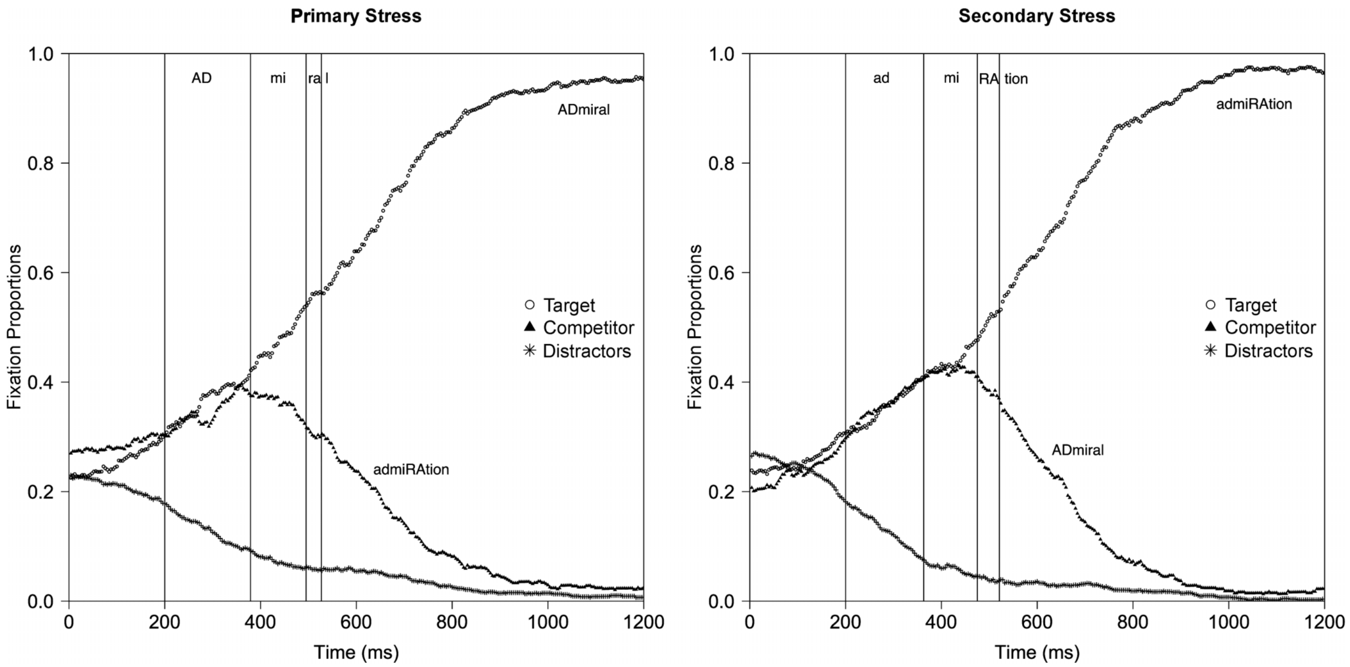
Eye movements were recorded with a desktop-mounted Eyelink 1000 system (SR Research Ltd., Kanata, Ontario, Canada). The eye tracker was first calibrated to each participant. Participants kept the head in a headrest throughout the experiment. During the experiment, each trial began with a fixation cross, shown for 500 ms in the center of the screen. A black screen followed for 200 ms before a set of four printed-word response alternatives was displayed (Lucida Sans Typewriter, size 20), with each word centered in one of the four quadrants of the screen. Fixations of the participants' right eye on these words were recorded (at 1 kHz) while they listened over Sennheiser HD 280 PRO headphones (Wedemark, Germany) to the sentence "Click on the word ____," ending with one of the displayed words (e.g., *admiral*) as the target word. Displays consisted of the target word (e.g., *admiral*), a competitor word (*admiration*), and two unrelated distractor words (*converter, convergence*). Auditory materials were presented at a fixed comfortable listening level and were timed such that acoustic onset of the target word was always 1,800 ms after display onset. Participants' task was to click on the printed version of the word spoken at the end of the sentence. The display stayed on until a response was given or for 2 s after the auditory stimulus was completed. The response window ended 5 s after acoustic onset of the target word. The intertrial stimulus interval was 440 ms.

The main experiment comprised one practice block and four main blocks. The practice block consisted of two presentations of six practice sets. In each main block, each of the 24 critical sets and the eight filler sets were shown once in a randomized order. Within each block, half of the critical words selected as targets had primary lexical stress and half had secondary lexical stress on the first syllable. Target word selection was counterbalanced across participants. In the first block, auditorily presented target words on critical trials always came from the critical stress pairs (e.g., *admiral* or *admiration*). With each repetition of a set in a new block, each word from a set was equally likely to be the target. Responses to both items within a critical pair could thus be collected (Reinisch et al., 2010). The order of Blocks 2 to 4 was counterbalanced across participants. Presentation order within a block was always newly randomized. Positions of the words on the screen were pseudo-randomized such that targets and competitors occurred equally often in each position within each block. Drift corrections were completed before the main experiment and after every eighth trial.

Results

Eye fixation data for the first presentation of each critical word as target were analyzed if a correct response was given and no fixations occurred outside of the screen (i.e., 97.19% of all critical trials with first presentations were analyzed). The data were sampled down to 250 Hz. Correct responses were defined as responses given with the computer mouse within a Euclidian distance of 120 pixels from the center of the target word (99.3% overall accuracy). Fixations within 191 pixels of a word were categorized as being on that printed word. This threshold was the largest distance possible without counting a fixation as belonging to two words. Proportions of fixations on critical targets, competitors, and averaged-over unrelated distractors were calculated for each target stress type (i.e., primary vs. secondary) separately for time windows corresponding to first and second syllables. Time windows were shifted by 200 ms—an average estimated time to program an eye movement (e.g., Matin, Shao, & Boff, 1993). Figure 1 shows the results by stress type of the target word on critical trials. Vertical lines indicate (from left to right) the onset of the target word, the average offset of the first and second syllables, and the offset of the segmental overlap between pairs. This figure suggests that fixations on target and competitor words diverged after the first syllable when targets had primary lexical stress on their first syllables (left panel) but not until after the second syllable when targets had initial secondary lexical stress (right panel).

Figure 1. Fixation proportions over time to target (circles), competitor (triangles), and distractors (crosses) separated by whether the first syllable of the target has primary lexical stress (left panel) or secondary lexical stress (right panel). The vertical lines show (from left to right) the average onset of the target word, the average offset of the first and second syllables, and the average offset of the segmental overlap between pairs. All measures were shifted by 200 ms.



For the statistical analyses, fixation proportions were transformed to empirical logit (Dixon, 2008). Target preference was defined as the difference of log-transformed fixation proportions of targets and competitors. Competitor preference was calculated as difference in log-transformed fixation proportions of competitors and distractors. Mixed-effect models were fit to these two dependent measures using the maximum likelihood criterion. Analyses on target preference were conducted separately by stress type of the target. For analyses on competitor preference, stress type was a contrast-coded fixed factor (primary lexical stress = −0.5; secondary lexical stress = 0.5). Subjects and items were evaluated as random factors. Models evaluating competitor preference included a by-subject slope adjustment for stress type (Barr, Levy, Scheepers, & Tily, 2013). The p values were estimated using Satterthwaite approximations.

Analyses of target preference showed no such preference during the first syllable regardless of whether the targets had primary lexical stress ($\beta = 0.13$, $SE = 0.28$, $p = 0.64$) or secondary lexical stress ($\beta = -0.02$, $SE = 0.32$, $p = 0.96$). Targets were, however, fixated significantly more during the second syllable if they had primary lexical stress on their first syllable ($\beta = 0.69$, $SE = 0.30$, $p = 0.03$). This suggests that listeners used suprasegmental cues located in the first syllables of accentuated words with initial primary lexical stress online to recognize these words. No such target preference for this time window was found for targets with secondary lexical stress ($\beta = 0.02$, $SE = 0.39$, $p = 0.97$).

Analyses of competitor preference showed that the degree of competition was not modulated by stress type during the first syllable ($\beta = 0.09$, $SE = 0.29$, $p = 0.75$) or during the second syllable ($\beta = 0.54$, $SE = 0.33$, $p = 0.11$). Words that were phonologically overlapping with the targets competed the same for recognition regardless of whether they had primary or secondary lexical stress on their first syllable. Only a general preference of the competitors over distractors was found for both time windows and all stress types (all $p < .0001$), indicating general phonological competition.

Discussion

In English, lexical stress is a viable cue to the identity of a word. English listeners rely mostly on segmental cues to stress but can also use suprasegmental cues to determine the stress pattern of a word—at least if the word is accentuated (e.g., Cooper et al., 2002; Mattys, 2000). Our study is the first to show that English listeners evaluate this suprasegmental information online, as soon as the information becomes available, during spoken-word recognition, and quickly enough to speed up recognition before segmental information disambiguates the target word. Specifically, suprasegmental information on the initial primary-stressed syllable of words with phrase-level prominence facilitates spoken-word recognition. Suprasegmental information about lexical stress thus modulates the time course of spoken-word recognition in English.

Suprasegmental information, however, did not affect the degree of lexical competition. This differs from previous findings for Dutch, where words with primary lexical stress were stronger competitors than words with secondary lexical stress (Reinisch et al., 2010). This difference between the two languages dovetails nicely with those found in cross-modal fragment priming, demonstrating that Dutch listeners are more efficient in using suprasegmental information to lexical stress for word recognition than English listeners (Cooper et al., 2002; van Donselaar et al., 2005). Together, these results suggest that English listeners weigh segmental cues more than suprasegmental cues and are therefore less effective than Dutch listeners in using suprasegmental cues to inhibit segmentally overlapping competing words.

Similar to the patterns of results observed in Dutch for the auditory and visual recognition of spoken words (Jesse & McQueen, 2013; Reinisch et al., 2010), English listeners used suprasegmental information only on syllables with primary lexical stress—not secondary stress—to facilitate spoken-word recognition. It is important to note that this pattern is not simply attributable to a bias toward favoring words with primary lexical stress on the first syllable (van Leyden & van Heuven, 1996) that might emerge from a prevalence of such words in the English lexicon (Cutler & Carter, 1987). Such a bias would have been reflected in a global preference of primary-stressed words irrespective of the stress pattern of the target word. Rather, as argued previously, our results further confirm that this pattern has at least partially a perceptual nature (cf. van Heuven & Menert, 1996). One possibility is that this asymmetry is driven by the influence of lexical stress on intelligibility: Stressed syllables are produced with more precise and extensive articulation (e.g., Scarborough, Keating, Mattys, Cho, & Alwan, 2009), are more informative about the identity of a word, and are more reliably recognized than unstressed syllables (e.g., Altman & Carter, 1989; Lieberman, 1963; McAllister, 1991; Mehta & Cutler, 1988). Another possible explanation for this asymmetry can be the placement of pitch accents on primary-stressed syllables. As mentioned in the introduction, the discussion on cross-linguistic differences in the use of suprasegmental cues to lexical stress largely has been limited to words with phrase-level prominence (e.g., Cooper et al., 2002; van Donselaar et al., 2005; Fry, 1958; Mattys, 2000; Reinisch et al., 2010). Critical words in these studies, and to facilitate comparisons in our study, have been recorded in isolation or in a focal position within a sentence. Primary-stressed syllables therefore received nuclear pitch accents, highlighting their salience to the listener (Bolinger, 1958). Phrase-level prominence enhances listeners' ability to make lexical stress distinctions in auditory and visual speech (Jesse & McQueen, 2013; van Heuven, 1988). In contrast, initial syllables with secondary lexical stress carried prenuclear pitch accents (e.g., Bolinger, 1958), which are acoustically different from nuclear pitch accents found on primary-stressed syllables. The suprasegmental cues provided by the secondary-stressed syllables were, however, not used by English listeners (and Dutch listeners in Reinisch et al., 2010) online during word recognition.

Given the fact that phrase-level prominence is associated with lexical stress in these studies, the question arises of how pitch accents can facilitate spoken-word recognition. One explanation is that listeners use only the suprasegmental information associated with the prominence of the word. Listeners may perceive the pitch movement on the syllable with a nuclear pitch accent and therefore interpret the syllable as carrying primary lexical stress, thereby facilitating word recognition. Another explanation is that pitch accents are additional cues to primary lexical stress. According to some phonological theories, lexical stress and prominence are a unitary concept, falling onto the same continuum of prominence (Chomsky & Halle, 1968; Liberman & Prince, 1977; Selkirk, 1984). In this view, the same suprasegmental parameters are modulated to indicate either type of emphasis but to different degrees. Other phonological accounts consider lexical stress and prominence as separate entities and autonomous phenomena (Bolinger, 1958; Gussenhoven, 1991; Shattuck-Hufnagel, Ostendorf, & Ross, 1994). In line with this view is that pitch movement may mark only prominence and not lexical stress (Beckman & Edwards, 1994; Plag et al., 2011; Sluijter & van Heuven, 1996). In relation to the present study, this could mean that listeners used either the suprasegmental cues to lexical stress for word recognition that were strengthened by accentuation or additional cues only available through the pitch accent.

The current findings cannot resolve these opposing views, but our acoustical analyses indicate the availability of pitch accents (i.e., significant pitch excursion differences between primary- and secondary-stressed syllables). The use of pitch accents would be consistent with an analysis of Cooper et al.'s (2002) data (Cutler, Wales, Cooper, & Janssen, 2007) showing that English listeners relied only on fundamental frequency (F0) to perceive stress patterns in online spoken-word recognition but that Dutch listeners used F0 and amplitude (and duration in Reinisch et al., 2010). This possibly suggests that English listeners may rely on pitch accents in online spoken-word recognition, whereas Dutch listeners can also utilize other suprasegmental cues directly related to lexical stress. More research clearly is needed before we can determine the contribution of these cues to lexical stress during spoken-word recognition. Studies on stress perception (that do not tap into online spoken-word recognition) have painted a mixed picture regarding the cues used and their relative weight assigned by English listeners in offline processing (e.g., Beckman, 1986; Chrabaszcz, Winn, Lin, & Idsardi, 2014; Fry, 1958; Howell, 1993; Lai, 2008; Mattys, 2000; Morton & Jassem, 1965; Zhang & Francis, 2010). However, the possibility that English listeners can use pitch accents during online processing is supported by results from other studies using the visual world paradigm showing that English listeners can capitalize on pitch accents online for reference resolution (e.g., Dahan, Tanenhaus, & Chambers, 2002; Ito & Speer, 2008).

Whatever role pitch accents may play, we can conclude that the suprasegmental information available in the first, primary-stressed syllables of accentuated words is utilized by English listeners immediately to speed up spoken-word recognition. This has clear implications for theories of spoken-word recognition (e.g., McClelland & Elman, 1986; Norris & McQueen, 2008) because there is a need to incorporate the use of suprasegmental prosodic information into these accounts. It is important to clarify when and how prosodic information is available during online processing for the various levels of language processing it informs (for an overview, see Cutler, Dahan, & van Donselaar, 1997). In agreement with previous findings (e.g., Dahan et al., 2002; Ito & Speer, 2008; Reinisch et al., 2010; Salverda et al., 2003), our results highlight the importance of prosodic information in spoken-word recognition. The collective evidence strongly suggests that prosodic information affects the early moments of spoken-word recognition, possibly through parallel analyses of segmental and suprasegmental cues (cf. Cho, McQueen, & Cox, 2007; Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Salverda et al., 2003). Lexical activation would thus be modulated immediately by both segmental and prosodic information from the input, with cues weighted depending on their relevance in the respective language and listening condition. In addition, listeners interpret this information as soon as it becomes available in relation to information from the preceding context (e.g., for use of context in interpreting prosodic information, see Brown, Salverda, Dilley, & Tanenhaus, 2015; Reinisch, Jesse, & McQueen, 2011b). Current models need to explain how both segmental and suprasegmental auditory speech information and visual speech information (e.g., see Jesse & McQueen, 2013) about prosody constrain the word recognition process.

In summary, we have shown that English listeners use suprasegmental information about lexical stress in accentuated words immediately during online spoken-word recognition. Words can thus be recognized by their stress pattern from stress-mismatching, but segmentally matching competitors before these competitors are ruled out later through segmental mismatch with the input. However, only the suprasegmental information available in primary-stressed syllables of the accentuated words can be exploited. Our findings further call for the need to incorporate the use of prosodic information in models of spoken-word recognition.

Acknowledgments

This work is supported by National Institutes of Health Grant R01-DC012300 (awarded to Ying-Yee Kong). The authors thank Michael Bartoli and Robert Moura for their assistance in collecting the data for this study and Anne Cutler for stimulating this work. Parts of this work were presented at the 169th Meeting of the Acoustical Society of America.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439. doi:10.1037/a0019901

Altman, G., & Carter, D. M. (1989). Lexical stress and lexical discriminability: Stressed syllables are more informative, but why? *Computer Speech & Language, 3,* 265–275. doi:10.1016/0885-2308(89)90022-3

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. doi:10.1016/j.jml.2012.11.001

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht, the Netherlands: Foris.

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes, 11*(1–2), 17–68. doi:10.1080/016909696387213

Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form* (pp. 1–33). Cambridge, United Kingdom: Cambridge University Press.

Bolinger, D. (1958). A theory of pitch accent in English. *Word, 14,* 109–149.

Bond, Z. S. (1981). Listening to elliptic speech: Pay attention to stressed vowels. *Journal of Phonetics, 9,* 89–96.

Braun, B., Lemhöfer, K., & Mani, N. (2011). Perceiving unstressed vowels in foreign-accented English. *The Journal of the Acoustical Society of America, 129,* 376–387. doi:10.1121/1.3500688

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. doi:10.1037/a0038689

Campbell, N., & Beckman, M. (1997). Stress, prominence and spectral tilt. In A. Botinis, G. Kouroupetroglou, & G. Crayiannis (Eds.), *Intonation: Theory, models and applications. Proceedings of an ESCA Workshop, Athens, Greece*. Baixas, France: European Speech Communication Association.

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically-driven detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics, 35,* 210–243. doi:10.1016/j.wocn.2006.03.003

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York, NY: Harper & Row.

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research, 57,* 1–12.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access. I. Adult data. *Journal of Memory and Language, 51,* 523–547. doi:10.1016/j.jml.2004.07.001

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. (1986). Forbear is not a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29,* 201–220.

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

Cutler, A., & Clifton, C., Jr. (1984). The use of prosodic information in word recognition. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183–196). Hillsdale, NJ: Erlbaum.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40,* 141–201.

Cutler, A., & Pasveer, D. (2006). Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. In R. Hoffman & H. Mixdorff (Eds.), *Proceedings of the Third International Conference on Speech Prosody* (pp. 250–254). Dresden, Germany: TUD Press.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetics Sciences* (pp. 1913–1916). Dudweiler, Germany: Pirrot.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47,* 292–314.

Davies, M. (2008). *The corpus of contemporary American English: 450 million words, 1990–present*. Retrieved from http://corpus.byu.edu/coca/

de Pijper, J. R. (1983). *Modelling British–English intonation*. Dordrecht, the Netherlands: Foris.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59,* 447–456. doi:10.1016/j.jml.2007.11.004

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America, 97,* 1893–1904.

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. doi:10.1162/089892904323057281

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America, 27,* 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1,* 126–152. doi:10.1177/002383095800100207

Gussenhoven, C. (1991). The English rhythm rule as an accent deletion rule. *Phonology, 8*(1), 1–35.

Gussenhoven, C. (2011). Sentential prominence in English. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 2780–2806). Malden, MA: Wiley-Blackwell.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies.* Chicago, IL: University of Chicago Press.

Howell, P. (1993). Cue trading in the production and perception of vowel stress. *The Journal of the Acoustical Society of America, 94,* 2063–2073.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language, 57,* 460–482.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58,* 541–573. doi:10.1016/j.jml.2007.06.013

Jesse, A., & McQueen, J. M. (2013). Suprasegmental lexical stress cues in visual speech can guide spoken-word recognition. *The Quarterly Journal of Experimental Psychology, 67,* 793–808. doi:10.1080/17470218.2013.834371

Lai, Y. (2008). *Acoustic realization and perception of English lexical stress by Mandarin learners* (Doctoral dissertation). University of Kansas, Lawrence.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8,* 249–336.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America, 32,* 451–454. doi:10.1121/1.1908095

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6,* 172–187.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36.

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

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics, 62,* 253–265.

Mattys, S. L., & Samuel, A. G. (2000). Implications of stress-pattern differences in spoken-word recognition. *Journal of Memory and Language, 42,* 571–596. doi:10.1006/jmla.1999.2696

McAllister, J. (1991). The processing of lexically stressed syllables in read and spontaneous speech. *Language and Speech, 34*(1), 1–26.

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

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

Mehta, G., & Cutler, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech, 31,* 135–156.

Morton, J., & Jassem, W. (1965). Acoustic correlates of stress. *Language and Speech, 8,* 159–181.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115,* 357–395. doi:10.1037/0033-295X.115.2.357

Plag, I., Kunter, G., & Schramm, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics, 39,* 362–374. doi:10.1016/j.wocn.2011.03.004

Poellmann, K., Mitterer, H., & McQueen, J. M. (2014). Use what you can: Storage, abstraction processes, and perceptual adjustments help listeners recognize reduced forms. *Frontiers in Psychology*. doi:10.3389/fpsyg.2014.00437

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. doi:10.1080/17470210903104412

Reinisch, E., Jesse, A., & McQueen, J. M. (2011a). Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech, 54,* 147–165. doi:10.1177/0023830910397489

Reinisch, E., Jesse, A., & McQueen, J. M. (2011b). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 978–996. doi:10.1037/a0021923

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90,* 51–89. doi:10.1016/S0010-0277(03)00139-2

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition, 105,* 466–476. doi:10.1016/j.cognition.2006.10.008

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1108–1117. doi:10.1037/a0019901

Scarborough, R., Keating, P., Mattys, S. L., Cho, T., & Alwan, A. (2009). Optical phonetics and visual perception of lexical and

phrasal stress in English. *Language and Speech, 52,* 135–175. doi:10.1177/0023830909103165
Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
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.
Sluijter, A. M., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America, 100,* 2471–2485.
Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45,* 412–432. doi:10.1006/jmla.2000.2783
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. doi:10.1016/j.jml.2011.08.001
van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology: Section A, 58,* 251–273. doi:10.1080/02724980343000927
van Heuven, V. J. (1988). Effects of stress and accent on the human recognition of word fragments in spoken context: Gating and shadowing. In W. A. Ainsworth & J. N. Holmes (Eds.), *Proceedings of Speech '88, 7th FASE Symposium* (pp. 811–818). Edinburgh, United Kingdom: Institute of Acoustics.
van Heuven, V. J., & Hagman, P. J. (1988). Lexical statistics and spoken word recognition in Dutch. In P. Coopmans & A. Hulk (Eds.), *Linguistics in the Netherlands* (pp. 59–69). Dordrecht, the Netherlands: Foris.
van Heuven, V. J., & Menert, L. (1996). Why stress position bias? *The Journal of the Acoustical Society of America, 100,* 2439–2451.
van Leyden, K., & van Heuven, V. J. (1996). Lexical stress and spoken word recognition: Dutch vs. English. In C. Cremers & M. den Dikken (Eds.), *Linguistics in the Netherlands* (Vol. 13, pp. 159–170). Amsterdam, the Netherlands: Benjamins.
Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics, 38,* 260–271. doi:10.1016/j.wocn.2009.11.002
Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32,* 25–64.

Appendix

Word pairs (N = 24) that contrast primary stress and secondary stress on the first syllable

Stress pair	Primary stress (initial syllable)	Word frequency	Pitch excursion (semitones/s)	Secondary stress (initial syllable)	Word frequency	Pitch excursion (semitones/s)
1	violate	728	13	violin	474	8
2	terminal	541	26	terminology	170	13
3	uniform	1,521	16	universal	1,613	7
4	parasol	13	27	parasitic	34	27
5	catapult	39	11	catastrophic	648	12
6	analog	101	10	analytic	26	12
7	continent	606	16	continuity	241	5
8	interval	74	18	intervene	791	15
9	paramount	436	32	paranoia	213	14
10	mediate	110	11	mediocre	153	10
11	allegory	18	5	allegation	1,019	9
12	intellect	194	13	integration	504	22
13	animal	3,570	12	animation	441	4
14	horrify	20	24	horizontal	123	5
15	diagram	100	11	diabetes	976	6
16	interface	103	30	interfere	552	16
17	orthodox	399	13	orthodontic	2	9
18	coroner	606	26	coronation	110	11
19	televise	40	37	telescopic	10	30
20	insulin	225	15	insulation	172	14
21	elephant	713	9	elevation	127	6
22	competent	615	18	competition	3,695	7
23	admiral	1,413	11	admiration	326	10
24	portable	329	26	portabella	12	9
Average		521	18		518	12

Note. Spoken word frequency is taken from Davies (2008). Pitch excursion is calculated as the range of fundamental frequency in semitones over the duration of the vowel in the first syllable and expressed in semitones per second.