Perceiving unstressed vowels in foreign-accented English

Bettina Braun[a)]
Max-Planck-Institute for Psycholinguistics, P.O. Box 310, NL-6500AH Nijmegen, The Netherlands

Kristin Lemhöfer
Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands

Nivedita Mani[b)]
Department of Developmental Science, University College London, 2 Wakefield Street, London WC1N1PF, United Kingdom

(Received 19 February 2010; revised 15 September 2010; accepted 17 September 2010)

This paper investigated how foreign-accented stress cues affect on-line speech comprehension in British speakers of English. While unstressed English vowels are usually reduced to /ə/, Dutch speakers of English only slightly centralize them. Speakers of both languages differentiate stress by suprasegmentals (duration and intensity). In a cross-modal priming experiment, English listeners heard sentences ending in monosyllabic prime fragments—produced by either an English or a Dutch speaker of English—and performed lexical decisions on visual targets. Primes were either stress-matching ("ab" excised from *absurd*), stress-mismatching ("ab" from *absence*), or unrelated ("pro" from *profound*) with respect to the target (e.g., ABSURD). Results showed a priming effect for stress-matching primes only when produced by the English speaker, suggesting that vowel quality is a more important cue to word stress than suprasegmental information. Furthermore, for visual targets with word-initial secondary stress that do not require vowel reduction (e.g., CAMPAIGN), resembling the Dutch way of realizing stress, there was a priming effect for both speakers. Hence, our data suggest that Dutch-accented English is not harder to understand *in general*, but it is in instances where the language-specific implementation of lexical stress differs across languages.
 [DOI: 10.1121/1.3500688]

PACS number(s): 43.71.Sy, 43.71.Es, 43.71.An, 43.71.Hw [MAH] Pages: 376–387

I. INTRODUCTION

In an era of globalization, it is not the exception anymore that people are confronted with foreign-accented speech. A number of factors influence the perceived severity of a nonnative speaker's foreign accent (e.g., native language of the speaker, age of acquisition, amount of exposure, phonetic similarity between native and nonnative language; Best, 1995; Flege, 1995; Broselow, 1999; Flege *et al*., 1999, Best *et al*., 2001). With the current status of English as the world's prime lingua franca, native English listeners especially, have to deal with a large variety of nonnative accents. Anecdotal evidence differentiating accents that are more or less difficult to understand is well documented. However, the specific phonetic aspects of accents that determine the ease of understanding remain, as yet, poorly understood.

Previous research indicates that native speakers are highly sensitive to the presence and strength of foreign accents (e.g., Flege, 1984; Anderson-Hsieh and Koehler, 1988; Magen, 1998). Furthermore, nonnative accents are—at least initially—harder to process than standard native speech. For instance, in a recent "artificial" foreign-accent study, Braun *et al*. (in press) manipulated the intonation contour of Dutch sentences to make them sound prosodically nonnative, while leaving their segmental and rhythmic structure intact. Using word monitoring and cross-modal priming techniques, they showed that sentence processing in native Dutch listeners was slowed down by the unfamiliar intonation contour, compared to the natural one. While the study by Braun *et al*. (in press) demonstrates the detrimental effect of intonational foreign accent, it is conceivable that deviations on the word level have similar or even stronger effects on on-line speech processing. In the present study, we will investigate how nonnative (Dutch) phonetic implementation of word stress affects spoken word recognition by English native speakers. However, before describing our experiment in more detail, we will review the aspects of word stress in English and Dutch that are relevant in the present context. In particular, we will focus on the acoustic correlates of different levels of lexical stress in English and the factors that influence stress perception.

Very generally, word stress is defined for each word in the mental lexicon. It is an abstract marker that makes one syllable more prominent than others. In English, word stress can distinguish between otherwise identical words, as in the verb *to record* compared to the noun *the record*. In contrast to the paradigmatic features vowel quality and lexical tone, lexical stress is primarily a syntagmatic feature. In other words, a given syllable does not usually have an absolute value for "strength" or "prominence," but a syllable may be

[a)]Author to whom correspondence should be addressed. Present address: University of Konstanz, Universitätsstr. 10, Fach 186, D-78467 Konstanz, Germany. Electronic mail: bettina.braun@uni-konstanz.de
[b)]Present address: Georg-August-Universität Göttingen, Gossler Str. 14, 37073 Göttingen, Germany.

stronger or weaker only in comparison with neighboring syllables. In the English word *gymnast*, for instance, the first syllable is strong, while the second is weak. In addition to this strong–weak distinction, the English and Dutch stress systems also have a paradigmatic aspect (e.g., Beckman and Edwards, 1994). Compare the English words *gymnast* and *tempest*, for instance. While both have a strong–weak sequence, the weak syllables in the two words differ in absolute strength. In *gymnast*, the weak syllable is produced with a full vowel, while in *tempest*, it is reduced to the central vowel schwa /ə/, which makes it weaker than a syllable with a full vowel (see Liberman and Prince, 1977; Trommelen and Zonneveld, 1999).

There are hence three levels of stress in English, often termed primary stress, secondary stress, and unstressed.[1] Each content word in the lexicon contains only one syllable that receives primary stress and which is the most prominent syllable of that word. (Note that some function words such as *the* and *an* do not normally contain primary stress—although they can attract it when highlighted.) If the respective word is accented at the utterance level, the pitch accent is aligned with the primary stressed syllable. In polysyllabic words, there may be one or more syllables with secondary stress assigned on rhythmic considerations and syllable weight. Secondary stressed syllables are somewhat less prominent than primary stressed syllables, but more prominent than unstressed syllables. In phonological terms, all stressed syllables are the head of phonological feet. Secondary stressed syllables may also receive an accent, for instance, to avoid an accent clash due to a close-by accent (e.g., accenting the first syllable of the word *Chinese* in *Chinese food*). Moreover, syllables with secondary stress are considered as nonreducible, meaning that these syllables maintain their full vowel quality. Unstressed syllables, on the other hand, never receive a pitch accent and are segmentally strongly reduced to /ə/ or /ɪ/[2] (e.g., Delattre, 1969). A complication arises in disyllabic words like *gymnast* above. The first syllable is described as secondary stressed by some authors (e.g., Beckman and Edwards, 1994; Trommelen and Zonneveld 1999) but as unstressed with an unreduced vowel by others (e.g., Fear *et al*., 1995). In this article, we do not offer data for adjudicating between one or the other. Terminologically, we will term syllables that are weaker than the primary stressed syllable but realized with a full vowel as secondary stressed. In other words, only syllables containing a schwa are referred to as unstressed.

Dutch phonology assumes the same three stress levels as English. In striking contrast to English, however, segmental reduction of unstressed syllables is not similarly obligatory in Dutch (e.g., Kager, 1989, page 275; Sluijter and van Heuven, 1996a) and dependent on the original vowel quality (/e/ being more prone to reduction than other vowels), the lexical frequency (more reduction in higher frequent words), and speech style (more reduction in less formal speech). In Dutch, as in English, the assignment of secondary stress is foremost based on rhythmic considerations and syllable weight (see Kager, 1989, pages 276–283).

Because the present study will investigate the perception of syllables with different stress levels in native and nonnative English, we will now review what is known on the acoustic characteristics that guide both word stress production and perception in English.

A. Acoustic cues to different degrees of stress in English

Most studies on the acoustic cues to English word stress have concentrated on the distinction between syllables carrying primary stress and unstressed syllables. In most of these studies, stress is confounded with accentuation, i.e., phrase-level prominence. If a target word is produced in isolation or in focus, primary stressed vowels are also accented (realized with a pitch accent) while unstressed vowels are not accompanied by such a pitch movement. Under such conditions, primary stressed vowels are characterized by increased f0, longer duration, higher intensity, and more peripheral articulation (Fry, 1955; Lieberman, 1960; Delattre, 1969; Nakatani *et al*., 1981; van Bergem, 1993; Lai, 2008, pages 22–46). For a mainly articulatory study, Beckman and Edwards (1994) recorded /ˈpapa/ in accented and unaccented (postnuclear) position in three different speech rates produced by two native speakers of English. They analyzed the first syllable of the target word (unaccented vs accented on the utterance level) and compared it to the second syllable in terms of syllable duration, as well as duration, displacement, and peak velocity of the lower-lip movement into the vowel. Their results showed that unstressed syllables had shorter durations than stressed syllables (regardless of their accentual status) and that their opening movement was smaller and slower.

There are also a number of studies investigating the acoustic cues that differentiate syllables with secondary stress from those with primary stress or unstressed syllables in English (Nakatani *et al*., 1981; Fear *et al*., 1995; Braun *et al*., 2008; Yuan *et al*., 2008). Fear *et al*. (1995), for instance, recorded five sets of word quadruples such as *audiences*, *auditoria*, *addition*, and *audition* in two different speaking rates by 12 native speakers of standard southern British English. The first syllables in these words were either primary stressed (*audiences*, henceforth P), secondary stressed (*auditoria*, S), unstressed and reduced (*addition*, R), or unstressed but unreduced (*audition*, U). Primary stressed syllables received phrase-level accent, and the other stress levels were unaccented. Acoustic measurements showed that duration was significantly different for all four categories ($P > S > U > R$), while intensity and spectral quality differentiated all categories except for primary and secondary stress ($P = S > U > R$ for intensity, $P = S < U < R$ for amount of centralization). One recent corpus study compared duration and f0 in a large number of primary stressed, secondary stressed, and unstressed (and reduced) vowels in English (Yuan *et al*., 2008). Based on linear regression models, they found that primary stressed vowels differed from secondary stressed and unstressed reduced vowels in f0 (possibly owing to the association between primary stress and phrase-level accentuation). However, in terms of duration, unstressed reduced vowels were shorter than both primary and secondary stressed vowels. In an orthogonally designed experiment involving both native and nonnative speakers of English, Braun *et al*. (2008) compared the spectral and suprasegmental differences between unstressed, reduced vowels (e.g., "ab" in "absurd"), and primary stressed vowels on the one hand (e.g., "ab" in "absence" so-called reduced set, comparable to the R- and P-groups in Fear *et al*., 1995) and between secondary and primary stressed vowels on the other (e.g., "campaign" vs "campus" so-called unreduced set, comparable to the U- and P-groups in Fear *et al*., 1995). Primary stressed syllables were accompanied by f0 movement, secondary stressed and unstressed syllables were not. Results showed that for native English speakers, unstressed reduced vowels were more centralized than primary stressed vowels, but there was no difference in spectral quality between primary and secondary stressed vowels (measured in terms of Euclidean distances in F1 and F2 in Bark from a speaker-specific schwa). Both sets yielded a main effect of stress on duration, i.e., primary stressed vowels were longer than secondary stressed and unstressed vowels. Spectral tilt was not affected by stress in neither set.

To conclude, primary stressed vowels in English are produced with longer duration, higher intensity, steeper spectral tilt, and more peripheral vowel quality than unstressed vowels. In most studies, secondary stressed vowels appear to group with unstressed (reduced) vowels (or at least do differ from primary stressed vowels) with respect to duration, but they group with primary stressed vowels when it comes to vowel quality.

B. Acoustic cues to different degrees of stress in Dutch and Dutch-accented English

Apart from the status of vowel quality in signaling the stressed–unstressed distinction, the acoustic cues to word stress are very similar in English and Dutch (van Bergem, 1993; Sluijter and van Heuven, 1996a, 1996b). As regards the production of secondary stress in Dutch, secondary stressed syllables are shorter than primary stressed syllables, but longer than unstressed syllables (Rietveld *et al*., 2004). We are not aware of any study comparing spectral quality in secondary stressed vowels to primary stressed or unstressed ones.

A recent production study (Braun *et al*., 2008) found that native speakers of Dutch apply their native way of implementing lexical stress (i.e., producing only slightly more centralized vowels) to the pronunciation of their second language (L2), English. In the reduced set, the initial syllable was either unstressed or primary stressed (*absurd-absence*). In the unreduced set, the initial syllable was either primary or secondary stressed (e.g., *campus-campaign*). Results showed that in the reduced set, Dutch speakers did not segmentally reduce the unstressed vowels of English words (like the vowel in the first syllable of *absurd*) as much as native English speakers did, while producing *stressed* vowels (like the first vowel in *absence*) less peripherally compared to the native group. They made more use of suprasegmentals (duration, spectral tilt) though, than English natives. In other words, the difference in the segmental characteristics (i.e., vowel quality) of primary stressed and unstressed syllables was less for Dutch speakers of English compared to native English speakers. In the unreduced set, on the other hand, Dutch speakers of English did not differ in how they signaled word stress, neither in spectral characteristics nor in suprasegmental features. Thus, while the difference between primary and secondary stress is signaled very similarly across languages (unreduced set), Dutch speakers implement the difference between stressed and unstressed reduced vowels more by ways of suprasegmental rather than by segmental features.

C. Perceptual cues to different degrees of stress in English

What are the primary *perceptual* cues to word stress in English? The answer to this question is not straightforward, owing again to the confound between word stress and accent. In a series of three perception experiments, Fry (1958) used synthesized stress minimal pairs such as *object* (noun or verb, depending on primary stress location) and systematically varied the suprasegmental cues f0, intensity, and duration (the acoustic correlates of pitch, loudness, and length). Both vowels were synthesized with a full vowel quality. On the basis of his results, Fry concluded that both duration and intensity were efficient in signaling a change in percept, but that duration was a somewhat stronger cue. Fundamental frequency strongly interacted with utterance intonation (pitch accent type) and outweighed duration as a stress cue (although f0 is an accentuation rather than a stress cue). Lai (2008, pages 68–98) tested the perception of word stress in resynthesized "dada" syllables by native English listeners and by beginning and advanced Mandarin Chinese learners of English using a stress detection task. English listeners were shown to be sensitive to changes in vowel quality, duration, and f0. In the case of conflicting cues (duration vs f0), listeners relied more strongly on duration than on f0 [which is at odds with the findings by Fry (1958)]. Unfortunately, Lai did not investigate the relative importance of segmental (vowel quality) vs suprasegmental (duration, f0) stress cues for native English listeners. In a cross-modal fragment priming study, Cooper *et al*. (2002) tested whether suprasegmental information is used in on-line word recognition. They used auditory word fragments (monosyllabic and disyllabic ones in two different experiments) which were segmentally ambiguous between a secondary or primary stressed syllable (e.g., "mu" taken from *music* or from *museum* or "admi" taken from *admiral* or *admiration*). These fragments hence differed in suprasegmental, but not in segmental stress cues. These primes, or unrelated control primes (e.g., "im" taken from *immerse*), preceded the presentation of visual targets (e.g., *music*, *admiration*), on which a lexical decision had to be made. Results showed stronger priming for stress-matching than for stress-mismatching auditory primes, relative to the unrelated control primes. For monosyllabic primes (Exp. 1a), the stress-mismatching condition did not even show a significant priming effect. In other words, whether or not a visual target was preactivated by the prime depended on the overlap of prime and target in terms of suprasegmental features, suggesting that suprasegmental stress cues are used during on-line speech recognition.

While these studies show that suprasegmental information is an important stress cue for native English listeners, a number of studies claim that the main cue to stress perception in English is vowel quality. Cutler and Clifton (1984), using a speeded semantic decision task (Exp. 3), reported that mis-stressing involving a change in vowel quality (from unstressed, i.e., reduced, to primary stressed or the other way round) had a more detrimental effect on word recognition latencies than mis-stressing without change in vowel quality (from secondary to primary stressed or the other way round). Fear *et al*. (1995) cross-spliced the first syllables of words such as *audiences*, *auditoria*, *addition*, and *audition* and had participants judge the naturalness of the resulting cross-spliced words on a scale from 1 to 5. Overall, judgments were most strongly influenced by vowel quality, followed by intensity and duration. Participants rated cross-splicings between primary stressed, secondary stressed, and unstressed unreduced vowels as identical to the original, but cross-splicings involving an unstressed reduced vowel were rated as significantly different from the other groups. These findings were interpreted as showing that English listeners were primarily sensitive to a change in vowel quality.

In summary, the picture that seems to emerge from the literature is that both types of cues, segmental and suprasegmental ones, are important in the perception of word stress. In particular, the importance of vowel quality as a cue to English stress seems to be undisputed. The present study aims at investigating problems in comprehension when word stress is signaled by suprasegmental stress cues but not by segmental ones, like in Dutch-accented English.

D. The present study

Given that Dutch speakers of English implement word stress in a different way from native speakers, particularly where vowel quality is concerned, the question arises what effect this has on English speakers' comprehension of Dutch-accented speech. Our focus will therefore be on the comparison of unstressed syllables, which are usually reduced to schwa in English (but not in Dutch), and primary stressed ones. If it is true that English listeners rely primarily on segmental cues like vowel quality, recognition of words containing unstressed (reduced) vowels should be significantly hampered, due to the Dutch tendency to *not* reduce these vowels. On the other hand, if English listeners are also sensitive to suprasegmental cues, the lack of vowel reduction in Dutch speakers might be compensated by the use of suprasegmental cues, possibly resulting in no comprehension problems at all.

Furthermore, we will include a condition comparing primary stressed with secondary stressed vowels to control for the effect of Dutch accent *per se*. Because the Dutch implementation of stress in these word pairs resembles the English pattern (only little or no vowel reduction and a similar degree of suprasegmental cues), having a Dutch speaker pronounce these words should be less detrimental to word recognition than might be the case for words with initial unstressed syllables.

Besides studying the effects of foreign-accented speech on comprehension, the perception of these different stress levels will allow us to substantiate prior studies on English listeners' reliance on suprasegmental and segmental cues to word stress. Since vowel reduction is a cue for the distinction between primary stressed and unstressed syllables in native English but not in Dutch-accented English, we can test directly whether English listeners rely only on vowel quality in recognition. If they rely similarly on suprasegmental cues, lacking vowel reduction in Dutch-accented English might not be detrimental to speech comprehension. In comparing primary stressed to secondary stressed vowels, we can test English listeners' use of suprasegmental cues when spectral information does not signal the degree of stress.

These issues will be investigated using the cross-modal fragment priming paradigm, a method to study word perception that has proved to be sensitive to manipulations of word stress (Soto-Faraco *et al*., 2001; Cooper *et al*., 2002; van Donselaar *et al*., 2005).

II. EXPERIMENT

In the present study, we used the cross-modal priming paradigm to investigate native English speakers' perception of Dutch-accented vs native English speech. Participants heard a spoken sentence ending in a one-syllable word fragment (e.g., *He didn't know the word "ab"*) and were subsequently shown a letter string presented visually on the computer screen (e.g., ABSURD[3]), on which they performed a lexical decision task. The fragment prime matched or mismatched the visual target ("ab" taken from *absurd* or from *absence*) or was unrelated to it ("pro" taken from *profound*). The carrier sentence and prime fragment were spoken by either a native English speaker or by a Dutch speaker of English. Because Dutch speakers do not reduce the vowel in the unstressed syllable as much as a native listener might expect, the fragment "ab" pronounced with a Dutch accent might not be as good a prime for the target ABSURD as when spoken by a native English speaker, resulting in longer lexical decision latencies for the target. Thus, we expect that stress-matching primes facilitate target recognition relative to the control prime condition, but this facilitation should be larger for primes spoken by an English speaker compared to those pronounced by a Dutch speaker.

The "unreduced" word set (included to control for the effect of Dutch accent *per se* and to study the use of suprasegmental stress cues when no segmental cues are available) contained words with initial secondary stress, for which there was only a phonetic vowel reduction or vowel centralization for Dutch speakers of English (e.g., *campaign* [ˌkæmˈpɛɪn]). Because the implementation of stress in these words is mainly characterized by suprasegmental features and resembles the Dutch one, having a Dutch speaker pronounce these words should be less detrimental to word recognition than what might be the case for words from the "reduced" set.

A. Participants

Eighty native speakers of British English, unaware of the purpose of the experiment, participated for a small fee. They had no self-reported hearing problems and normal or corrected-to-normal vision. Participants were recruited and tested in the United Kingdom to reduce the possibility of experience with Dutch-accented English. They were chosen from the subject pool at University College London with a mean age of 21.1 years (range: 18–36 years; 36 male, 44 female). Half of

TABLE I. Examples of the experimental conditions.

Word set	First syllable of target	Condition	Auditory prime (at end of carrier sentence)	Visual target	*N*
Reduced	Unstressed	Stress-match	/əb/	ABSURD	10
Reduced	Unstressed	Stress-mismatch	/ˈæb/	ABSURD	10
Reduced	Unstressed	Control	/ˈprɔ/ or /prə/	ABSURD	20
Reduced	Stressed	Stress-match	/ˈæb/	ABSENCE	10
Reduced	Stressed	Stress-mismatch	/əb/	ABSENCE	10
Reduced	Stressed	Control	/ˈprɔ/ or /prə/	ABSENCE	20
Unreduced	Unstressed	Stress-match	/ˈkæm/	CAMPAIGN	10
Unreduced	Unstressed	Stress-mismatch	/ˌkæm/	CAMPAIGN	10
Unreduced	Unstressed	Control	/ˈdɪ/ or /də/	CAMPAIGN	20
Unreduced	Stressed	Stress-match	/ˈkæm/	CAMPUS	10
Unreduced	Stressed	Stress-mismatch	/ˌkæm/	CAMPUS	10
Unreduced	Stressed	Control	/ˈdɪ/ or /də/	CAMPUS	20

the participants received the recordings spoken by the English native speaker as auditory stimuli, while the other half received the sentences spoken by the Dutch speaker.

B. Materials

1. Words

Forty disyllabic word pairs that differed in stress placement were chosen as visual targets. Twenty word pairs formed the "reduced set" in which the first syllable of the two words in a pair were orthographically identical, but the initial vowel contained a /ə/ when it was unstressed and a full vowel when it was stressed (e.g. *absurd-absence*, see Table IV in the Appendix for the full list).[4] The other half of these word pairs formed the "unreduced set" where the first syllable of both words in a pair also differed in stress placement, but this difference was not indicated by a vowel quality change, i.e., the first syllable in both words was phonemically identical (e.g., *campaign-campus*, see Table V in the Appendix for the full list). In both sets, to minimize coarticulatory differences, the first phoneme of the second syllable of the two words in a pair had the same place of articulation, and except for one pair also the same manner of articulation. The two members of each pair were chosen on the basis of maximum similarity in terms of lexical frequency, as well as number and frequency of cohort competitors (i.e., words that share the same first syllable and stress pattern). However, due to the structure of the English lexicon, initial unstressed syllables containing a /ə/ always had more competitors and therefore a higher competitor frequency than words that are stressed on the first syllable and contain a full vowel. Also, words with an initial unstressed syllable were more frequent than words with stress on the first syllable. The two members of a pair in the "unreduced" group were matched for number of competitors, but the member with primary stress on the first syllable had a higher cohort frequency than the one with primary stress on the second syllable. The lexical characteristics of the materials are summarized in Table VI in the Appendix.

The selected words were used as visual targets and combined with word fragment primes such that three experimental conditions were formed: Stress-matching prime, stress-mismatching prime, or unrelated prime. In the stress-matching prime condition, the first syllable of the target itself served as auditory prime fragment (e.g., /əb/ from *absurd* as prime for the target ABSURD). In the stress-mismatching condition, the first syllable of the other member of the word pair served as a prime (e.g., /ˈæb/ from *absence* as prime for ABSURD). Finally, in the unrelated condition, visual targets were preceded by syllable primes from a different word pair (e.g., /ˈprɔ/ from *profound* as prime for ABSURD). Half of these unrelated primes were excised from words stressed on the first syllable, the other half from words with an unstressed first syllable. Each participant saw one quarter of the critical targets in the stress-matched condition ($n = 20$, half from the unreduced and half from the reduced set), one quarter in the stress-mismatched conditions ($n = 20$), and half of the targets with unrelated primes ($n = 40$). An overview of the word and prime conditions with examples is given in Table I.

2. Nonwords and fillers

For use as visual targets requiring a "no" response in the lexical decision task, 40 nonword pairs were constructed with the same (presumed) stress and vowel reduction characteristics as the word targets (e.g., *stranique-stranning*; *bambeel: bambage*). These nonwords were created by combining the first and second syllables of existing words (e.g., *stranique*, combined from *strategic* and *unique*), using syllable combinations that were likely to result in a given stress pattern. In fact, the speakers did not have to be instructed about which stress pattern to use. None of the first syllables used in the nonwords occurred in the critical word conditions. As with the word targets, nonword targets were preceded by stress-matching, stress-mismatching, or unrelated primes to avoid any confound of prime-target overlap and required response.

To counter strategic responses, we reduced the proportion of phonologically related prime-target combinations within the experiment from 50% to 33.3% by including an additional set of unrelated fillers. This additional set consisted of the first syllables of 80 English words that had not been used in the conditions above as filler primes, as well as 80 additional visual targets (40 nonwords and 40 words). All of these prime-target combinations were phonologically unrelated (e.g., /æf/-MINGLE).

This totaled in 40 stress-matching trials (20 of which with words as targets and 20 with nonwords), 40 stress-mismatching trials, and 160 unrelated trials.

3. Carrier sentences

The fragment primes were embedded in semantically non-constraining carrier sentences (e.g., *The name of the ship was…, He couldn't spell the word…, The last word in the book was…*). The sentence frames contained only monosyllabic words or disyllabic words without phonological vowel reduction to avoid familiarizing the listeners with the critical aspect of Dutch-accented English (i.e., insufficient vowel reduction in unstressed syllables).[5] For the recording, each stimulus pair was assigned to two different sentence frames, so that sentence frames could be counterbalanced across participants and target pairs.

The sentence frames were recorded four times each, i.e., ending in all possible prime words for the given target pair. The recording always included the complete last words (e.g., *The name of the ship was "absurd"*), with the second syllable of the prime word removed from the recording afterwards, i.e., prior to presentation in the perception experiment.

4. Recording and acoustic analyses

The complete sentences were recorded by a female native speaker of southern British English and a female Dutch speaker of English, who had been chosen as a typical speaker of Dutch-accented English on the basis of prior speech signal analyses (see below). The English speaker was 38 years old and originated from London. The Dutch speaker was 22 years old and had been learning English for eight years at the time of recording. Self-ratings on foreign accent, amount of experience in English reading and speaking, as well as frequency of English usage are presented in Table VII (in the Appendix) to give an impression on her proficiency level. Sentences were recorded in a sound-attenuated cabin at the Max-Planck-Institute and were directly digitized onto a personal computer (sampling rate 44.1 kHz, 16 bit, stereo). Care was taken that primary word stress was placed on the correct syllable.

To confirm that the experimental materials chosen in the current experiment and produced by the two speakers did indeed display the typical differences in stress implementation already observed by Braun *et al*. (2008), we analyzed the segmental and suprasegmental characteristics of the vowels in the fragment primes. The vowels in the initial syllables were manually annotated using the PRAAT software package (Boersma and Weenik, 2009). For all monophthongs ($N = 18$ in the unreduced set, $N = 14$ in the reduced set), the frequency of the first two formants in Bark-scale (cf. Zwicker, 1961)[6] at the midpoint of the vowels were automatically extracted. The speaker-specific F1 and F2 values for /ə/ were estimated by averaging over five productions in the function word "the". Average F1 for the English speaker's /ə/ was 517.5 Hz compared to 452.3 Hz for the Dutch speaker. Average F2 for the English speaker's /ə/ was 1654.1 Hz compared to 1707.7 Hz for the Dutch speaker. Euclidean distances in Bark between F1 and F2 of each vowel and F1 and F2 of the speaker-specific /ə/ were calculated.

The English speaker completely elided /ə/ in the reduced set words *grenade*, *cravat*, *supply*, and *career*. For the remaining items, the average Euclidean distances from the speaker-specific /ə/ were subjected to a multilevel logistic regression model (see Baayen *et al*., 2008) with *Reduction type* (reduced or unreduced set), *Primary stress position* (initial or second syllable), and *Speaker* (English or Dutch) as fixed factors and *Item* as random factor. Results showed a significant three-way interaction between these factors ($p < 0.05$). The reduced and unreduced sets were subsequently analyzed separately to clarify the nature of the three-way interaction. In the reduced set, there were significant main effects of *Speaker* ($\beta = 0.32$, $p < 0.005$), *Primary Stress position* ($\beta = 1.12$, $p < 0.0001$), and a significant interaction between the two ($\beta = 0.61$, $p < 0.0001$). As expected, the Dutch speaker produced unstressed syllables in the reduced set (e.g., /əb/ in *absurd*) less centrally than the English speaker (average distance from speaker-specific /ə/ was 1.34 Bark for the Dutch speaker and 1.09 Bark for the English speaker, $p < 0.001$), while producing the stressed syllable in these pairs (e.g., /ˈæb/ in *absence*) less peripherally compared to the English speaker (average distance from speaker-specific /ə/ was 2.47 Bark for the Dutch speaker and 2.70 Bark for the English speaker, $p < 0.01$).[7] In the unreduced set, there were no main effects and no interaction (all p values > 0.3). Average distance from the speaker-specific /ə/ was 1.93 Bark for Dutch primary stressed vowels and 1.66 Bark for Dutch secondary stressed ones, compared to 2.30 Bark for English primary stressed vowels and 1.80 Bark for English secondary stressed ones). The mean values of F1 and F2 in the first vowel of the words in the reduced and unreduced sets for the two speakers are shown in Tables VIII and IX in the Appendix (for monophthongs only).

Furthermore, the suprasegmental features duration, intensity, and spectral tilt were analyzed in the same way as described above, with the same factors. For duration, the effect of *Speaker* approached significance ($\beta = 8.3$, $p = 0.054$). There was a strong effect *of Primary Stress position* ($\beta = 52.6$, $p < 0.0001$) and an interaction between *Speaker, Primary Stress position*, and *Reduction type* ($\beta = 206$, $p < 0.05$). Again, we analyzed reduced and unreduced words separately to investigate the three-way interaction. For the reduced group, there was a main effect of *Speaker* (the English speaker's vowels were on average 8.3 ms longer than the Dutch speaker's vowels) and a main effect of *Primary Stress position* (primary stressed vowels were on average 52.7 ms longer than unstressed vowels), but there was no interaction. For the unreduced group, there were effects of *Speaker* ($\beta = 12.5$, $p < 0.05$), *Primary Stress position* ($\beta = 42.6$, $p < 0.05$), and an interaction between the two ($\beta = 20.8$, $p < 0.05$). The Dutch speaker made a larger duration difference between primary and secondary stressed vowels than the English native (42.6 ms for the Dutch speaker compared to 21.8 ms for the English speaker).

For spectral tilt (energy in frequency band from 600 to 5000 Hz divided by energy in band from 0 to 600 Hz), there were main effects of *Speaker* ($\beta = 0.3$, $p < 0.05$), *Primary Stress position* ($\beta = 1.79$, $p < 0.0001$), *Reduction type* ($\beta = 1.27$, $p < 0.0005$), and an interaction between *Primary Stress position* and *Reduction type* ($\beta = 1.58$, $p < 0.005$). For the reduced group, the English speaker had a significantly steeper spectral tilt than the Dutch speaker ($\beta = 0.4$, $p < 0.005$) and unstressed vowels had a significantly steeper tilt than stressed vowels ($\beta = 1.75$, $p < 0.0001$). For the unreduced group, there was only an effect of *Speaker* ($\beta = 0.3$, $p < 0.05$), in the same direction as for the reduced group.

These acoustic analyses replicate the findings of Braun *et al*. (2008). With respect to suprasegmental stress cues, the Dutch speaker is comparable (in terms of spectral tilt) or even more pronounced (in terms of duration) than the English speaker in marking the different stress levels. With respect to vowel quality, the Dutch speaker does not differ from the English speaker in signaling the contrast between English primary and secondary stress (words in the unreduced set). In contrast, in the reduced set, stressed vowels are less peripheral in Dutch-accented English and unstressed vowels are less centralized than in native English.

C. Procedure

Participants were tested one-by-one in a quiet room. They were seated in front of a laptop computer with a 14-inch screen and wore headphones, through which auditory stimuli were presented in stereo. They were instructed that they would hear a sentence ending in a word fragment, followed by a letter string presented on the screen. They were asked to indicate as quickly and correctly as possible whether this letter string was an existing English word or not. Right-handed participants pressed the right button on a button box for a "word" response and the left button for a "nonword" response. The button box was reversed for left-handed participants, who received the reverse instruction. There was a practice block of five trials before the experiment proper began.

Each visual target appeared in the middle of the screen at the offset of the prime fragment (white lowercase letters in 72 pt Arial on black background) and remained on screen until the response was given, or until a timeout of 2000 ms had passed. Response latencies were measured in milliseconds relative to the appearance of the visual target. The next trial began after an intertrial interval of 600 ms. There were four blocks of 60 trials each, separated by a pause that the participant could end by pressing a button. The experiment was controlled with NESU (Nijmegen Experiment SetUp).

Prime condition (stress-match, stress-mismatch, control) and member of word pair (initial or final stress) were counterbalanced across participants, resulting in eight parallel lists. Participants were randomly assigned to one of these lists (five participants per list). Each participant saw every target word once (e.g., both *absence* and *absurd*) and received one member of each stimulus pair in a related prime condition (stress-match or stress-mismatch) and the other one in the control prime condition. The control primes were "unused" word fragments from the other conditions (i.e., word fragments that did not appear in a stress-match or stress-mismatch condition for the same participant). This way, the same set of primes contributed to the related and unrelated condition, avoiding artifacts due to material selection. The two members of each stimulus pair always occurred in different halves of the experiment and were separated by at least 20 trials. Randomization was restricted such that no more than four word or nonword targets and no more than four segmental-match or mismatch trials occurred in a row, to avoid response preparation effects.

III. RESULTS

Two participants in the English speaker condition were excluded because of high error rates (more than 25% errors). Furthermore, two items were excluded (in all conditions) because the target had multiple pronunciations (*access* and *polish*). Trials with the visual targets *brocade*, *ballast*, *innings*, and *oboe* were removed because they resulted in more than 35% errors. Furthermore, trials containing prime fragments from the words *grenade*, *cravat*, *supply*, and *career* were discarded, because, as mentioned above, the vowels in the first syllable were completely elided and therefore did not represent a mismatching prime. Reaction times (RTs) of the remaining trials were log-normalized and 23 trials with log-RTs larger than 7.5 (longer than 1800 ms) were removed as outliers after inspection of the density distribution. This left 5454 data points (85.2% of the overall data) for analysis.

A. Error analyses

The mean error rate was 2.8%. Correct and incorrect responses were subjected to a binomial logistic regression model (e.g., Baayen *et al*., 2008) with *Condition* (stress-match, stress-mismatch, or control), *Speaker* (English or Dutch), *Reduction type* (unreduced or reduced set), and *Primary stress position* of the target as fixed factors, as well as *Subject* and *Target word* as crossed random factors. Compared to averaged by-items and by-subjects analyses, mixed-effects modeling is more robust with respect to missing data. Results showed a main effect *of Reduction type* ($z = 2.30$, $p < 0.05$) and of *Condition* ($z = 2.28$, $p < 0.05$), but no interactions (all $p > 0.3$). Participants produced significantly less errors in unreduced set words (1.7%) than in reduced set words (4.1%; $\beta = -0.88$, $p < 0.05$). Furthermore, there were significantly less errors in stress-match trials (1.7%) than in control (3.1%, $\beta = -0.60$, $p < 0.05$) and stress-mismatch trials (3.3%, $\beta = -0.68$, $p < 0.05$).

B. Reaction time analyses

Reaction times for correct responses (5302 data points) were analyzed using a multilevel regression model with *Condition* (stress-match, stress-mismatch, or control), *Speaker* (English or Dutch), *Reduction type* (unreduced or reduced set), and *Primary stress position* of the target word (initial or second) as fixed factors (and interactions thereof), as well as *Subject* and *Target word* as crossed random factors. Furthermore, we included predictors that have previously been shown to affect lexical decision latencies, such as lexical frequency of the visual target, its number of characters, position of the trial in the experiment, log-RT to the preceding filler trial, as well as number and frequency of prime competitors. Predictors that were not significant at $p < 0.1$ were removed if this did not deteriorate the fit of the model (as estimated by the log-likelihood ratio, a measure of the predictive power of the statistical model). The most parsimonious model was refitted, and data points with residuals larger than 2.5 standard deviations were removed as outliers. Resulting *p*-values were

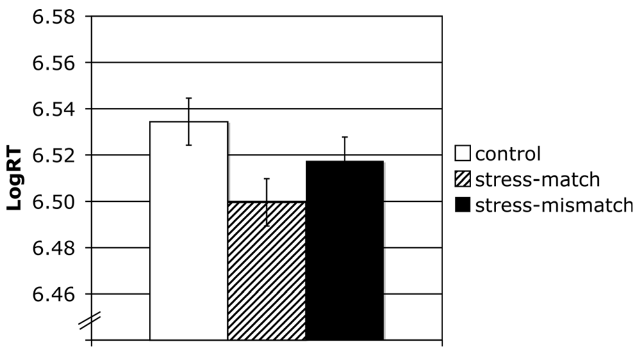

FIG. 1. Mean values and standard errors for trials in the unreduced set, computed for a median trial number of 124, and a mean log-RT to the preceding trial of 6.44.

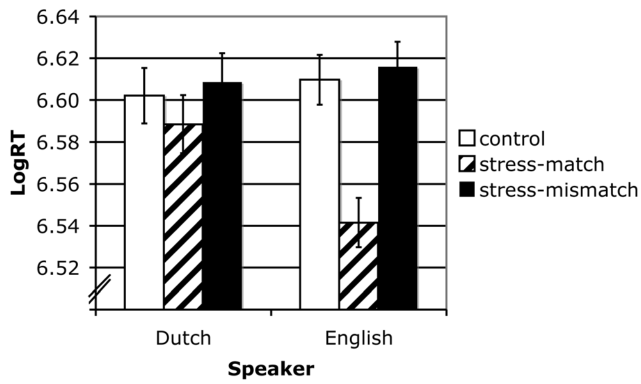

FIG. 2. Mean values and standard errors for trials in the reduced set and computed by the statistical model, computed for a median trial number of 124, a mean frequency of 5.5 and a mean log-RT to the preceding trial of 6.66 for the Dutch speaker and 6.60 for the English speaker.

estimated as the posterior probability of a Markov Chain Monte Carlo (MCMC) simulation with 10 000 runs.

Following this procedure, we removed number of characters, primary stress position of the visual target, as well as number and frequency of prime competitors from the model. Log-likelihood of the full model was 76.2 compared to 68.6 of the final model [$\chi^2(15) = 15, p > 0.4$].

In addition to expected effects of lexical frequency (the higher the lexical frequency, the faster the responses), position of the trial in the experiment (the earlier the trial, the faster the responses), and log-RT to the preceding filler trial (the faster the reaction time to the preceding filler trial, the faster the response in the actual experimental trial), results showed a significant three-way interaction between *Condition*, *Speaker*, and *Reduction type* [$F(1,5102) = 3.74, p < 0.05$]. There were no interactions between the control variables (frequency, position in the experiment, etc.), and the three critical factors (all p values > 0.2). In what follows, we describe separate analyses of the data in the reduced and unreduced set.

For the *unreduced* set—in which stress cues were primarily suprasegmental in both languages—there was a main effect of *Condition* ($p < 0.005$), but no effect of *Speaker* and no interaction (both $p > 0.5$; see Fig. 1). Responses in the stress-matching condition (6.50 on average which corresponds to 665 ms)[8] were significantly faster than in the control condition (6.53 on average, i.e., 688 ms, $p < 0.001$). There was, however, no difference in RTs between trials in the stress-mismatching (6.52 on average, i.e., 677 ms) and control conditions ($p > 0.1$) nor in RTs between trials in the stress-matching and the stress-mismatching condition ($p > 0.1$). Additional analyses showed that the same result holds for the two speakers individually. The estimates of the MCMC sampling, the upper and lower bounds, as well as the p-values are summarized in Table II.

TABLE II. Estimates, lower and upper bounds, and p-values based on a MCMC simulation with 1000 runs for trials from the unreduced set. The intercept is based on stress-matching trials.

	Mean estimate	Lower bound	Upper bound	p (MCMC)
Intercept (Stress-match)	5.5582	5.3313	5.7965	0.0001
Condition				
(Control)	0.0347	0.0147	0.0548	<0.001
(Stress-mismatch)	0.0181	−0.0062	0.0402	n.s. (0.13)
Position in experiment	−0.0002	−0.0004	−0.0001	0.0001
Previous log-RT	0.1493	0.1151	0.1822	0.0001

For the *reduced* set—in which stress cues differed across speaker—there was an interaction between *Condition* and *Speaker*, see Fig. 2. When the primes were produced by a native English speaker, RTs to stress-matching trials (6.54 on average, corresponding to 693 ms)[9] were over 50 ms shorter than RTs to both control (6.61 on average, corresponding to 743 ms, $p < 0.01$) and stress-mismatching trials (6.54 on average, corresponding to 747 ms, $p < 0.01$), while RTs in stress-mismatching and control trials did not differ from each other ($p > 0.6$). However, when the primes were produced by a Dutch speaker of English, there was no difference between stress-matching, stress-mismatching, and control trials (6.60 on average, i.e., 734 ms, $p > 0.4$; stress-match: 6.59 or 726.7 ms, stress-mismatch: 6.61 or 741.5 ms, control: 6.60 or 736.6 ms). The mean estimates of the MCMC sampling, the upper and lower bounds, as well as p-values for significant predictors and interactions are shown in Table III. Note that for this subset also there was no effect of stress position of the target word ($p > 0.5$) and no interactions with it ($p > 0.4$). Log-likelihood of the model including target word stress was

TABLE III. Estimates, lower and upper bounds, and p-values based on a MCMC simulation with 1000 runs for trials from the reduced set. The intercept is based on stress-matching trials from the Dutch speaker.

	Mean estimate	Lower bound	Upper bound	p (MCMC)
Intercept (Stress-match, Dutch)	5.5935	5.3177	5.8538	<0.0001
Condition				
(Control)	0.0141	−0.0186	0.0603	n.s. (0.4)
(Stress-mismatch)	0.0204	−0.0185	0.0603	n.s. (0.3)
Speaker (English)	−0.0414	−0.1078	0.0243	n.s. (0.3)
Frequency	−0.0109	−0.0198	−0.0022	<0.05
Position in experiment	−0.0002	−0.0004	−0.0001	<0.005
Previous log-RT	0.1540	0.1170	0.1926	<0.0001
Condition*speaker (control, English)	0.0540	0.0071	0.1001	<0.05
Condition*speaker (stress-mismatch, English)	0.0536	−0.0023	0.1083	<0.05

42.8 compared to 43.6 in the simpler model reported here [$\chi^2(6) = 1.6, p > 0.9$].

IV. CONCLUSIONS

The current study investigated factors contributing to the difficulty in understanding foreign-accented speech. Specifically, we examined the influence of improper phonetic stress implementation by nonnative speakers (i.e., when word stress is produced on the correct syllable but by the wrong acoustic means), while controlling for other more general effects of nonnative accent. Rather than investigating whether a foreign accent can be detected or how strongly it is perceived, this is one of the first studies to provide a direct examination of the extent to which a specific aspect of a foreign accent hampers speech perception.

Our results show that recognition of foreign-accented words by native English speakers suffers considerably when word stress is not implemented in an "English" way (i.e., by means of vowel quality). Unsurprisingly, a significant priming effect (i.e., shorter latencies to stress-matching compared to control primes) was found when the primes were produced by a native English speaker. Thus, the auditory matching prime preactivated the respective target and speeded up the subsequent processing of its printed form. However, when the primes were produced by a Dutch speaker of English, a significant priming effect was not found in the critical reduced set. That is, hearing the first syllable from words like *absurd* pronounced by a Dutch speaker did not aid the subsequent processing of *absurd* as a target. Significant priming of stress-matching primes pronounced by the Dutch speaker was only found in the unreduced set (e.g., *campaign*, comparing secondary stress vs primary stress), where differences in word stress were produced largely similarly across the two speakers. This replicates earlier findings that vowel quality is an important perceptual cue to the distinction between primary stressed and unstressed vowels in English (e.g., Cooper *et al.*, 2002).

For the unreduced set, as with the primes produced by English speakers, the results point to a graded priming effect with shorter latencies to stress-matching primes compared to stress-mismatching primes, which in turn had shorter latencies than the unrelated primes (replicating the findings of Cooper *et al.*, 2002). This suggests that suprasegmental information is also used to some extent in on-line speech recognition.

In summary, our results show that native English listeners had difficulties only with the Dutch way of producing English unstressed and primary stressed vowels (the reduced set), but not with the Dutch way of signaling English primary and secondary stress (the unreduced set). This suggests that Dutch-accented English was not harder to understand than native English *in general*, but only when the language-specific implementation of lexical stress differed across languages. While some previous studies report that native speakers of English are much more sensitive to segmental than to suprasegmental stress cues (e.g., Cutler and Clifton, 1984; Fear *et al.*, 1995), this study is the first to show that English listeners are hampered by the absence of segmental cues to lexical stress.[10] Clearly, the suprasegmental cues to word stress (duration, spectral tilt), which were used by the Dutch speaker to the same or even to a larger degree than by the native speaker, could not override the effect of vowel quality. Hence, these results suggest that spectral information is indeed vital for the perceptual distinction between unstressed and primary stressed vowels.

Since the latter conclusion stems from the difference in the priming effect between reduced and unreduced words, we now consider the direction of effects in the reduced set in more detail. Overall, we found that English listeners were hampered by the Dutch way of implementing stress in words in the reduced set. Notably, recognition of words with stress on the second syllable was equally affected as that of words with stress on the first syllable, i.e., *absurd* as well as *absence*. This suggests that Dutch speakers' insufficient reduction to /ə/ is as harmful as an improper quality in the full vowel. This symmetry in effects is surprising, given that the English /ə/ is often described as spectrally very variable (e.g., Koopmans-van Beinum, 1994; Flemming and Johnson, 2007), and even as a speech sound without a specific articulatory target (e.g., Browman and Goldstein, 1992) that is frequently and strongly assimilated to its consonantal and vocalic context [see Barry (1998), among others, for a target undershoot account of /ə/]. Despite its lack of articulatory and acoustic specificity, the English /ə/ appears to have a very specific *auditory* mental representation. As a consequence, a Dutch speaker's slightly centralized—but not fully /ə/-like—unstressed vowel failed to activate words with unstressed syllables in English, despite its pronounced suprasegmental reduction.

As stated in the Introduction, another vowel occurring in unstressed syllables is /ɪ/. Unlike /ə/, it preserves a full vowel quality and hence assumes a hybrid status in English (as it can occur in both stressed and unstressed syllables). In the current study, we focused on vowel quality differences and therefore included syllables containing the vowel /ɪ/ (e.g., *differ-define*) in the unreduced set. Future research will have to show if this vowel behaves differently than genuine full vowels.

Other than the Dutch speakers' production of vowel quality, English natives did not have greater difficulties understanding Dutch-accented English, relative to English produced by a native English speaker. This was surprising, since the speaker had a strong foreign accent (most easily recognizable in devoicing of voiced fricatives, incorrect th-articulation, and devoicing of final obstruents), despite being highly proficient in English. One explanation for this finding is that the other (segmental, rhythmic, intonational) characteristics of our speaker's nonnative accent do not seem to be as relevant for English listeners as vowel quality. Alternatively, it is possible that listeners used the preceding utterance context to tune into the characteristics of the Dutch speaker, i.e., segmental and rhythmic features (but recall that the carrier sentences did not contain reduced words, thereby preventing prior familiarization with the critical feature of vowel reduction). Taken together, the results of the current study suggest that one of the difficulties English listeners have in understanding Dutch-accented English concerns the Dutch use of schwa in implementing stress contrasts.

The findings, therefore, have important implications for the representation of stress information by English listeners and Dutch speakers of English. The English unstressed syllables we tested appear to be stored with the neutral vowel schwa. As a consequence, English listeners fail to recognize unstressed syllables that are not produced with this vowel (in the same way as they fail to recognize stressed vowels with an improper vowel quality) which will lead to a mismatch. Future research will have to address the difference between unstressed reduced and unstressed unreduced vowels (here grouped with secondary stressed syllables), especially when these do not differ in syllable weight. The mental representations seem to be different for Dutch speakers of English. Insufficient spectral reduction of this word-initial vowel in Dutch-accented English suggests that it is represented as a full vowel, which is then slightly reduced in unstressed positions, just as unstressed Dutch vowels are.

The present study shows, therefore, that language-specific stress implementation is an important factor in determining the intelligibility of foreign-accented speech. Conceivably, the reported difficulties in understanding foreign-accented speech generalize to other nonnative speaker groups whose native language does not make use of phonological vowel reduction such as German or French (Delattre, 1969). In practical terms, our results suggest that teaching English as a second language should lay special emphasis on the English way of implementing stress (in particular, the fact that unstressed vowels are reduced to schwa).

ACKNOWLEDGMENTS

Parts of this research were conducted while K.L. was at the Max-Planck-Institute for Psycholinguistics in Nijmegen. We thank the two anonymous reviewers for valuable comments on an earlier version, which helped us to improve the manuscript.

APPENDIX: MATERIALS, RATINGS, AND ACOUSTIC MEASURES

TABLE IV. Materials in the reduced set.

Primary stress on first syllable	Phonemic IPA transcription	Unstressed first syllable	Phonemic IPA transcription
absence	/ˈæb.səns/	absurd	/əb.ˈsɝd/
access	/ˈæk.səs/	accept	/ək.ˈsɛpt/
advent	/ˈæd.vənt/	advance	/əd.ˈvɑːns/
apple	/ˈæ.pɫ/	applause	/ə.ˈplɔːz/
ballast	/ˈbæ.ləst/	balloon	/bə.ˈluːn/
carriage	/ˈkæ.rIdʒ/	career	/kə.ˈrɪəɹ/
compound	/ˈkɔm.ˌpaʊnd/	complaint	/kəm.ˈpleɪnt/
convent	/ˈkɔn.vənt/	convey	/kən.ˈveɪ/
craven	/ˈkreɪ.vən/	cravat	/krə.ˈvæt/
fatal	/ˈfeɪ.tɫ/	fatigue	/fəˈtiːg/
gallop	/ˈgæ.ləp/	gazelle	/gə.ˈzɛɫ/
granny	/ˈgræ.ˌnɪ/	grenade	/grə.ˈneɪd/
matter	/ˈmæ.tɚ/	mature	/mə.ˈtjʊəɹ/
polish	/ˈpɔ.lɪʃ/	polite	/pə.ˈlaɪt/
proper	/ˈprɔ.pɚ/	propose	/prə.ˈpoʊz/
racket	/ˈræ.ˌkɪt/	raccoon	/rə.ˈkuːn/
substance	/ˈsʌb.stəns/	subscribe	/səb.ˈskraɪb/
supper	/ˈsʌ.pɚ/	supply	/sə.ˈplaɪ/
tonic	/ˈtɔ.ˌnɪk/	tonight	/tə.ˈnaɪt/
trapper	/ˈtræ.pɚ/	trapeze	/trə.ˈpiːz/

IPA, International Phonetics Association.

TABLE V. Materials in the unreduced set. Secondary stress marks in the IPA transcription of the last column are set in brackets to indicate that this annotation reflects our terminological decision to group spectrally unreduced syllables with secondary stressed ones.

Primary stress on first syllable	Phonemic IPA transcription	Secondary stress on first syllable	Phonemic IPA transcription
archives	/ˈɑː.ˌkaɪvz/	arcade	/(ˌ)ɑː.ˈkeɪd/
booking	/ˈbʊ.kɪŋ/	bouquet	/(ˌ)bʊ.ˈkeɪ/
broker	/ˈbroʊ.kɚ/	brocade	/(ˌ)broʊ.ˈkeɪd/
campus	/ˈkæm.pəs/	campaign	/(ˌ)kæm.ˈpeɪn/
sicken	/ˈsɪ.kən/	cigar	/(ˌ)sɪ.ˈgɑːɹ/
differ	/ˈdɪ.fɚ/	define	/(ˌ)dɪ.ˈfaɪn/
discount	/ˈdɪs.ˌkaʊnt/	discard	/(ˌ)dɪs.ˈkɑːd/
diver	/ˈdaɪ.vɚ/	diverse	/(ˌ)daɪ.ˈvɝs/
donor	/ˈdoʊ-nɚ/	donate	/(ˌ)doʊ.ˈneɪt/
humor	/ˈhjuː.mɚ/	humane	/ˌhjuː.ˈmeɪn/
image	/ˈɪ.mɪtʃ/	immense	/(ˌ)ɪ.ˈmɛns/
index	/ˈɪn.ˌdɛks/	induce	/(ˌ)ɪn.ˈdjuːs/
innings	/ˈɪ.nɪŋs/	inert	/(ˌ)ɪ.ˈnɝt/
mainly	/ˈmeɪn.lɪ/	maintain	/ˌmeɪn.ˈteɪn/
oboe	/ˈoʊ.ˌboʊ/	obese	/(ˌ)oʊ.ˈbiːs/
ordered	/ˈɔəɹ.dəd/	ordeal	/(ˌ)ɔəɹ.ˈdiːl/
pretty	/ˈprɪ.tɪ/	pretend	/(ˌ)prɪ.ˈtɛnd/
robot	/ˈroʊ.ˌbɔt/	robust	/(ˌ)roʊ.ˈbʌst/
rooted	/ˈru.tɪd/	routine	/(ˌ)ru.ˈtiːn/
transit	/ˈtræn.zɪt/	transcend	/ˌtræn.ˈsɛnd/

TABLE VI. Lexical characteristics of the materials: Mean lemma frequency in occurrences per million (o.p.m), number of cohorts competitors based on first syllable, summed frequency of cohort group, and number of characters. Standard deviations in brackets.

Primary stress position	Mean frequency in o.p.m.	Number of cohort competitors for the first syllable	Summed cohort frequency in o.p.m.	Number of characters
Reduced set				
First	30.6	128.7	14 103	6.85
Syllable	(31.8)	(270.6)	(12 029)	(0.93)
Second	42.2	187.4	30 751	6.50
Syllable	(52.1)	(143.3)	(66 413)	(1.05)
Unreduced set				
First	27.4	121.0	34 020	5.95
Syllable	(38.8)	(152.8)	(99 631)	(1.05)
Second	26.7	132.2	16 404	6.65
Syllable	(31.0)	(236.2)	(3316)	(0.99)

TABLE VII. Self-ratings of the Dutch speaker with respect to her experience with English.

	Ratings (1 very low/little; 7, very high/strong)
Frequency of reading English	5
Frequency of speaking English	2
Frequency of English TV/radio usage	5
Self rating of foreign accent	3
Amount of experience with reading English	6
Amount of experience writing English	6
Amount of experience speaking English	5

TABLE VIII. Mean F1 and F2 values for each initial monophthong for each speaker in the reduced set.

Primary stress on first syllable	Dutch F1 (Hz)	Dutch F2 (Hz)	English F1 (Hz)	English F2 (Hz)	Unstressed first syllable	Dutch F1 (Hz)	Dutch F2 (Hz)	English F1 (Hz)	English F2 (Hz)
absence	777	1810	932	1555	absurd	616	1721	481	1496
access	742	1896	935	1554	accept	562	1865	581	1809
advent	742	1830	848	1643	advance	623	1695	434	1878
apple	816	1828	735	1315	applause	615	1534	475	1650
ballast	742	1778	870	1398	balloon	453	1517	532	1575
carriage	628	1842	788	1387	career	419	1842		
compound	652	1136	612	934	complaint	344	1099	334	1389
convent	683	1177	693	1096	convey	498	1270	399	1518
craven					cravat				
fatal					fatigue				
gallop	773	1791	841	1655	gazelle	440	1890	425	1866
granny	673	1741	762	1305	grenade	442	1649		
matter	720	1766	634	1480	mature	404	1871	409	1732
polish	714	1188	642	1069	polite	608	1393	564	1554
proper	675	1259	605	1046	propose	444	1260	396	1426
racket	705	1701	908	1478	raccoon	561	1556	453	1199
substance	612	1364	663	1304	subscribe	411	1457	475	1669
supper	647	1367	730	1401	supply	395	1431		
tonic	745	1268	506	994	tonight	405	1818	414	1794
trapper	690	1604	860	1449	trapeze	460	1523	475	1539

TABLE IX. Mean F1 and F2 values for each initial monophthong for each speaker in the unreduced set.

Primary stress on first syllable	Dutch F1 (Hz)	Dutch F2 (Hz)	English F1 (Hz)	English F2 (Hz)	Secondary stress on first syllable	Dutch F1 (Hz)	Dutch F2 (Hz)	English F1 (Hz)	English F2 (Hz)
archives	740	1303	708	1192	arcade	701	1454	685	1256
booking	454	1125	451	1070	bouquet	420	1157	394	1566
broker					brocade				
campus	683	1836	884	1747	campaign	674	1835	630	1597
sicken	457	2202	470	2271	cigar	427	2141	435	1961
differ	468	2179	457	2240	define	391	2046	451	1692
discount	411	2228	405	1968	discard	416	1888	378	2015
diver					diverse				
donor					donate				
humor					humane				
image	445	2223	450	1180	immense	438	2254	484	2145
index	449	2174	434	2275	induce	418	2147	403	2248
innings	493	2046	452	2245	inert	497	2213	475	2274
mainly					maintain				
oboe					obese				
ordered	568	1126	421	655	ordeal	590	1301	397	767
pretty	479	1907	463	1797	pretend	399	2020	377	1815
robot					robust				
rooted	401	1385	413	1826	routine	418	1775	382	1783
transit	662	1800	598	1699	transcend	631	1653	584	1745

[1]This distinction between secondary stress in the second vowel of *gymnast* and an unstressed vowel in *tempest* is dependent on phonological theory and has not yet been tested empirically.

[2]In this paper, we are mostly concerned with unstressed syllables produced with the central vowel /ə/ in English and not with those produced with /ɪ/. Therefore, weak syllables with /ɪ/ will be treated as secondary stressed here, simply because they maintain a vowel quality other than schwa.

[3]Visual targets are highlighted with capitals in the text but were *not* shown with capitals during the experiment.

[4]The pair "granny-grenade" does not fulfill this constraint, but was later excluded for other reasons.

[5]The use of cliticized words such as *couldn't* is not critical, as an elided schwa does not give away how Dutch speakers implement phonological vowel reduction in English.

[6]$F1\_bark = (26.81 * F1\_Hz)/(1960 + F1\_Hz) - 0.53$.

[7]The less peripheral productions of English stressed vowels by the Dutch speaker were mainly due to the vowel target /æ/, which is often produced like a /ɛ/ by Dutch speakers.

[8]Mean values are based on the estimates from the statistical model and are calculated for the mean log-RT to the preceding filler trial (6.44) and the median trial number (124).

[9]Mean values are based on the estimates from the statistical model (median trial number of 124, mean frequency of 5.5, and mean log-RT to the preceding filler trial of 6.66 for the Dutch speaker and 6.60 for the English speaker).

[10]This study could not provide a direct proof of the obtained effects arising from stress (rather than vowel) misperception. Given the structure of the English lexicon, a dissociation of vowel quality and stress is not possible. However, previous studies have already shown that English listeners rely mostly on vowel quality when making explicit stress judgments (e.g., Exp3 in Cooper *et al.*, 2002).

Anderson-Hsieh, J., and Koehler, K. (**1988**). "The effect of foreign accent and speaking rate on native speaker comprehension," Lang. Learn. **38**, 561–613.

Baayen, H. R., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**, 390–412.

Barry, W. J. (**1998**). "Time as a factor in the acoustic variation of Schwa," in *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, pp. 3071–3074.

Beckman, M. E., and Edwards, J. (**1994**). "Articulatory evidence for differentiating stress categories," in *Laboratory Phonology III: Phonological Structure and Phonetic Form*, edited by P. A. Keating (Cambridge University Press, Cambridge), pp. 7–33.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 171–204.

Best, C. T., Mc Roberts, G. W., and Goodell, E. (**2001**). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system," J. Acoust. Soc. Am. **109**, 775–794.

Boersma, P., and Weenik, D. (**2009**). "PRAAT: Doing phonetics by computer (Version 5.1.05) [Computer program]." Retrieved May 1, 2009, from http://www.praat.org/#.

Braun, B., Lemhöfer, K., and Cutler, A. (**2008**). "English word stress as produced by English and Dutch speakers: The role of segmental and suprasegmental differences," in *Proceedings of Interspeech 2008*, pp. 1953–1953.

Braun, B., Dainora, A., and Ernestus, M. (**2010**). "An unfamiliar intonation contour slows down on-line speech comprehension," Lang. Cogn. Processes (in press).

Broselow, E. (**1999**). "Stress, epenthesis and segment transformation in Selayarese loans," in *Proceedings of the 25th Annual Conference of the Berkeley Linguistics Society*, edited by S. S. Chang, L. Liaw, and J. Ruppenhofer (Berkeley Linguistics Society, Berkeley, CA), pp. 211–225.

Browman, C., and Goldstein, L. (**1992**). "'Targetless' schwa: An articulatory analysis," in *Papers in Laboratory Phonology II*, edited by D. R. Ladd and G. Doherty (Cambridge University Press, Cambridge), pp. 26–67.

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

Cutler, A., and Clifton, C. E. (**1984**). "The use of prosodic information in word recognition," in *Attention and Performance X*, edited by H. Bouma and D. G. Bouwhuis (Erlbaum, Hillsdale, NJ), pp. 183–196.

Delattre, P. (**1969**). "An acoustic and articulatory study of vowel reduction in four languages," Int. Rev. Appl. Linguist. Lang. Teach. **7**, 294–325.

Fear, B. D., Cutler, A., and Butterfield, S. (**1995**). "The strong/weak syllable distinction in English," J. Acoust. Soc. Am. **97**, 1893–1904.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York, Timonium, MD), pp. 233–272.

Flege, J. E. (**1984**). "The detection of French accent by American listeners," J. Acoust. Soc. Am. **76**, 692–707.

Flege, J. E., MacKay, I. R. A., and Meador, D. (**1999**). "Native Italian speakers' perception and production of English vowels," J. Acoust. Soc. Am. **106**, 2973–2987.

Flemming, E., and Johnson, S. (**2007**). "Rosa's roses: Reduced vowels in American English," J. Int. Phonetic Assoc. **37**, 83–96.

Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.

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

Kager, R. (**1989**). *A Metrical Theory of Stress and Destressing in English and Dutch* (Foris Publications, Dordrecht), pp. 275–283.

Koopmans-van Beinum, F. J. (**1994**). "What's in a schwa? Durational and spectral analysis of natural continuous speech and diphones in Dutch," Phonetica **51**, 68–79.

Lai, Y. (**2008**). "Acoustic realization and perception of English lexical stress by Mandarin learners," Thesis, University of Kansas, Kansas, pp. 22–46, 68–98.

Lieberman, P. (**1960**). "Some acoustic correlates of word stress in American English," J. Acoust. Soc. Am. **22**, 451–454.

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

Magen, H., S. (**1998**). "The perception of foreign-accented speech," J. Phonetic **26**, 381–400.

Nakatani, L. H., O'Connor, K. D., and Aston, C. H. (**1981**). "Prosodic aspects of American English speech rhythm," Phonetica **38**, 84–105.

Rietveld, A. C. M., Kerkhoff, J., and Gussenhoven, C. (**2004**). "Word prosodic structure and vowel duration in Dutch," J. Phonetics **32**, 349–371.

Sluijter, A., M. C., and 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*, Philadelphia, pp. 630–633.

Sluijter, A. M. C., and Van Heuven, V. J. (**1996b**). "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **100**, 2471–2485.

Soto-Faraco, S., Sebastián-Gallés, N., and Cutler, A. (**2001**). "Segmental and suprasegmental mismatch in lexical access," J. Mem. Lang. **45**, 412–432.

Trommelen, M., and Zonneveld, W. (**1999**). "Word stress in West Germanic languages," in *Word Prosodic Systems in the Languages of Europe*, edited by H. v. d. Hulst (Mouton de Gruyter, Berlin), pp. 478–515.

van Bergem, D. R. (**1993**). "Acoustic vowel reduction as a function of sentence accent, word stress and word class," Speech Commun. **12**, 1–23.

van Donselaar, W., Koster, M., and Cutler, A. (**2005**). "Exploring the role of lexical stress in lexical recognition," Q. J. Exp. Psychol. **58A**, 251–273.

Yuan, J., Isard, S., and Liberman, M. (**2008**). "Different roles of pitch and duration in distinguishing word stress in English," in *Proceedings of the 9th Annual Conference of the International Speech Communication Association*, Brisbane, Australia, p. 885.

Zwicker, E. (**1961**). "Subdivision of the audible frequency range into critical bands," J. Acoust. Soc. Am. **33**, 248.