Cross-linguistic Influences on Sentence Accent Detection in Background Noise

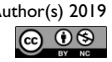
Odette Scharenborg, Sofoklis Kakouros, Brechtje Post, Fanny Meunier

Article

Language *and* Speech

Cross-linguistic Influences on Sentence Accent Detection in Background Noise

Language *and* Speech
2020, Vol. 63(1) 3–30


Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0023830918819573
journals.sagepub.com/home/las


Odette Scharenborg
Radboud University Nijmegen, The Netherlands

Sofoklis Kakouros
Radboud University Nijmegen, The Netherlands;
Aalto University, Finland

Brechtje Post
University of Cambridge, UK

Fanny Meunier
University Côte d'Azur, France

Abstract

This paper investigates whether sentence accent detection in a non-native language is dependent on (relative) similarity between prosodic cues to accent between the non-native and the native language, and whether cross-linguistic differences in the use of local and more widely distributed (i.e., non-local) cues to sentence accent detection lead to differential effects of the presence of background noise on sentence accent detection in a non-native language. We compared Dutch, Finnish, and French non-native listeners of English, whose cueing and use of prosodic prominence is gradually further removed from English, and compared their results on a phoneme monitoring task in different levels of noise and a quiet condition to those of native listeners. Overall phoneme detection performance was high for the native and the non-native listeners, but deteriorated to the same extent in the presence of background noise. Crucially, relative similarity between the prosodic cues to sentence accent of one's native language compared to that of a non-native language does not determine the ability to perceive and use sentence accent for speech perception in that non-native language. Moreover, proficiency in the non-native language is not a straightforward predictor of sentence accent perception performance, although high proficiency in a non-native language can seemingly overcome certain differences at the prosodic level between the native and non-native language. Instead, performance is determined by the extent to which listeners rely on

Corresponding author:
Odette Scharenborg, Multimedia Computing Group, Delft University of Technology, Van Mourik Broekmanweg 6, Delft 2628 XE, the Netherlands.
Email: o.e.scharenborg@tudelft.nl

local cues (English and Dutch) versus cues that are more distributed (Finnish and French), as more distributed cues survive the presence of background noise better.

Keywords
Sentence accent detection, phoneme detection, native and non-native listening, noise, acoustic cues, prosody, cross-linguistic influence

Introduction

Sentence accent plays an important role in speech comprehension (Akker & Cutler, 2003; Cutler, Dahan, & van Donselaar, 1997). For instance, compare the following two sentences, which consist of the same words but have different sentence accent (denoted by upper case), and consequently have a different meaning:

a. The GIRL was cleaning the table.
b. The girl was cleaning the TABLE.

Where in sentence *a* it is emphasized that it was the *girl*, rather than for instance, a boy, who was cleaning the table, in sentence *b* it is emphasized that the *table* was cleaned, and not some other object. Additional accent placements are also possible in the given example with each distinct placement conveying a different meaning to the listener. Sentence accent thus expresses semantic focus. Rapid and effective processing of accent placement in an utterance is highly important for the efficient comprehension of meaning (see for a review, Cutler et al., 1997) as it is pivotal in understanding the important parts of a speaker's message.

Throughout the literature, a number of terminological variants can be encountered that are used synonymously with the term *accent*. These include prominence, stress, emphasis, and prosodic focus (see also Wagner et al., 2015, for a discussion). In this work, we will use the overarching term "prominence" to refer to the percept of an element (e.g., a word) standing out from its context within a phrase or utterance (see, e.g., Terken & Hermes, 2000), whereas "accent" refers to the acoustic prosodic properties that tend to be associated with it. Prominence can be expressed in different ways, such as through a change in word order (Ladd, 1996; Vallduví, 1992; or probabilistically on the basis of listeners' prior experience with the language, such as, for example, operationalized by word frequency (Cole, Mo, & Hasegawa-Johnson, 2010). Acoustically, from the physical signal perspective, relative changes in energy, F0, duration, and spectral tilt have been found to correlate with the perception of prosodic prominence (e.g., Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; Fry, 1955, 1958; Lieberman, 1960; Sluijter & van Heuven, 1996; Venditti, Jun, & Beckman, 1996).

In optimal listening conditions, native listeners of a language are able to exploit the prosodic cues in the speech signal signaling upcoming sentence accent, which allows them to actively focus their attention to those parts of the sentence where accent will fall (for English native listeners, see Akker & Cutler, 2003). Non-native listeners, at least those with a high proficiency in the non-native language, have been shown to be able to detect sentence accent (native Mandarin Chinese listeners of English: Rosenberg, Hirschberg, & Manis, 2010; non-native listeners of German with several different native language backgrounds: Wagner, 2005) and to use similar acoustic prosodic cues as native listeners for accent detection (native Dutch listeners of English: Akker & Cutler, 2003; non-native listeners of German with several different native language backgrounds: Wagner,

2005). Nevertheless, high-proficiency non-native listeners display a reduced efficiency in using prosodic information signaling sentence accent for the processing of incoming speech (native Dutch listeners of English: Akker & Cutler, 2003). Moreover, differences in the operationalization of focus between a native (Basque) and non-native language (English) lead to increased difficulty of handling non-native accentual focus structures in perception (Garcia Lecumberri, 1995).

Research on accent detection has so far only been carried out in clean listening conditions, even though background noise is prevalent in everyday listening conditions (but see Carroll & Ruigendijk, 2016; Van Zyl & Hanekom, 2011, for native prosody perception in noise). The presence of background noise is known to negatively affect speech perception, with worse recognition performance for non-native listeners compared to native listeners (for a review, see Garcia Lecumberri, Cooke, & Cutler, 2010). This worse performance for non-native listeners can largely be explained by phonetic differences between the native and non-native languages (Scharenborg, Coumans, & van Hout, 2018) and differences in the amount of exposure to the (non)native language (Karaminis & Scharenborg, 2018). Moreover, there is accumulating evidence that non-native listeners use higher-level information less effectively, that is, information above the word level such as contextual linguistic or prosodic information, to compensate for loss of information at lower processing levels during speech recognition than native listeners (Bradlow & Alexander, 2007; Cutler, Garcia Lecumberri, & Cooke, 2008; Scharenborg, Kolkman, Kakouros, & Post, 2016). Consequently, accent detection in the presence of background noise might be more difficult for non-native listeners than for native listeners. Scharenborg and colleagues (Scharenborg et al., 2016) indeed found that the presence of background noise reduced native English and non-native listeners' (Dutch and Finnish listeners of English) ability to exploit sentence accent for speech processing. However, all listener groups were (still) able to use prosodic information signaling upcoming sentence accent, although the non-native listeners did so to a lesser extent than the native listeners.

The degrading effect of background noise is due to its masking of local, acoustic cues (e.g., Cooke, 2009). Consequently, accent-related prosodic cues that are (more) local, such as energy and duration, can be obscured. Several prosodic cues that correlate with accent, however, have more widely distributed spectral properties that relate to the perception of the prosodic cues (e.g., F0, tilt; referred to as "non-local" cues). For instance, F0 can preserve information through its harmonic components whereas tilt is based on the overall frequency distribution across the spectrum (that defines the slope of the spectrum). These non-local prosodic cues, such as fundamental frequency (F0), might be more robust to noise sources and are expected to better survive the degrading effect of background noise as cues may survive in different frequency regions (e.g., Garcia Lecumberri et al., 2010). Listeners from different language backgrounds use different prosodic cues to detect sentence accent, depending on the way accent is expressed in their native language (Garcia Lecumberri, 1995). Possibly, the relative similarity between the prosodic cues to accent in the three languages in the study by Scharenborg and colleagues (i.e., English, Dutch, and Finnish; Scharenborg et al., 2016) helped sentence accent perception for the non-native listeners. The finding that background noise impeded perception for native and non-native listeners in a similar way could then be due to the use of similar, more local prosodic cues in all three languages.

The current study investigates these possibilities by comparing sentence accent detection in a non-native language by non-native listeners whose native language is dissimilar in different degrees from English with respect to the use of prosodic cues for the cueing of sentence accent, that is, Dutch, Finnish, and French, thus extending the study by Scharenborg and colleagues (Scharenborg et al., 2016). Moreover, we investigate cross-linguistic differences in the precise role of local and non-local prosodic cues to accent detection. Specifically, we ask the questions:

1. whether sentence accent detection in a non-native language is dependent on (relative) similarity between prosodic cues to accent between the non-native and the native language;
2. whether cross-linguistic differences in the cueing of sentence accent lead to differential exploitation of prosodic information signaling sentence accent in the presence of background noise. We investigate this point by pulling apart the role of preceding prosodic cues and the role of accent on sentence accent detection.
3. whether any found cross-linguistic differences in the exploitation of prosodic information signaling sentence accent in the presence of background noise can be explained by cross-linguistic differences in the use of local and non-local prosodic cues to sentence accent.

The English-Dutch language pair allows us to investigate the influence of prosodic information on non-native spoken-word recognition without vital mismatches at the phonological level and with a reasonably small mismatch at the sound level, as Dutch and English prosodic cues and prosodic structures for sentence accent and prosodic processing are highly similar (see, for a detailed comparison, Akker & Cutler, 2003). The most important prosodic cues for accent expression and detection in Dutch and English are those of duration (particularly for Dutch), energy, and spectral tilt (spectral tilt has been primarily observed to correlate with Dutch accent; see Sluijter & van Heuven, 1996, but see also Campbell & Beckman, 1997; van Kuijk & Boves, 1999), which are arguably more local cues, with F0, an arguably non-local cue, being less important (Fry, 1955, 1958; Gussenhoven, 1983; Sluijter & van Heuven, 1996). In Finnish, the most important acoustic cue to accent is F0, whereas energy and duration are less important; instead, word order is an important cue for prominence (Vainio & Järvikivi, 2006). French and English both use prosodic prominence to convey focus, but it has been shown that in general, Germanic versus Romance languages differ in their use of prosody to encode focus (e.g., Cruttenden, 1997, 2006; Ladd, 1990, 1996, 2008; Lambrecht, 1994). In particular, it has been shown that whereas English employs prosodic prominence very often, and in a variety of different types of focus contexts (such as corrective, contrastive, or parallelism), French only uses prosodic prominence to encode focus in one particular type of context (i.e., corrective focus; see Jun & Fougeron, 2000). Such cross-language variations for prosodic prominence are well-documented in speech production studies but their repercussions on the comprehension side are poorly understood. Interestingly, at another prosodic level, French native speakers present a striking specificity as they have been described as "deaf" to lexical stress. They show strong difficulties to detect suprasegmental distinctions in word-level stress (Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Dupoux et al., 2008; Peperkamp & Dupoux, 2002). This effect even persists for French native speakers who are late learners of Spanish, where Spanish does use accent to contrast between words (Dupoux et al., 2008). In French, stress does not carry lexical information, but predictably falls on the word's final vowel. It has been suggested that consequently, French speakers do not (need to) process stress to identify lexical items, which is reflected as "deafness." The situation for relative phrasal prominence is somehow different as prominence in French is present but to a much lesser degree than English (see Frost, 2011, for a discussion). In perceptual studies, French listeners have been reported to rely more heavily on F0 to perceive accent than English listeners, and less on duration (Frost, 2011) and intensity (Séguinot, 1977) (arguably more "local" cues).

Sentence accent perception was investigated using a phoneme monitoring task (see e.g., Akker & Cutler, 2004; Cutler, 1976; Shields, McHugh, & Martin, 1974). As reviewed in Akker and Cutler, due to the rapid processing of accented syllables, word-initial phonemes are detected more quickly in words receiving sentence accent compared to words in an unaccented condition (Cutler & Foss, 1977; Shields et al., 1974). This rapid processing is on the one hand due to greater spectral clarity for words in accented position compared to words in an unaccented position (Koopmans-van

Table 1. The number of participants and their mean age and mean LexTale score per listener/language group.

Listener group	*N*	Age		LexTale		Fixed effect estimates		
		Mean	*SD*	Mean	*SD*	β	*SE*	*p*
English	47	20.8	2.7	98.6	2.6	98.560	1.257	<0.001
Dutch	42	23.0	4.2	84.5	11.2	−11.762	1.798	<0.001
Finnish	45	27.3	6.7	86.8	9.7	−18.279	1.975	<0.001
French	32	29.8	8.6	80.3	8.9	−14.018	1.830	<0.001

All participants had a LexTale score of minimally 63, which corresponds (roughly) to an upper intermediate or B2 level of proficiency (Lemhöfer & Broersma, 2012). The column "Fixed effect estimates" shows the results of the linear regression analysis with the English listener group on the intercept.
SD, standard deviation; *SE*, standard error; *p*, significant difference.

Beinum & van Bergem, 1989), and on the other hand due to the listener's ability to direct their attention to parts of the sentence that will receive sentence accent (Cutler, 1976; Cutler & Darwin, 1981). Given the general performance differences between native and non-native listeners, we expect a worse performance, that is, fewer detected target phonemes, for the non-native listener groups compared to the native listeners, especially at the harder levels of background noise. At the same time, if Finnish listeners are able to apply their native accent detection strategies, they would use F0, the cue that is expected to survive the background noise to a large extent, and consequently they might suffer less from the presence of background noise than the non-native Dutch listeners for whom F0 is not a primary cue to accent detection. Due to the "deafness" to prominence reported in the literature, and consequently having no native accent detection strategies which can be influenced by the presence of background noise, we expect the French listeners to show the least deterioration from the clean listening condition to the noisy background listening conditions.

2 Methods

2.1 Participants

Table 1 provides an overview of the four listener groups with their mean age (with *SD*). General English proficiency was assessed using the standardized test of vocabulary knowledge, LexTale (Lemhöfer & Broersma, 2012). The mean LexTale scores (with *SDs*) are added to Table 1. Scores between 80% and 100% correspond to an "upper & lower advanced/proficient user" and a C1 and C2 level of proficiency according to the Common European Framework of Reference for Languages (note Lemhöfer and Broersma do not differentiate between C1 and C2 levels). All participants had a LexTale score of minimally 63 which corresponds (roughly) to an upper intermediate or B2 level of proficiency (Lemhöfer & Broersma, 2012). All non-native English participants were taught English in high school for minimally 6 years. The native English students were recruited from the University of Cambridge, United Kingdom; the Dutch students were recruited from the subject pool of the Radboud University Nijmegen, the Netherlands; the Finnish listeners were recruited from Aalto University, Finland; and the French participants were recruited from the University Côte d'Azur, France. The English and Finnish listeners were the same listeners as those from the previous study (Scharenborg et al., 2016), while the Dutch listeners partially overlapped with those from the same study with the inclusion of additional, newly recruited participants. The sample size of the French listeners was smaller than the other three listener groups due to a difficulty finding enough French listeners with a good proficiency in English. None of the participants reported a history of language,

speech, or hearing problems. The participants were paid for their participation. The difference on the LexTale task between the native and all non-native listener groups was significant according to a one-way analysis of variance (see Table 1 for the estimates of the fixed effect Language group; the English listener group is on the intercept). Subsequent regression analyses with the other language groups on the intercept showed that the Dutch and Finnish listener groups did not differ significantly from one another (respectively, 84.5 vs. 86.8; $p = 0.224$), whereas the LexTale score of the French listeners (80.3) was significantly lower than that of the Dutch ($\beta = 4.2023$, $SE = 2.023$, $p = 0.037$) and the Finnish ($\beta = 6.517$, $SE = 1.993$, $p = 0.001$) listener groups.

2.2 Materials

2.2.1 Target phonemes and sentences. For the phoneme monitoring task, 48 experimental sentences were constructed. This set was adapted and extended from the set of 24 experimental sentences created by Akker and Cutler (2003). Akker and Cutler (2003) used /b, d, g/ in their experiments with Dutch non-native listeners of English; however, /g/ only appears in loan words in Dutch and /b, g/ only in loan words in Finnish. Because /p, t, k/ appear in all four languages under consideration, these were chosen as target phonemes. Note, however, that the acoustic properties of the voiceless stops differ for the languages. In English, word-initial voiceless stops are aspirated and thus have a "long voicing lag" (voicing starts late after the release; Lisker & Abramson, 1964), while Dutch, Finnish, and French are highly similar in that they have a "short voicing lag," and they have no prevoicing. Dutch voiceless stops have little or no aspiration (van Alphen & Smits, 2004), whereas Finnish and French voiceless stops are entirely unaspirated (Lein, Kupisch, & van de Weijer, 2016; Suomi, Toivanen, & Ylitalo, 2008). Nevertheless, word-initial voiceless stops are similar in the three non-native languages, so differences in target phoneme detection rates are not likely to be due to differences in the production of word-initial stops between the languages.

Target phonemes always appeared word-initially and in syllables containing word stress. Target-bearing words were nouns consisting of up to four syllables (14 monosyllabic words, 24 bisyllabic words, eight trisyllabic words, and two four-syllabic words). The sonority of the syllable increased from the start of the syllable onwards in all targets (see, e.g., Clements, 1990; Gussenhoven & Jacobs, 2011), leading to a maximally salient onset of syllables. Because all target phonemes appear word-initially, and thus syllable-initially, syllable length is not expected to have an influence, because the onset of syllables (or words) attracts the attention of the listener (Gussenhoven & Jacobs, 2011). Target-words were not perfectly matched for lexical frequency as frequency has not been found to influence initial phoneme monitoring (Eimas & Nygaard, 1992; Foss, Harwood, & Blank, 1980).

Appendix A gives an overview of all 48 target-bearing words and the experimental sentences in which they appeared. The experimental sentences had a similar syntactic structure, were semantically unpredictable, and contained only one target phoneme per sentence (indicated on the computer screen before auditory presentation of each sentence). Target-bearing words could appear early or late in the sentence but always minimally four words from the start of the sentence. Examples of an early and late target phoneme position (indicated with a capital letter):

a. The owner of the Pawn shop checked the customer's items.
b. The actions of the crew led to the Test lab's evacuation.

An additional set of 48 filler distractor sentences was created (see Appendix B for an overview). This set was adapted and extended from the set of 24 distractor sentences created by Akker and Cutler (2003). Filler sentences were added to reduce listener's chances to predict the location of the sentence accent and of the target phoneme. To that end, the filler sentences had a similar structure

to the experimental sentences but in half of them (12) the target-bearing word had a different position in the sentence compared to the experimental sentences or the target phoneme appeared in a different type of word, while the other half of the filler sentences (12) did not contain a target phoneme. All sentences were recorded by a male native speaker of British English, using the front internal microphone on a Samson Zoom H2 recorder. All recordings were made at 44.1 kHz, 16 bit, stereo, in a quiet room.

2.2.2 Background noise. Four signal variants were used in the experiment for each recording: clean (no noise was added; original recording), and three levels of noise. Different noise types give differential results on (non-native) listening (Broersma & Scharenborg, 2010; Cooke, Garcia Lecumberri, & Barker, 2008; Cooke, Garcia Lecumberri, Scharenborg, & van Dommelen, 2010; Garcia Lecumberri & Cooke, 2006). One distinction often used to describe the type of masking by background noise is energetic versus informational masking (Shinn-Cunningham, 2008). Energetic masking occurs due to the direct interaction of the background noise and the speech signal at the periphery and in the same ear (Mattys et al., 2009), while informational masking is the effect of background noise after the effect of energetic masking has been taken into account, for example, when audible linguistic information from the masker interferes with perception of the target speech (e.g., Garcia Lecumberri et al., 2010; Mattys et al., 2009), and is known to have an effect on attentional resources and cognitive load. As we are primarily interested in the effect of masking of the acoustic cues on the uptake and use of prosodic cues we chose an energetic masker. Speech-shaped noise (SSN) is such a pure energetic masker, which is often used in research on the effect of background noise on speech processing (see for a review, Garcia Lecumberri et al., 2010; Scharenborg et al., 2018). It has a fixed spectrum and no significant temporal modulations (Cooke et al., 2010). The three standard noise ratios (*SNR*s) that were used were +5 dB, 0 dB, and −5 dB. The SSN noise was automatically added to all experimental and filler sentences using a PRAAT script (Boersma & Weenink, 2005). All sentences had 200 ms of leading and trailing SSN noise. A Hamming window was applied to the noise, with a fade in of 10 ms for the leading noise and a 10-ms fade-out for the trailing noise.

2.3 Prosodic contexts

To investigate the exploitation of prosodic information signaling sentence accent to aid speech perception in the presence of background noise, sentence accent was manipulated so that the target-bearing words could occur in two prosodic contexts. All sentences contained prosodic context preceding the target-bearing word signaling sentence accent on the upcoming target-bearing word; however, in the "deaccented" condition, the target-bearing word was in fact deaccented, that is, incongruent with the preceding context, whereas it was accented in the "accented" condition, that is, congruent with the preceding context. To create the two prosodic contexts all sentences were recorded several times with an early and a late focal sentence accent (reflecting narrow focus on the words in upper case), and subsequently manipulated. Three productions were used to create the stimuli:

a. The remains of the **CAMP** were found by the tiger hunter.
b. The remains of the **camp** were found by the TIGER hunter.
c. The remains of the **CAMP** were found by the tiger hunter.

Following the cross-splicing procedure used in Akker and Cutler (2003), for the deaccented condition, the target-bearing word (in bold) from sentence **b** was spliced into sentence **a**. For the

accented condition, the target-bearing word from sentence **c,** which is a different rendition of the same sentence as in **a,** was spliced into sentence **a**. This procedure ensured that both the deaccented and the accented conditions had identical prosodic information preceding the target-bearing words. Differences between the two conditions can thus only be attributed to absence or presence of sentence accent on the target-bearing word.

In addition to each experimental sentence being recorded in two prosodic contexts, all 48 experimental sentences were also recorded with prosody that did not signal (upcoming) sentence accent on the target-bearing word. These "prosodically neutral" sentences were used as a second type of filler sentence to further reduce the likelihood that listeners could predict the location of the target phoneme.

2.4 Experimental procedure

To counterbalance the three target phonemes, the two positions of the target-bearing word (early versus late) and the four listening conditions, 24 separate experimental lists were created. Each list contained all 48 experimental (equal eight experimental sentences for each of three target phonemes) and 48 distractor sentences. In each list, 12 experimental sentences were presented in each of the four background noise conditions. Within each set of 12 experimental sentences, the target phoneme, position of the target-bearing word, and the two prosodic contexts were evenly distributed. The filler sentences were distributed over the experimental lists following the same procedure.

To ensure that listeners processed the sentences for comprehension, and not just focused on detecting the target phoneme, participants were first instructed that they were participating in an experiment on sentence comprehension, and were told they would be tested on the content of the sentences after the experiment. During the experiment, they were asked to listen within a sentence for the presence of a target sound that was specified for each sentence separately. The target phoneme appeared as a printed letter on the screen for 1 s before auditory presentation of the stimulus sentence. Listeners were asked to press the space bar as fast as possible upon hearing the target phoneme, and were explicitly instructed to treat the letter as a sound, that is, letter "k" on the screen corresponded to sound /k/, and thus listeners should press the space bar also when the word in which the target sound /k/ appeared started with a "c" as, for example, in "camp." Participants were tested individually in a sound-proof booth. They were randomly assigned to one of the 24 experimental lists. Audio stimuli were presented binaurally through headphones. Participants were comfortably seated in front of a computer screen in a sound-proof booth.

After the experiment, participants were presented with 48 written sentences from the main task (equally sampled from all noise and prosodic conditions) in which one word was left blank, and had to indicate which of four alternative word choices they thought had appeared in the sentence in the main experiment. All four alternative word choices fit semantically into the sentence context. The word recognition task confirmed that both the native (47.8% correct, averaged over all noise and prosodic conditions) and the non-native (Dutch: 40.2% correct; Finnish: 42.0% correct; French: 45.2% correct) participants had indeed engaged with the experimental materials, although the non-native listeners did, unsurprisingly, worse on the task than the native listeners. The results for this task were somewhat low—although well above chance level (which is at 25%), probably due to the relative difficulty of the task:

1. Due to the presence of background noise, some sentences were harder to understand and listeners might have misrecognized the target word.
2. Over the course of the experiment, listeners heard 96 different sentences so listeners are likely to have forgotten the details of each sentence.
3. There was a lack of semantic information in the sentence to help in determining the correct alternative word choice.

In view of the relative difficulty of the task, it is reassuring that the performance of the non-native listeners was not that much weaker than that of the native listeners.

2.5 Acoustic measures for stimulus properties

To investigate whether cross-linguistic differences in the acoustic cues used for sentence accent detection influences (a) sentence accent detection and (b) the effect the presence of background noise has on this ability, four acoustic prosodic features were computed that are known to correlate with the occurrence of sentence accent in speech: (a) energy (e.g., Kochanski, Grabe, Coleman, & Rosner, 2005); (b) F0 (e.g., Terken, 1991); (c) spectral tilt (e.g., Sluijter & van Heuven, 1996); and (d) duration (e.g., Fry, 1955; Lieberman, 1960). A total of 96 speech recordings (48 from each of the two prosodic contexts) were used for the acoustic features' extraction.

Each signal was downsampled from 44.1 kHz to 8 kHz. All features except duration were computed using windows of 25 ms with a frame shift of 10 ms. Signal energy was computed from the time-domain signal according to Eq. (1) (where x is the input signal, w the length of the analysis window, t the current sample index, and τ the window shift; see, e.g., Kakouros & Räsänen, 2016); F0 was computed using the YAAPT pitch tracking algorithm (Zahorian & Hu, 2008); spectral tilt by computing the mel frequency cepstral coefficients and taking the first (C1) mel frequency cepstral coefficients (e.g., Kakouros, Räsänen, & Alku, 2017); and word duration was obtained from manual segmentations. All raw feature values were subsequently normalized:

1. Energy was logarithmically normalized as loudness perception is known to be on a logarithmic scale (see, e.g., Fletcher & Munson, 1933).
2. F0 was semitone-normalized (see, e.g., Fant & Kruckenberg, 2004) relative to the minimum F0 in each utterance according to Eq. (2) to account for intra-talker variation (where P denotes pitch in Hz, n the current frame, $Pmin$ the minimum F0 in the utterance, and P' the semitone normalized contour).
3. Tilt was exponentially normalized—in this case, the exponential function provides a near linear scaling of the tilt estimates to positive real numbers for ease of interpretation (see Figure 1 for an example of the computed features).

$$E(n) = \sum_{\tau=-\frac{w}{2}}^{\left(\frac{w}{2}\right)-1} \left|x(t+\tau)\right|^2 \tag{1}$$

$$P'(n) = 12 * log_2\left(P(n) / P_{\min}\right) \tag{2}$$

For all features, except word duration, four word-level statistical descriptors were computed for the target words: *mean, standard deviation, max*, and *range* (defined as the difference between the *max* and *min* during a word) (see, e.g., Kakouros & Räsänen, 2016). Note that statistical descriptors for individual words for duration cannot be computed, as for a given word there is a single value for the annotated duration (whereas during the same word there are several acoustic feature values). Words were selected as the units for evaluation of the statistical descriptors as they seem to represent the optimal level of analysis for prominence studies (see, e.g., Rosenberg & Hirschberg, 2009). For the energy measure, all four statistical descriptors were also computed over the

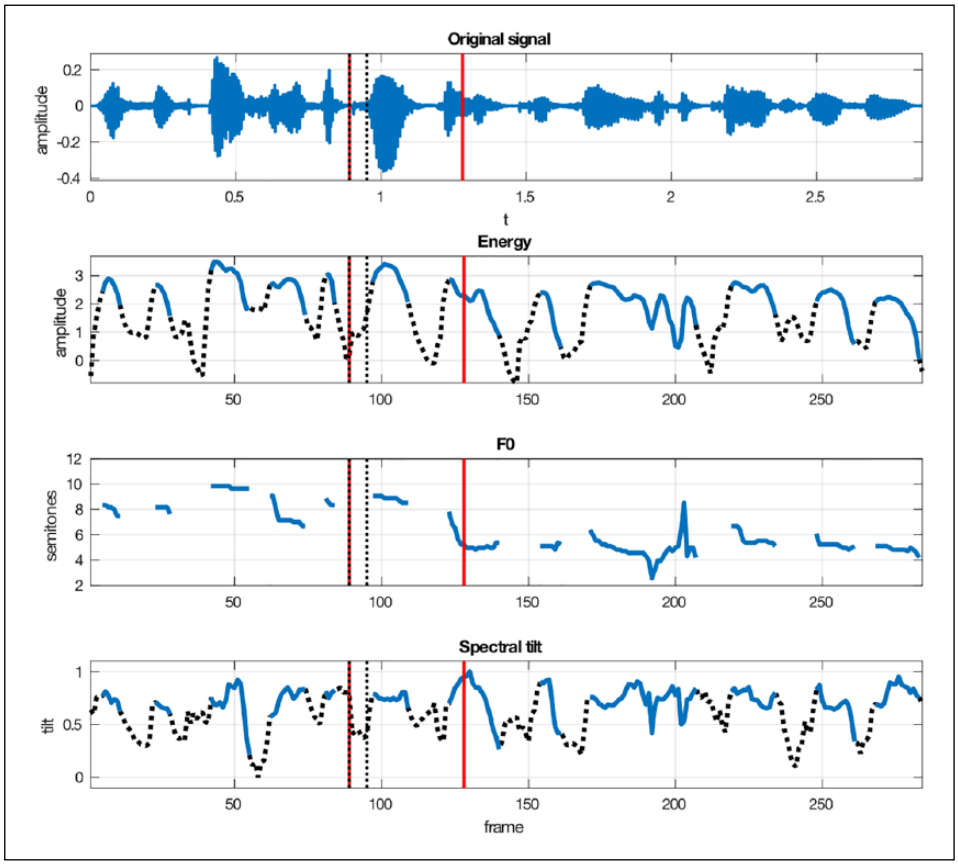


Figure 1. Original signal waveform, energy, F0, and spectral tilt for the deaccented condition of the sentence *The suspension of the /p/astor was confirmed at the college meeting*.
Vertical solid lines (red) demarcate the target word and vertical dashed lines (black) the target phoneme /p/. Horizontal dashed lines (black) indicate the feature values during unvoiced segments.
F0, fundamental frequency.

target-bearing phonemes. As all target phonemes are unvoiced plosives, spectral tilt and F0 were not relevant for the task and were not included in the analysis. Overall, as the *mean* and *max* descriptors provided large differences across all features, we only used these descriptors for the statistical analyses in the next section.

2.6 Acoustic analysis of the stimulus properties

To investigate acoustic cue use for sentence accent perception in native and non-native listening, we need to be able to tease apart the effects of language and noise on acoustic cue use in the two prosodic conditions. To that end, we compared the differences between the accented and deaccented condition for each acoustic parameter considered in our experimental setup. We first investigated possible differences in the acoustic features of (a) the target-bearing words (first analysis) and of (b) the three target phonemes (/p, t, k/; second analysis) between the two conditions (i.e., accented and deaccented). For the acoustic analysis our aim was to compare the differences

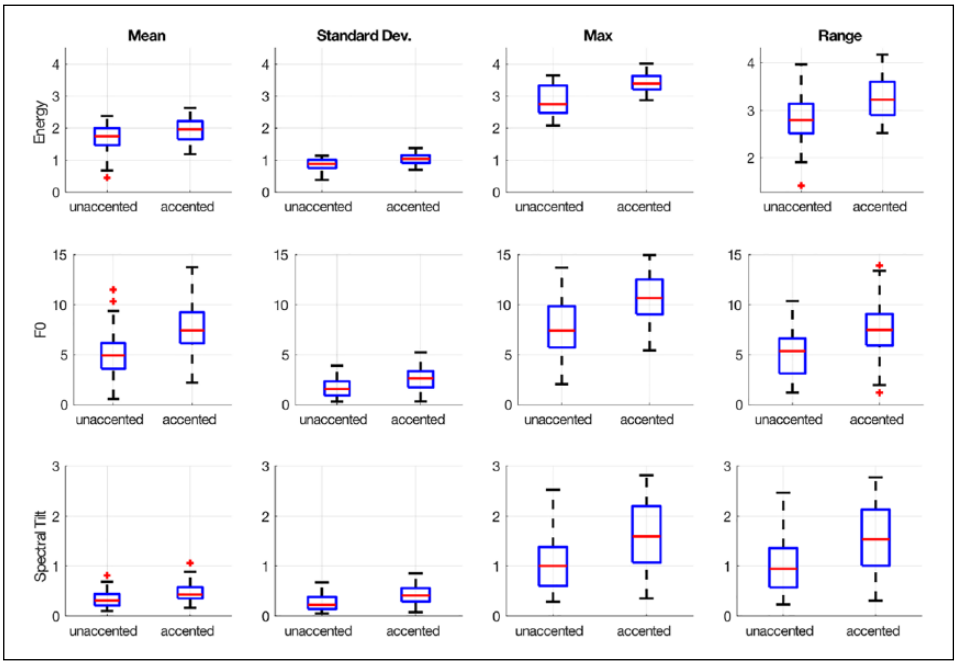


Figure 2. Boxplots for the target words for *mean, standard deviation, max, and range* of energy, F0, and spectral tilt.
Standard Dev., standard deviation; Max, maximum; F0, fundamental frequency.

between the accented and deaccented condition for each acoustic parameter considered in our experimental setup. As the normality assumption was not fulfilled for most of the examined acoustic parameters (tested using the Kolmogorov–Smirnov test) we selected a non-parametric test, the Wilcoxon rank-sum test, as an alternative. Cohen's d effect size was used to accompany the result of the rank-sum test but also to provide a way to exhibit the difference between the examined populations in a standardized scale. The d measure expresses the degree of separation of two populations with $d = 0$ indicating a complete overlap of the populations' distributions (no differences between the populations) and increasing magnitudes suggesting larger differences (e.g., $d > 0.8$ indicates a large difference; see Cohen, 1988).

For the first analysis, data were pooled over all target-bearing words within each prosodic condition. Similarly, in the second analysis, we also examined differences between the two prosodic conditions but in this case, instead of pooling the data together over all target words, data were pooled together over the individual phonemes (/p, t, k/). In addition, to ensure the correctness of our alignments (and therefore of the computed acoustic descriptors), we also examined the acoustic descriptors of the words surrounding the target-bearing word in the two prosodic conditions (as the annotations were carried out independently for each condition, the word boundaries might have slightly varied) and no significant differences were observed.

Figure 2 presents the results of the acoustic analysis of the target-bearing words in the two prosodic variants. The analysis indicated F0, energy, and tilt as the features with the largest differences for the target-bearing words across the two prosodic conditions. Durational differences, even though they were present, did not reach significance when compared to the rest of the features ($Z = -0.61$, $p = 0.55$, $d = 0.17$). The most significant differences were observed for max

Table 2. Effect size *d* for the differences between the two prosodic conditions for the word-level acoustic descriptors *max, mean, standard deviation*, and *range*, and for the target bearing-words.

	Maximum				Mean			
	Clean	+5 dB	0 dB	–5 dB	Clean	+5 dB	0 dB	−5 dB
Energy	**1.42*****	**0.80*****	**0.80*****	**0.64****	**0.62*****	0.29	**0.42***	0.16
F0	**1.32*****	**1.01*****	**0.99*****	0.35	**1.05*****	**0.67****	**0.56****	**0.03**
Spectral tilt	**0.84*****	**0.56****	**0.59****	**0.62****	**0.80*****	0.33	0.35	**0.41***
Duration	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
	SD				Range			
	Clean	+5 dB	0 dB	−5 dB	Clean	+5 dB	0 dB	−5 dB
Energy	**1.03*****	**0.59****	**0.50***	**0.51****	**0.93*****	**0.80*****	**0.59****	**0.59****
F0	**0.92*****	**0.83*****	**0.83*****	**0.73*****	**0.95*****	**0.71*****	**0.75*****	**0.69*****
Spectral tilt	**0.86*****	**0.52****	**0.55****	**0.52****	**0.85*****	**0.58****	**0.61****	**0.62****
Duration	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17

The results are presented for clean speech and for three noise levels (+5, 0, −5 dB *SNR*). Values in bold indicate statistically significant differences between the two prosodic conditions.
dB, decibel(s); *SD*, standard deviation; *SNR*, signal/noise ratio.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

energy ($Z = -5.21$, $p < 0.001$, $d = 1.42$), F0 ($Z = -5.54$, $p < 0.001$, $d = 1.33$), and spectral tilt ($Z = -4.18$, $p < 0.001$, $d = 0.94$) and also for the mean of the same features (energy: $Z = -2.77$, $p < 0.01$, $d = 0.62$; F0: $Z = -4.89$, $p < 0.001$, $d = 1.05$; spectral tilt: $Z = -3.84$, $p < 0.001$, $d = 0.80$) and range (energy: $Z = -4.10$, $p < 0.001$, $d = 0.94$; F0: $Z = -4.26$, $p < 0.001$, $d = 0.95$; spectral tilt: $Z = -4.15$, $p < 0.001$, $d = 0.94$) (see also Figure 2 for a more detailed presentation of the results). Overall, the target-bearing words in the accented condition carry more energy, have higher F0, and higher spectral tilt compared to the deaccented condition.

A further investigation of the acoustic differences between the two prosodic conditions, across the clean and noise corrupted versions of the recordings, revealed a gradual degradation of the separation *d* with increasing noise levels. For instance, for the *range* descriptor and energy *d* drops from 0.93 for clean speech to 0.59 for −5 dB *SNR* (see also Table 2 for a presentation of the effect sizes for the target words for the different signal variants for all statistical descriptors). Observing the separation (*d*) of the target words, it can be seen that, although substantially decreased, the overall separation between the two prosodic conditions remains statistically significant for energy, F0, and spectral tilt across most noise levels with, however, some variability that is dependent on the statistical measure utilized. For instance, although F0 has a large drop from $d = 1.326$ for clean speech to $d = 0.348$ for −5 *SNR* for the *max* descriptor, the respective change for the *range* descriptor is from $d = 0.948$ to 0.694. In all, acoustic differences, although reduced, are maintained across the different *SNR* levels for energy, F0, and spectral tilt. Finally, duration, as also observed earlier, has the lowest overall separation, with its effect size being non-significant for the target words between the two prosodic conditions. Duration is unaffected by the additive noise in our experiment as it is based on the annotated data and is therefore independent of the signal degradation applied on the speech recordings. Please note, the descriptors for clean signals were used in the statistical analyses below for sentence accent detection.

The second acoustic analysis, an evaluation of the two prosodic conditions over the target phonemes, did not reveal any significant differences between the conditions. These results are,

however, not unexpected, as all target phoneme categories (/k, p, t/) are unvoiced plosives, whereas effects relating to accentuation are typically observed at sonorant parts of the words that are typically found at the vowels of syllabic nuclei.

3 Results

3.1 Sentence accent detection

Two sets of statistical analyses were carried out. The first set of analyses investigates whether native and non-native listeners of English are able to exploit prosodic information signaling sentence accent to aid speech perception in the presence of background noise, and whether they use preceding prosodic cues and accent on sentence accent detection to do so. The second set of analyses investigates which acoustic cues are used to perform the task depending on native language and the impact of background noise on this use. To that end, logistic (binomial) regression statistical analyses on the target phoneme detections in the experimental sentences were carried out using generalized linear mixed-effect models (e.g., Baayen, Davidson, & Bates, 2008).

The dependent variable was the correct detection of a target phoneme (coded as 0 (incorrect) or 1 (correct)). To obtain the final, best-fitting model, we used a backward stepwise selection procedure, in which interactions and predictors that proved not significant at the 5% level were removed one-by-one from the model, and subsequent models were compared with their preceding one using the likelihood ratio test. The final model was selected by comparing Akaike information criterion (*AIC*) values on the basis of likelihood ratio tests and degrees of freedom (the number of factors). The model with the lowest *AIC* is the model that best fit the data (see e.g., Scharenborg, Weber, & Janse, 2014). Note, the model with the lowest *AIC* might contain non-significant factors, for example, when they are part of a significant interaction, and even non-significant interactions, which when removed result in a worse model. Fixed factors were Prosodic Condition (accented and deaccented, latter on the intercept), Listening Condition (clean on intercept, SNR +5, 0, −5 dB), and Language (English on intercept, Dutch, Finnish, and French; only in the first analysis), and their two-way and three-way interactions. We used Language as a factor rather than LexTale score as we were interested in differences between language groups rather than in the role of proficiency on sentence accent detection in background noise. Nevertheless, as proficiency in the non-native language might play a role in the uptake and use of specific acoustic cues, we entered the LexTale score (centered around 0 and normalized) as a fixed effect to the second set of analyses. Listening Condition was included as a nominal variable. Target-bearing Word, Target Phoneme, and Subject were entered as random factors. Random by-stimulus slopes and by-subject slopes for Listening Condition were added and tested through model comparisons in all analyses. In the second analysis, the acoustic features calculated at the target-bearing word level (centered around 0 and normalized) were additionally added as fixed factors and in interaction with Listening Condition, Language, and LexTale score: Energy_max, Energy_mean, Tilt_max, Tilt_mean, F0_max, F0_mean, Word Duration. Correlation scores between the max and mean of each acoustic measure are as follows: Energy_max and Energy_mean: Pearson $r = 0.62$, $p < 0.001$, Tilt_max and Tilt_mean: Pearson $r = 0.85$, $p < 0.001$, and F0_max and F0_mean: Pearson $r = 0.88$, $p < 0.001$.

3.1.1 Sentence accent detection in clean and noisy listening conditions. Figure 3 shows the average proportion of target phonemes that were correctly detected per listener group, for each of the listening conditions and prosodic conditions, plotted per language group. In Figure 3, the deaccented

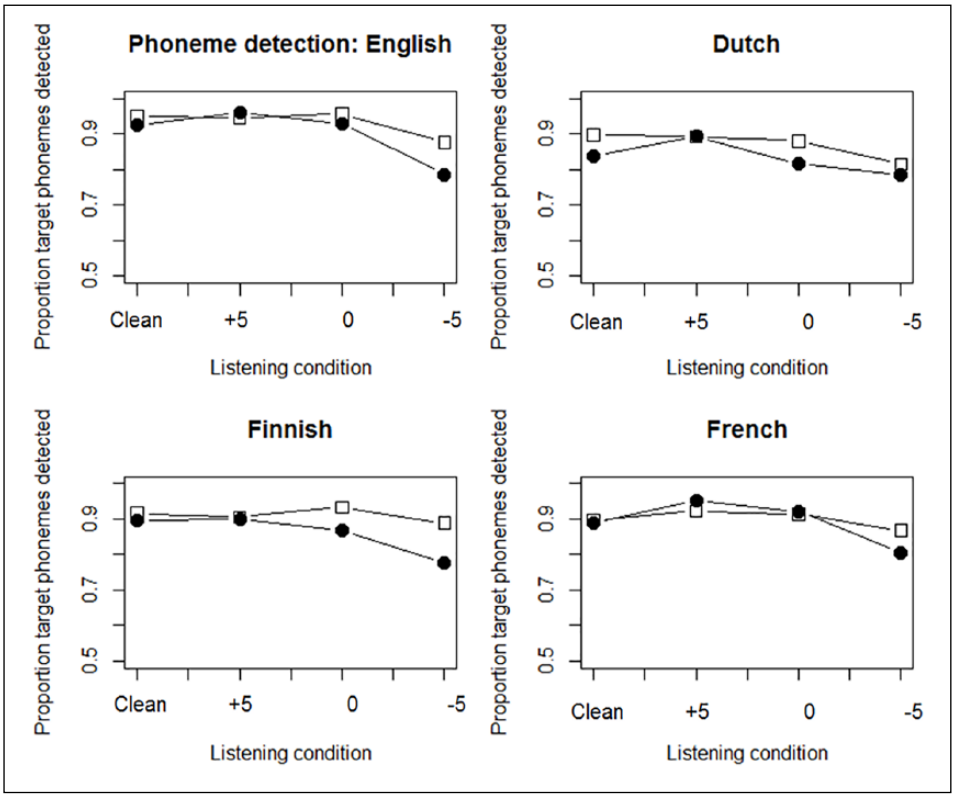


Figure 3. The proportion of detected target phonemes for the four listener groups and the four listening conditions, plotted per language group.
The deaccented condition is represented with the rounded markers and the accented condition with the squared markers.

condition is represented with the rounded markers, and the accented condition with the square markers. To show the effect of the presence of background noise, Figure 4 displays the same data but plotted per listening condition, aggregated over the two prosodic conditions. The results of the first statistical analysis are presented in Table 3, which shows the estimates of the fixed effects and their interactions in the best-fitting model for the phoneme detection results for all languages in all prosodic and listening conditions combined.

The results show a main effect of Language (English: 91.8% target phonemes detected, Finnish: 88.5%, French: 89.6%, Dutch: 85.3%); however, only the Dutch non-native listener group detected significantly fewer target phonemes than the English listener group (see also Figure 3 and Table 3 Language: Dutch). No significant difference was found between the other non-native listener groups and the English native listener group. We also observed a main effect of Listening Condition (90.4% target phonemes detected in the clean condition, *SNR* +5 dB: 92.2%, *SNR* 0 dB: 90.3%, *SNR* −5 dB: 82.5%) with significantly fewer target phonemes detected in the most difficult listening condition compared to the clean listening condition (Table 3, Listening Condition: *SNR* −5 dB; see also Figure 4). Finally, we found a main effect of Prosodic Condition with significantly more target phonemes detected in the accented condition compared to the deaccented condition (90.5% vs. 87.1%). The lack of an interaction between Language and Listening Condition showed that the detrimental effect of noise was similar for all listener groups.

Table 3. Fixed effect estimates for the best-fitting models of performance for the target phoneme detection accuracy analyses.

Fixed effect	β	*SE*	*p*
Intercept	1.904	0.313	**<0.001**
Prosodic condition	0.440	0.097	**<0.001**
Listening condition: *SNR* +5 dB	0.087	0.161	0.588
Listening condition: *SNR* 0 dB	−0.213	0.179	0.236
Listening condition: *SNR* −5 dB	−0.899	0.193	**<0.001**
Language: Finnish	−0.304	0.207	0.142
Language: French	−0.163	0.235	0.487
Language: Dutch	−0.525	0.209	**0.012**

$n = 5312$ observations. Bold indicates a significant effect.
SE, standard error; *p*, significant difference; *SNR*, signal/noise ratio; *dB*, decibel(s).

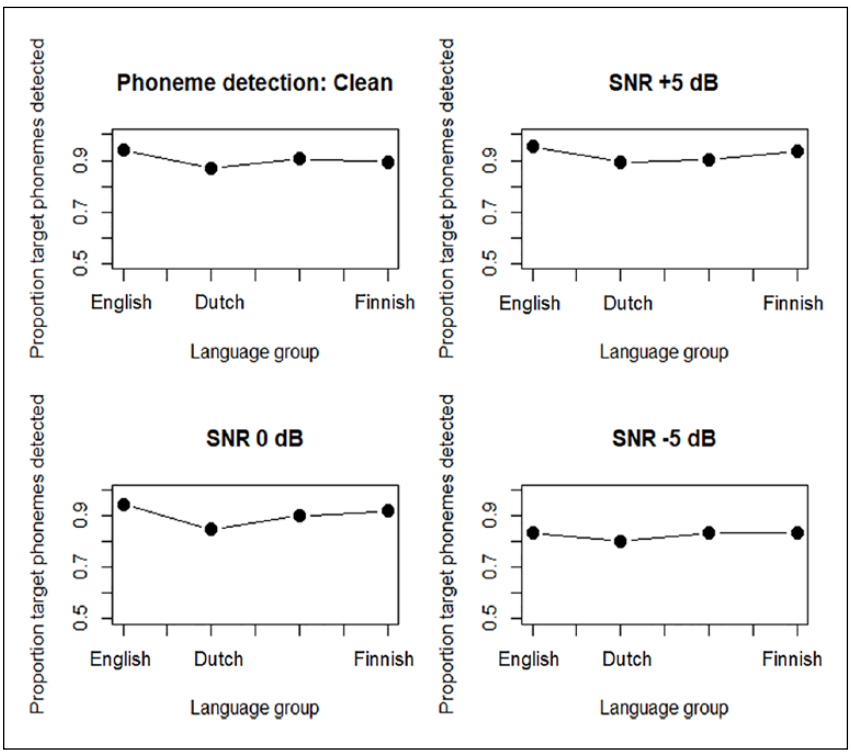


Figure 4. The proportion of detected target phonemes for the four listener groups plotted per listening condition, aggregated over the two prosodic conditions.
SNR, signal/noise ratio; dB, decibel(s).

Moreover, the analysis did not show an interaction between Language and Prosodic Condition, suggesting that the native and non-native listeners did not differ in their uptake of prosodic information signaling sentence accent. The random slope structure of the model included a target word random slope and a subject random slope for Listening Condition, indicating that target phoneme

Table 4. Fixed effect estimates for the best-fitting models of performance for the target phoneme detection accuracy analyses with acoustic measures, for all listener groups together.

Fixed effect	β	*SE*	*p*
Intercept	2.211	0.417	**<0.001**
Prosodic condition	0.286	0.142	**0.045**
Listening condition: *SNR* +5 dB	0.266	0.182	0.144
Listening condition: *SNR* 0 dB	−0.060	0.198	0.763
Listening condition: *SNR* −5 dB	−0.715	0.210	**<0.001**
Maximum energy	0.162	0.120	0.176
Language: Finnish	−0.368	0.212	0.082*
Language: French	−0.194	0.243	0.425
Language: Dutch	−0.517	0.217	**0.017**
Duration	−0.263	0.122	**0.031**
Listening condition: *SNR* +5 dB × maximum energy	−0.480	0.161	**0.003**
Listening condition: *SNR* 0 dB × maximum energy	−0.025	0.144	0.860
Listening condition: *SNR* −5 dB × maximum energy	0.227	0.139	0.103
Language: Finnish × duration	0.258	0.134	0.054*
Language: French × duration	0.302	0.167	0.070*
Language: Dutch × duration	−0.059	0.132	0.656

$n = 5162$. Bold indicates a significant effect; * indicates a marginal effect.
SE, standard error; *p*, significant difference; *SNR*, signal/noise ratio; *dB*, decibel(s).

detection decreased faster for some listeners and for some target words, than other listeners and target words for worse listening conditions.

In summary, the performance on the task was high for the native and the non-native listener groups. Only the Dutch non-native listeners detected significantly fewer target phonemes than the English native listener group. Both native and non-native listeners detected more target phonemes in the fully accented condition compared to the deaccented condition, that is, when not only the preceding context indicated upcoming sentence accent but when the target-bearing word also carried sentence accent. Moreover, the four listener groups did not differ in the extent to which they exploited prosodic information signaling sentence accent (as shown by the absence of a Language with Prosodic Condition interaction). The deteriorating effect of the presence of noise on sentence accent detection was also similar for all listener groups, and was only observed for the worst listening condition. Interestingly, although the proficiency of the French listeners was significantly lower than that of the other non-native language groups, their performance on the phoneme detection task was similar to that of the native English listener group. Possibly this finding can be explained by differences in the uptake and use of different sentence accent cues by the different listener groups. This is investigated in the analyses in the next section.

3.1.2 Cross-linguistic acoustic cue use. The second set of analyses investigated whether there are cross-linguistic differences in the specific acoustic cues used for sentence accent detection, and whether there are cross-linguistic differences on the impact of background noise on the acoustic cues used for sentence accent detection. To that end, the acoustic measures were entered as fixed factors and in interaction with Listening Condition and Prosodic Condition into the statistical analyses, first of all language groups together and then of each language separately.

Table 4 shows the estimates of the fixed effects and their interactions in the best-fitting model of the language analysis. As we are interested in the role of the different acoustic measures, we will focus on these in our discussion of the results. We found a main effect of Duration with longer

Table 5. Fixed effect estimates for the best-fitting models of performance for the target phoneme detection accuracy analyses with acoustic measures, for the native English listener group.

Fixed effect	β	*SE*	*p*
Intercept	3.006	0.268	**<0.001**
Listening condition: *SNR* +5 dB	0.351	0.348	0.313
Listening condition: *SNR* 0 dB	0.037	0.315	0.907
Listening condition: *SNR* −5 dB	−1.152	0.269	**<0.001**
Maximum energy	0.267	0.220	0.224
Listening condition: *SNR* +5 dB × maximum energy	−0.696	0.360	0.053*
Listening condition: *SNR* 0 dB × maximum energy	−0.095	0.303	0.754
Listening condition: *SNR* −5 dB × maximum energy	0.253	1.178	0.239

$n = 1472$. Bold indicates a significant effect; * indicates a marginal effect.
SE, standard error; *p*, significant difference; *SNR*, signal/noise ratio; *dB*, decibel(s).

duration resulting in significantly fewer detected phonemes. Moreover, for the *SNR* +5 dB condition, a higher Maximum Energy resulted in fewer detected target phonemes (see Listening Condition: *SNR* +5 dB × Maximum Energy in Table 4). The best model contained a non-significant interaction between Language and Duration. However, removing this interaction resulted in a significant drop in the *AIC* (model with interaction $AIC = 3332.2$; model without interaction $AIC = 3335.9$).

Tables 5–8 show the estimates of the fixed effects and their interactions in the best-fitting model for the analyses of each of the four languages separately. The analyses showed that for the English listeners, significantly fewer target phonemes were detected in the most difficult listening condition compared to the clean listening condition. The best model contained a non-significant interaction between Listening Condition and Maximum Energy. Removing the Listening Condition and Maximum Energy interaction resulted in a significant drop in the *AIC* (model with interaction $AIC = 783.9$; model without interaction $AIC = 789.0$) but also a significant main effect of Maximum Energy ($\beta = 0.295$, $SE = 0.111$, $p = 0.008$), suggesting that for the English listeners, a higher Maximum Energy resulted in more detected target phonemes. Note, however, that for the *SNR* +5 dB condition, a higher Maximum Energy led to the reverse pattern, that is, fewer detected target phonemes, albeit not significantly so (see Listening Condition: *SNR* +5 dB × maximum energy in Table 5). For the Dutch listeners, significantly fewer target phonemes were detected in the most difficult listening condition. The Dutch like the English also used energy for phoneme detection: Significantly more target phonemes were detected for target-bearing words with higher Mean Energy. Moreover it appears that Duration is also significant for Dutch: Significantly fewer target phonemes were detected when the target-bearing word had a longer duration.

For the Finnish listeners, significantly fewer target phonemes were detected in the most difficult listening condition, while significantly more target phonemes were detected with increasing Maximum F0. For the French listeners, the statistical analysis showed that three acoustic measures played a role in phoneme detection, all in interaction with proficiency in the non-native listeners: Maximum Energy, Mean F0, and Mean Tilt. For French listeners with a higher proficiency, higher Mean F0 and higher Maximum Energy led to more target phoneme detections, whereas for the French listeners with a lower proficiency, higher Mean F0 and higher Maximum Energy led to fewer target phoneme detections. For the French listeners with a higher proficiency, a higher mean spectral tilt led to fewer target phoneme detections whereas for listeners with a lower proficiency, a higher mean spectral tilt led to more target phoneme detections. Note that removing Listening Condition from the model resulted in a significant drop in the *AIC* (model with Listening Condition $AIC = 619.6$; model without Listening Condition $AIC = 624.3$).

Table 6. Fixed effect estimates for the best-fitting models of performance for the target phoneme detection accuracy analyses with acoustic measures, for the Dutch non-native listener group.

Fixed effect	β	*SE*	*p*
Intercept	2.596	0.397	**<0.001**
Listening condition: *SNR* +5 dB	0.299	0.317	0.345
Listening condition: *SNR* 0 dB	−0.399	0.361	0.269
Listening condition: *SNR* −5 dB	−0.874	0.404	**0.031**
Mean energy	0.201	0.088	**0.023**
Duration	−0.322	0.098	**<0.001**

$n = 1319$. Bold indicates a significant effect.
SE, standard error; *p*, significant difference; *SNR*, signal/noise ratio; *dB*, decibel(s).

Table 7. Fixed effect estimates for the best-fitting models of performance for the target phoneme detection accuracy analyses with acoustic measures, for the Finnish non-native listener group.

Fixed effect	β	*SE*	*p*
Intercept	2.646	0.292	**<0.001**
Listening condition: *SNR* +5 dB	−0.095	0.263	0.718
Listening condition: *SNR* 0 dB	−0.101	0.279	0.718
Listening condition: *SNR* −5 dB	−0.856	0.272	**0.002**
Maximum F0	0.0355	0.105	**<0.001**

$n = 1472$. Bold indicates a significant effect; * indicates a marginal effect.
SE, standard error; *p*, significant difference; *SNR*, signal/noise ratio; *dB*, decibel(s).

In summary, where the English and Dutch listeners primarily used energy, in addition to duration for Dutch listeners, the Finnish listeners only used F0, and the French listeners used a combination of energy, F0, and spectral tilt. In the worst background noise condition significantly fewer target phonemes were detected by the English, Dutch, and Finnish listeners, but no such effect of background noise was found for the French listeners. Proficiency in English was only found to be a predictor for the French listeners, and was related to the uptake and use of the acoustic features.

4 Discussion

This paper investigates cross-linguistic differences in sentence accent detection in the presence of background noise using a phoneme detection task. Specifically, we ask the questions whether:

1. Accent detection in a non-native language is dependent on (relative) similarity between prosodic cues to accent between the non-native and the native language.
2. whether cross-linguistic differences in the cueing of sentence accent lead to differential exploitation of prosodic information signaling sentence accent in the presence of background noise. We investigate this point by pulling apart the role of preceding prosodic cues and the role of accent on sentence accent detection.
3. whether any found cross-linguistic differences in the exploitation of prosodic information signaling sentence accent in the presence of background noise can be explained by cross-linguistic differences in the use of local and non-local prosodic cues to sentence accent.

Table 8. Fixed effect estimates for the best-fitting models of performance for the target phoneme detection accuracy analyses with acoustic measures, for the French non-native listeners.

Fixed effect	β	*SE*	*p*
Intercept	2.602	0.350	**<0.001**
Listening condition: *SNR* +5 dB	0.950	0.492	0.054*
Listening condition: *SNR* 0 dB	0.191	0.338	0.564
Listening condition: *SNR* −5 dB	−0.507	0.314	0.106
LexTale	0.225	0.230	0.328
Maximum energy	0.193	0.341	0.571
Maximum F0	0.107	0.466	0.818
Mean F0	0.101	0.391	0.797
Mean tilt	−0.518	0.371	0.162
LexTale × maximum energy	0.613	0.263	**0.020**
LexTale × mean F0	0.862	0.340	**0.011**
LexTale × mean tilt	−0.875	0.266	**0.001**

$n = 840$. Bold indicates a significant effect; * indicates a marginal effect.
SE, standard error; *p*, significant difference; *SNR*, signal/noise ratio; *dB*, decibel(s); LexTale, standardized test of vocabulary knowledge.

We compared Dutch, Finnish, and French non-native listeners of English, whose native language is dissimilar in different degrees from English with respect to the use of prosodic cues for the cueing of sentence accent, and compared their results to those of native English listeners.

In line with previous results obtained in clean listening conditions (e.g., Akker & Cutler, 2003; Rosenberg et al. 2010; Wagner, 2005), we found that native and non-native listeners of English are able to exploit prosodic information signaling sentence accent, with only a performance difference for the Dutch non-native listeners who detected significantly fewer target phonemes than the English listeners, despite having a proficiency level that did not differ significantly from that of the Finnish listeners. We extended the results of a previous study (Scharenborg et al., 2016) and showed that also French listeners who are highly proficient in the non-native language are surprisingly good at exploiting sentence accent for phoneme detection—even though French listeners are described to be "deaf" to prominence (e.g., Dupoux et al., 2001, 2008; Peperkamp & Dupoux, 2002). In fact, the French listeners' performance was not significantly worse than that of the English native listeners despite having an average proficiency level that was not only significantly lower than that of the English native listeners, but also significantly lower than that of the two non-native listener groups. These results show that proficiency in a language is not a straightforward predictor of sentence accent detection performance.

Overall, the native and non-native listeners were more accurate to detect a target phoneme when not only the preceding context indicated upcoming sentence accent but when the target-bearing word also carried sentence accent, that is, the listeners were more accurate in detecting a target phoneme in the accented condition compared to the deaccented condition. No differences in the use of global prosodic context information, that is, preceding context and sentence accent placement, for sentence accent detection were observed between the four listener groups, which suggests that global prosodic context in a non-native language is an equally robust cue for native and non-native listeners.

When background noise was added to the stimuli, phoneme detection accuracies decreased for the native and non-native listeners in the worst listening condition. This deteriorating effect of the presence of noise on sentence accent detection was similar for all listener groups. Note, however,

that the language-dependent analyses did not show a significant effect for listening condition for the French listeners, although removing the factor "Listening Condition" from the model led to a significantly worse model. The similar size of the effect of the presence of background noise on native and non-native listening is in agreement with recent results on word recognition in the presence of background noise, which also found a similar effect of noise on native and non-native listening (Scharenborg et al., 2018).

Based on the differences in sentence accent cueing and differences in the use of acoustic cues for the perception of sentence accent, we expected differences in the acoustic cues used between the different language groups. Language-dependent analyses taking into account acoustic prosodic cues indeed showed differences in the acoustic cues used for sentence accent detection. In line with the literature (Fry, 1955, 1958; Gussenhoven, 1983; Sluijter & van Heuven, 1996), the analyses showed that both English and Dutch listeners primarily used energy for sentence accent detection with the Dutch listeners also using duration as a cue. The use of the durational cue, albeit unexpected given the lack of differences in duration between the two prosodic contexts, gives an indication of the potential language-dependent cue sensitivity of duration for the Dutch listeners (see Sluijter & Van Heuven, 1996). Again in line with the literature (Vainio & Järvikivi, 2006), Finnish listeners were found to primarily rely on the use of F0. The results for the French listeners were less in agreement with existing literature, which is perhaps not surprising given the contradictory prior results for French and the suggestion that French are "deaf" to prominence (e.g., Dupoux et al., 2001, 2008; Peperkamp & Dupoux, 2002). In our experiment we found that the French listeners used both energy (like the English and Dutch listeners) and F0 (similar to the Finnish listeners) information, and additionally spectral tilt, although this use was dependent on the French listener's proficiency in English. The French listeners thus showed a different pattern from the Finnish, who only used "non-local" F0, and the English and Dutch listeners, who only used "local" energy and duration, in that the French used both "non-local" F0 and spectral tilt and "local" energy. These results agree with Frost (2011), who found that French listeners rely more on F0 for accent detection than the English, but do not line up with those by Séguinot (1977)—who showed that French listeners relied less on energy than English listeners.

This differential use of local and non-local cues by the different non-native listener groups might be an explanation for the surprising results for the Dutch (worse performance than expected on the basis of their proficiency) and French (better performance than expected on the basis of their proficiency) listeners. Background noise masks local, acoustic cues (e.g., Cooke, 2009), which negatively affects speech comprehension (e.g., Garcia Lecumberri et al., 2010, and references therein; Scharenborg et al., 2018). The lack of an interaction between the listening conditions and the acoustic cues indicates that the same cues are used in clean and noisy listening conditions, and thus that listeners do not seem to adapt their cue use to changing listening conditions. All listener groups detected significantly fewer target phonemes in the worst listening condition, suggesting that indeed the background noise is masking important cues for the uptake of sentence accent. This means that for the two listener groups who only used local cues (energy and duration), that is, the English and Dutch listeners, important information is masked and thus can be used to a much lesser extent leading to a reduction in the ability to detect sentence accent and thus lower phoneme detection rates. The significantly lower detection rate for the Dutch listeners in the clean condition compared to the English listeners continues to be lower in the presence of background noise. This explains the lower results for the Dutch listeners. From this perspective, the results of the Dutch listeners are as expected and not surprising. The Finnish and French listeners use non-local cues, which are expected to survive the background noise better. The fact that both listener groups outperformed the Dutch non-native listener group, despite having a similar or even a significantly lower proficiency, indeed confirms that non-local

cues survive the background noise to a larger extent than local cues and that the Finnish and French listeners are thus able to continue to use these cues for successful sentence accent detection. The significantly lower performance of the Dutch listeners can thus be explained by the differences in cue use compared to the Finnish listeners. Finnish listeners suffer less from the presence of background noise because they use an acoustic cue that is better preserved in background noise.

The question remains why the French listeners, despite the significantly lower proficiency in English compared to the Finnish and Dutch listeners, perform similar to the English listeners. Due to French typically not employing prosodic prominence (apart from the case of corrective focus), French not having clearly distinguishable word-level prominence (e.g., Dupoux et al., 2001, 2008; Peperkamp & Dupoux, 2002), and also a relative phrasal prominence that is less present compared to, for example, English (Frost, 2011), French listeners might not have clear expectations about which acoustic configurations constitute a sentence accent. When learning a language that does have sentence accent, they have to learn to perceive sentence accent. However, because their perception sentence accent is not restricted by a "native filter," they can use any acoustic cue that is available to perceive of sentence accent. This explanation is in line with the finding that proficiency in the non-native language interacted with the use of several acoustic cues for the French listeners, but not for any of the other listener groups. And as argued above, the cues the French listeners use include the non-local cues spectral tilt and F0, which make them more robust against the presence of background noise than the Dutch and English listeners. This seems to indicate that (high) proficiency helps the non-native listener to overcome differences at the prosodic level between the native and non-native language (Scharenborg et al., 2016).

In conclusion, native and non-native listeners of English were found to be able to exploit sentence accent for improved target phoneme detection, irrespective of the prosodic cues used to cue sentence accent perception in one's native language. This confirms that non-native listeners can overcome certain differences at the prosodic level between the native and non-native language, at least at high-proficiency levels. Relative similarity between the prosodic cues to sentence accent of one's native language compared to that of a non-native language did not determine the ability to perceive and use sentence accent for speech perception. Of the non-native language groups, only the Dutch listeners performed worse than the native listeners; while based on the increasing dissimilarity of Dutch, Finnish, and French to English, the French were hypothesized to perform worst, they did not—despite their proficiency level also being significantly lower than that of the Dutch and Finnish listeners. The effect of background noise on sentence accent detection was found to be dependent on the acoustic information used. Cross-linguistic differences in the use of local and non-local cues to sentence accent detection fully explained the results, where non-local cues seem to survive the background noise better. These results suggest that when listening conditions are suboptimal, non-native listeners fall back to rely on their native prosodic cues, if available. In the case that the native language has no clear sentence accent marking, highly proficient non-native listeners learn to exploit different acoustic cues for the perception of sentence accent.

Acknowledgements

The authors thank Joop Kerkhoff for creating the Praat script, and Yvonne Flory and Elea Kolkman for running the English and Dutch experiments, respectively.

Part of the here-presented data have been presented at Speech Prosody 2018, Poznan, Poland. The results presented here supersede those reported at Speech Prosody 2018.

Odette Scharenborg is now at the Multimedia Computing Group at Delft University of Technology, the Netherlands.

Funding

This research is supported by a Vidi-grant from the Netherlands Organisation for Scientific Research (grant number 276-89-003) awarded to Odette Scharenborg.

References

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and non-native listening. *Bilingualism: Language and Cognition*, 6, 81–96.

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

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer [Computer program]. Version 5.1. Retrieved from http://www.praat.org/

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, *121*, 2339–2349.

Broersma, M., & Scharenborg, O. (2010). Native and non-native listeners' perception of English consonants in different types of noise. *Speech Communication*, *52*, 980–995.

Campbell, N., & Beckman, M. E. (1997). Stress, prominence, and spectral tilt. In A. Botinis, G. Kouroupetroglou, & G. Carayiannis (Eds.), *Intonation: Theory, Models, and Applications (Proceedings of an ESCA Workshop)* (pp. 67–70). Athens, Greece: ESCA.

Carroll, R., & Ruigendijk, E. (2016). ERP responses to processing prosodic phrasing of sentences in amplitude modulated noise. *Neuropsychologia*, *82*, 91–103.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I. Between the grammar and physics of speech* (pp. 283–333). Cambridge, UK: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*, 425–452.

Cooke, M. (2009). Discovering consistent word confusions in noise. In *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech-2009)*, *Brighton, UK* (pp. 1887–1890). Australia: Causal Productions Pty Ltd.

Cooke, M., García Lecumberri, M. L., & Barker, J. P. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, *123*, 414–427.

Cooke, M., Garcia Lecumberri, M. L., Scharenborg, O., & van Dommelen, W. A. (2010). Language-independent processing in speech perception: Identification of English intervocalic consonants by speakers of eight European languages. *Speech Communication*, *52*, 954–967.

Cruttenden, A. (1997). *Intonation*. Cambridge, UK: Cambridge University Press.

Cruttenden, A. (2006). The de-accenting of given information: A cognitive universal? In G. Bernini & M. L. Schwartz (Eds.), *Pragmatic organization of discourse in the languages of Europe*. (pp. 311–355). Berlin, Germany: Mouton de Gruyter.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, *20*, 55–60.

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

Cutler, A., & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics*, *29*, 217–224.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*, 1–10.

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *Journal of the Acoustical Society of America*, *124*, 1264–1268.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness." *Journal of the Acoustical Society of America*, *110*, 1606–1618.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition*, *106*, 682–706.

Eimas, P. D., & Nygaard, L. C. (1992). Contextual coherence and attention in phoneme monitoring. *Journal of Memory and Language*, *31*, 375–395.

Fant, G., & Kruckenberg, A. (2004). Intonation analysis and synthesis with reference to Swedish. In *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages (TAL-2004)*, *Beijing, China* (pp. 57–60) ISCA.

Fletcher, H., & Munson, W. A. (1933), Loudness, its definition, measurement and calculation. *The Journal of the Acoustical Society of America*, *5*, 82–108.

Foss, D. J., Harwood, D., & Blank, M. A. (1980). Deciphering decoding decisions, data and devices. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 165–199). Hillsdale, NJ: Erlbaum.

Frost, D. (2011). Stress and cues to relative prominence in English and French: A perceptual study. *Journal of the International Phonetic Association*, *41*, 67–84.

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.

Garcia Lecumberri, M. L. (1995). Perception of accentual focus by Basque L2 learners of English. In *Anuario del Seminario Julio de Urquijo (ASJU)* (pp. 581–598). Spain.

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America*, *119*, 2445–2454.

Garcia Lecumberri, M. L., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *53*, 864–886.

Gussenhoven, C (1983). Review of N. Willems, English intonation from a Dutch point of view (Dordrecht: Foris, 1981). In: *Toegepaste Taalkunde in Artikelen, Amsterdam: Vrije Universiteit*, *17*, 273–278.

Gussenhoven, C., & Jacobs, H. (2011). *Understanding phonology* (3rd ed.). New York, NY: Routledge.

Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, modeling and technology* (pp. 209–242). Dordrecht, The Netherlands: Kluwer.

Kakouros, S., & Räsänen, O. (2016). 3PRO—An unsupervised method for the automatic detection of sentence prominence in speech. *Speech Communication*, *82*, 67–84.

Kakouros, S., Räsänen, O., & Alku, P. (2017). Evaluation of spectral tilt measures for sentence prominence under different noise conditions. In *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech-2017)*, *Stockholm, Sweden* (pp. 3211–3215). Australia: Causal Productions Pty Ltd.

Karaminis, T., & Scharenborg, O. (2018). The effects of background noise on native and non-native spoken-word recognition: A computational modeling approach. In *Proceedings of the Cognitive Science Conference, Madison, WI*. Cognitive Science Society.

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

Koopmans-van Beinum, F. J., & van Bergem, D. R. (1989). The role of "given" and "new" in the production and perception of vowel contrasts in read text and in spontaneous speech. In *Proceedings of the European Conference on Speech Communication and Technology*, *Paris* (pp. 113–116). France: ESCA.

Ladd, D. R. (1990). Intonation: Emotion vs. grammar. Review of: *Intonation and Its Uses*, by Dwight Bolinger. *Language 66*, 806–816.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge, UK: Cambridge University Press.

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

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge, UK: Cambridge University Press.

Lein, T., Kupisch, T., & van de Weijer, (2016). Voice onset time production in adult simultaneous bilinguals (German-French) and the role of childhood. *International Journal of Bilingualism*. Retrieved from DOI: 10.1177/1367006915589424

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343.

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

Lisker, L., & Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*, 203–243.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress "deafness." In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 203–240). Berlin, Germany: Mouton de Gruyter.

Rosenberg, A., & Hirschberg, J. (2009). Detecting pitch accents at the word, syllable and vowel level. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT-2009), Companion Volume: Short Papers*, *Boulder, CO* (pp. 81–84). Stroudsburg, PA: Association for Computational Linguistics.

Rosenberg, A., Hirschberg, J., & Manis, K. (2010). Perception of English prominence by native Mandarin Chinese speakers. In *Proceedings of the Fifth International Conference on Speech Prosody (SPro-2010)*, *Chicago, IL* (paper 982). ISCA.

Scharenborg, O., Coumans, J., & van Hout, R. (2018). The effect of background noise on the word activation process in non-native spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Retrieved from http://dx.doi.org/10.1037/xlm0000441

Scharenborg, O., Kolkman, E., Kakouros, S., & Post, B. (2016). The effect of sentence accent on non-native speech perception in noise. In *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech-2016)*, *San Francisco, CA* (pp. 863–867). Australia: Causal Productions Pty Ltd.

Scharenborg, O., Weber, A., & Janse, E. (2015). The role of attentional abilities in lexically-guided perceptual learning by older listeners. *Attention, Perception, and Psychophysics*, *77*, 493–507.

Séguinot, A. (1977). L'accent d'insistance en français standard. In F. Carton, D. Hirst, A. Marschal, & A. Séguinot (Eds.), *Studia Phonetica 12: L'accent d'insistance: Emphatic stress* (pp. 1–58). Paris, France: Didier.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193–247.

Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, *102*, 250–255.

Shinn-Cunningham (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*, 182–186.

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

Suomi, K., Toivanen, J., & Ylitalo, R. (2008). Finnish sound structure. Studia Humaniora Ouluensia 9. Oulu: Oulun yliopisto.

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America*, *89*, 1768–1776.

Terken, J., & Hermes, D. (2000). The perception of prosodic prominence. In M. Horne (Ed.), *Prosody: Theory and experiment. Studies presented to Gösta Bruce* (pp. 89–127). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Vainio, M., & Järvikivi, J. (2006). Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics*, *34*, 319–342.

Vallduví, E. (1992). *The informational component*. New York, NY: Garland Press.

Van Alphen, A., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics*, *32*, 455–491.

Van Kuijk, D., & Boves, L. (1999). Acoustic characteristics of lexical stress in continuous telephone speech. *Speech Communication*, *27*, 95–111.

Van Zyl, M., & Hanekom, J. J. (2011). Speech perception in noise: A comparison between sentence and prosody recognition. *Journal of Hearing Science*, *1*, 54–56.

Venditti, J. J., Jun, S.-A., & Beckman, M. E. (1996). Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 287–311). Hillsdale, NJ: Erlbaum.

Wagner, P. (2005). Great expectations—Introspective vs. perceptual prominence ratings and their acoustic correlates. In *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech-2005)*, *Lisbon, Portugal* (pp. 2381–2384). Australia: Causal Productions Pty Ltd.

Wagner, P., Origlia, A., Avesani, C., Christodoulides, G., Cutugno, F., D'Imperio, M., … Moniz, H. (2015). Different parts of the same elephant: A roadmap to disentangle and connect different perspectives on prosodic prominence. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS-2015)*, *Glasgow, Scotland*. London: International Phonetic Association.

Zahorian, S. A., & Hu, H. (2008). A spectral/temporal method for robust fundamental frequency tracking. *Journal of the Acoustical Society of America*, *123*, 4559–4571.

Appendix A. Overview of all 48 target-bearing words and the experimental sentences in which they were embedded.

Carrier sentence	Target-bearing word
The documentary about poverty interested the common viewers.	common
The material for the tires was tested by the car manufacturer.	car
The statement of the crown witness led to the pickpocket's arrest.	crown
The area near the coast responded to the tornado warning.	coast
The young man on the corner was wearing the paper hat.	paper
The company of tap dancers negotiated with the pop star's agent.	pop star
The ducklings in the pond were fighting for the cake crumbs.	pond
The woman with the parrot went into the teacher's office.	parrot
The actions of the crew led to the test lab's evacuation.	test lab
The rising price of packaging worried the tennis racket manufacturer.	tennis
The group of students from Thailand photographed the crocodile wrestler.	Thailand
The mayor of the town will check on his popularity ratings.	town
The members of the Pomo tribe lived off the cotton trade.	cotton
The entrepreneurs in the transport sector suffered most from the coalmine's closure.	coalmine
The flavor of the coffee was ruined by the polluted water.	coffee
The investigator of the crime was shot by the taxi driver.	crime
The manager of the kitchen staff was upset about the pumpkin soup.	pumpkin
The visitors of the theme park came to look at the pygmy hippo's young.	pygmy
The archaeologists from Poland went excavating in the Kalahari desert.	Poland
The man in the purple suit was welcomed by the tailor's assistant.	purple
The farm with the cornfields was close to the toy factory.	toy
The article on prostate cancer was praised at the 12th Medical Congress.	twelfth
The reporter from *The Telegraph* was writing about the cat lover.	telegraph
The flight to the tropics crashed near Palm Island.	tropics
The bones of the triceratops were found by the Cuban archaeologist.	Cuban
The owner of the potato farm refused to go to the counselor's party.	counselor
The villa with the carport must belong to the tobacco farmer.	carport

(Continued)

Appendix A. (Continued)

Carrier sentence	Target-bearing word
The member of the cabinet was involved in the plagiarism incident.	cabinet
The old man with the tattoos had worked as a prison guard.	prison
The mistress of the king was given a pearl necklace.	pearl
The members of parliament were outraged by the tabloid's allegations.	parliament
The suspension of the pastor was confirmed at the college meeting.	pastor
The watcher on the patio saw the tournament finals.	tournament
The voice of the caller was hard to hear through the telephone connection.	telephone
The attendant at the toll booth looked through the passenger window.	toll
The chauffeur of the teenager refused to pick up her cousin's dog.	teenager
The shed for the tools was built in the cabin's garden.	cabin
The children in the primary school were drawing on the crayon boxes.	crayon
The remains of the camp were found by the tiger hunter.	camp
The personnel officer of the company interviewed the polo player.	company
The demonstrators at the train station were arrested by the police officer.	police
The box with the contraband was hidden in the parking lot.	parking
The group of tourists from Prague received information about the Tate Gallery.	Prague
The owner of the pawn shop checked the customer's items.	pawn
The shed near the park ranger's house was made of teak wood.	teak
The program about cocaine was meant for television broadcast.	television
The help of the tutor was appreciated by the pupil's mother.	tutor
The group of turtle enthusiasts wrote about Korean mythology.	turtle

The target phoneme is the first phoneme of the target-bearing word.

Appendix B. Overview of all 48 filler sentences; words in capitals denote words carrying sentence accent.

Filler sentence	Target phoneme	Target-bearing word
The villagers were against the expansion of the CAMPSITE.	*k*	campsite
The second-hand CAR was bought by the geology student.	*k*	car
The PAINTINGS in the gallery turned out to be forgeries.	*p*	paintings
The ambassador's wife wanted to order a new TABLE.	*t*	table
The value of the bonds was altered with the devalued CURRENCY.	*k*	currency
The delegation from the USA established PEACE between the Arab nations.	*p*	peace
The PERPETRATOR of the bombing was a member of the Afghan army.	*p*	perpetrator
The TOURISTS in the underground were greatly amused by the Japanese graffiti.	*t*	tourists
The participant in the running relay race sprained her KNEE.	*k*	*
His OLD Volvo broke down while he was on holiday in Kent.	*p*	*
The victim of the gang's vicious attack had cuts on his BACK.	*t*	*

Appendix B. (Continued)

Filler sentence	Target phoneme	Target-bearing word
The FLUTE ensemble performed in the school building.	*t*	*
The mother of three daughters wrote to the BOARDING school.	*k*	*
The STUDENT from Australia went to the book signing.	*k*	*
The General of the Army was killed during the CIVIL uprising.	*p*	*
The bridge nearest the CITY was sabotaged by the Libyan army.	*t*	*
The CAPTAIN of the expedition refused to turn back to the nearest harbor.	*k*	captain
The blonde woman won a prize at the CONTEST.	*k*	contest
The intern at the lab researched the rat's sleeping PATTERN.	*p*	pattern
The army officer was not happy about his POSTING to the desert war.	*p*	posting
The PEOPLE on the square rallied against gay marriage.	*p*	people
The figure skating TEAM from Brazil lost the world championship.	*t*	team
The cricket grounds of the village were maintained by the TEST match player.	*t*	test
The rich playboy's TIME was mostly spent on his luxury yacht.	*t*	time
The director knew nothing of the SCANDAL's outcome.	*k*	*
The celebrity was arrested for driving under INFLUENCE.	*p*	*
The board of the health center went to the conference on DONOR kidneys.	*p*	*
The members of the Women's Association held a collection for the VICTIMS of the flood.	*t*	*
The personnel manager of the department store fired the LAZY salesman.	*k*	*
The university vice-chancellor's opinion was reported on the EVENING news.	*p*	*
The association of online shoppers objected to the NEW delivery procedure.	*t*	*
The entrance to the dormitory was located near the BAPTIST church	*t*	*
The CHECKING of the ballots was interrupted when the machine failed.	*k*	*
The ROAD to the bridge was flooded by the rain.	*k*	*
The SCIENCE students listened to the violin concerto by Beethoven.	*p*	*
The Californian SENATOR proposed the motion to dismiss.	*t*	*
The ATTITUDE of the businessman aroused his associates' anger.	*k*	*
The head of staff of the BANK was facing a dollar shortage.	*p*	*
The FATHER of the deaf child was given instructions on the baby monitor's use.	*p*	*
The report of the SKIING accident was aired on the evening news.	*t*	*

(Continued)

Appendix B. (Continued)

Filler sentence	Target phoneme	Target-bearing word
The old folks' CLUB had gone on its regular Friday afternoon bus trip.	*k*	club
The CLUB of derby fans attended the new driver's first race.	*k*	club
The members of the board asked about the budget CUTS.	*k*	cuts
The chicken farmers voted against the minister's new fertilizer POLICY.	*p*	policy
The PRIEST from Denver did not attend the Easter celebration.	*p*	priest
The Basque TERRORISTS were responsible for the assassination.	*t*	terrorists
The residents of the neighborhood were annoyed by the noisy TRAFFIC.	*t*	traffic
The brain surgeon was unable to remove the smallest of the TUMORS.	*t*	tumors

The column denoted "Target phoneme" indicates the phoneme displayed on the computer screen. The "*" in the column "Target-bearing word" indicates that no target phoneme was present.