The Effect of Background Noise on the Word Activation Process in Nonnative Spoken-Word Recognition

Article *in* Journal of Experimental Psychology Learning Memory and Cognition · August 2017

DOI: 10.1037/xlm0000441

3 authors:


Odette Scharenborg
Delft University of Technology


Juul Coumans
Open Universiteit Nederland


Roeland van Hout
Radboud University

The effect of background noise on the word activation process in non-native spoken-word recognition

Odette Scharenborg[1,2], Juul M. J. Coumans[1], Roeland van Hout[1]

[1] Centre for Language Studies, Radboud University Nijmegen, The Netherlands

[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

10.1037/xlm0000441

Corresponding author:

Odette Scharenborg

Centre for Language Studies

Radboud University Nijmegen

Erasmusplein 1

6525 HT Nijmegen

The Netherlands.

E-mail: O.Scharenborg@let.ru.nl

Abstract

This paper investigates two questions: 1) does the presence of background noise lead to a differential increase in the number of simultaneously activated candidate words in native and non-native listening; 2) do individual differences in listeners' cognitive and linguistic abilities explain the differential effect of background noise on (non-)native speech recognition? English and Dutch students participated in an English word recognition experiment, in which either a word's onset or offset was masked by noise. The native listeners outperformed the non-native listeners in all listening conditions. Importantly, however, the effect of noise on the multiple activation process was found to be remarkably similar in native and non-native listening. The presence of noise increased the set of candidate words considered for recognition in both native and non-native listening. The results indicate that the observed performance differences between the English and Dutch listeners should not be primarily attributed to a differential effect of noise, but rather to the difference between native and non-native listening. Additional analyses showed that word-initial information was found to be more important than word-final information during spoken-word recognition. When word-initial information was no longer reliably available word recognition accuracy dropped and word frequency information could no longer be used suggesting that word frequency information is strongly tied to the onset of words and the earliest moments of lexical access. Proficiency and inhibition ability were found to influence non-native spoken-word recognition in noise, with a higher proficiency in the non-native language and worse inhibition ability leading to improved recognition performance.

Keywords: non-native spoken-word recognition, noise, competitor space, proficiency, inhibition ability.

1 Introduction

A common observation is that communication in the presence of background noise is more difficult in a non-native than in the native language - even for those who have a high proficiency in the non-native language involved (see for experimental evidence, e.g., Bradlow & Alexander, 2007; Brouwer, van Engen, Calandruccio, & Bradlow, 2012; Cooke, Garcia Lecumberri, & Barker, 2008; Ezzatian, Avivi-Reich, & Schneider, 2010; Kilman, Zekveld, Hällgren, & Rönnberg, 2014; Mayo, Florentine, & Buss, 1997; Meador, Flege, & Mackay, 2000; Van Wijngaarden, Steeneken, & Houtgast, 2002). The main reason for this problem seems obvious: Imperfect knowledge of the non-native language and a degraded speech signal due to the presence of background noise interact strongly to our disadvantage when we listen to a non-native language (Bradlow & Pisoni, 1999; Garcia Lecumberri, Cooke, & Cutler, 2010; Mayo et al., 1997). A common observation is that some people have more difficulty in such situations than others. The aim of this paper is to investigate the effect of background noise on the processes underlying non-native spoken-word recognition, and why some listeners cope better with non-native listening in noise than others.

Most influential models of spoken-word recognition assume that during spoken-word recognition all words that partially overlap with the input are activated simultaneously (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Gow & Gordon, 1995; Luce & Pisoni, 1998; Slowiaczek, Nusbaum, & Pisoni, 1987; Zwitserlood, 1989), and they subsequently compete for recognition (for a review see, McQueen, 2005). So upon hearing the word ‘shape’, partially overlapping words such as ‘shade’ and ‘cape’ will also be activated and compete for recognition. Listening in a non-native language and the presence of background noise both have a detrimental effect on this mapping process of the acoustic signal to a listener’s segmental and word representations. Because of differences in phoneme inventories between languages (e.g., Dutch does not have the /æ/ as in English *match*), non-native listeners' sound categories are less well specified or even absent, resulting in inaccurate sound perception (see for an overview the papers in Bohn & Munro, 2007). Consequently, there is an increase in the number of words which partially overlap with the speech input, and the number of word candidates that is simultaneously activated and considered for recognition increases due to the inaccurate perception of non-native sounds (e.g., Broersma, 2012; Cutler, Weber, & Otake, 2006; Pallier, Colomé, & Sebastián-Gallés, 2001). Not only additional words from the non-native language (e.g., *mention*) are spuriously activated but also words from the native language can be activated (e.g., Dutch *mes*, English: *knife*; Spivey & Marian, 1999; Weber & Cutler, 2004).

The presence of background noise interferes with the intelligibility of the speech signal due to its masking of the available acoustic information. Consequently, similar to the situation of non-native listening, the phonological match between the speech signal and the listener's segmental representations is decreased. This has been shown to lead to a decrease in phoneme perception accuracy in noise for both native and non-native listeners, but more so for the latter group (see for an overview, Garcia Lecumberri et al., 2010). The effect of this phonological mismatch due to a degraded speech signal on spoken-word recognition is however not entirely clear. Potentially, the number of simultaneously activated word candidates increases due to the target word's speech signal being a partial match with several other candidate words (referred to as the *many-additional-competitor scenario* by Chan & Vitevitch, 2015). Another possibility is that the target word is a good match to a different, erroneous word (referred to as the *single-strong-competitor scenario* by Chan & Vitevitch, 2015), resulting in the activation of one erroneous word candidate rather than the activation of many spurious word candidates. Moreover, the effect of noise on the number of simultaneously activated word candidates could be different for native and non-native listeners, in line with the differential effect of noise found for phoneme recognition (e.g., Cooke et al., 2010; Broersma & Scharenborg, 2010). Since the effect of noise on phoneme perception seems to be more detrimental for non-native listeners than for native listeners, the increase in number of simultaneously activated candidate words might therefore be larger in non-native listening than in native listening.

During spoken-word recognition, both candidate words that overlap with the onset of the spoken word (onset competitor, 'shade' in the example above) and words that overlap with the offset of the spoken word (offset competitor, 'cape' in the example above) are activated. Native listeners thus use both word-initial and word-final information for candidate word selection and word recognition (Slowiaczek, Nusbaum, & Pisoni, 1987; van der Vlugt & Nooteboom, 1987), although word-initial information seems to be more important in evaluating lexical candidates than word-final information, at least in clear listening conditions (Allopenna, Magnuson, & Tanenhaus, 1998; Marslen-Wilson & Zwitserlood, 1989; McQueen & Huettig, 2012). In the presence of background noise, the activations of candidate words have been shown to change during competition in native listening (Ben-David, Chambers, Daneman, Pichora-Fuller, Reingold, & Schneider, 2011; McQueen & Huettig, 2012), e.g., offset competitors have been found to be activated more compared to clean listening conditions (Brouwer & Bradlow, in press; McQueen & Huettig, 2012). Background noise thus modulates the competition process. There are indications that this debilitating effect of noise on the competition process is increased in non-native listening (Drozdova, van Hout, & Scharenborg, 2015).

This paper investigates the effect of background noise on the multiple activation process in non-native and native listening, focusing on the effect of background noise on the (spurious) activation of onset and offset competitors. In particular, the question we aim to answer is whether the presence of background noise leads to a differential increase in number of simultaneously activated candidate words in native and non-native listening, even after accounting for differences in recognition accuracy between the native and non-native listeners. Moreover, despite there being considerable individual variation in performance in non-native listening in noise, so far little attention has been paid to individual differences in non-native listening in noise. A second question of this paper is therefore whether individual differences exist in the effect of background noise on non-native speech recognition depending on a non-native listener's cognitive and linguistic abilities.

To address these questions, a free-response word identification experiment was conducted with two listener groups, a group of English native listeners and a group of Dutch non-native listeners of English. The set of answers provided by the English and Dutch listener groups, respectively, then served as a proxy to the size of the competitor space for the particular listener groups. In the experiment, words were presented in a clean listening condition, and, crucially, in listening conditions where their onsets or offsets were masked by noise at three increasingly difficult signal-to-noise ratios (SNRs) after which listeners had to type in the word they thought they heard. When the word onset was masked, word-initial information was no longer reliably available; while when word offset was masked, word-final information was no longer reliably available. Word accuracies were analysed to investigate the effect of background noise on native and non-native spoken word recognition. The size of the set of misrecognised words, or misperceptions, was subsequently analysed to investigate the effect of the presence of noise on the multiple activation process and the size of the competitor space during native and non-native word recognition. Finally, individuals' cognitive and linguistic abilities were entered as predictors in the statistical analyses of the non-natives to investigate individual differences in the effect of noise on the cognitive processes underlying non-native word recognition.

We are aware of only one study with the particular aim to investigate the contribution of proficiency in the non-native language and cognitive abilities on the level of individuals on non-native speech perception in noise. Kilman and colleagues (2014) found that, in their study, the proficiency in English of their Swedish participants was the most important factor in explaining differences in non-native speech perception in noise. Moreover, the cognitive skill they investigated, working memory, did not show a strong link with non-native speech perception in noise. Other studies also found an effect of proficiency on non-native speech perception in noise, although these studies did not investigate proficiency on an individual level but rather through group comparisons between native and non-native listeners (e.g., Mayo et al.,1997; Cooke, Garcia Lecumberri, Scharenborg, & van Dommelen, 2010). In the current study, individual listeners' non-native proficiency was assessed by using the LexTALE test (Lemhöfer & Broersma, 2012) and entered as a predictor of word recognition performance in the statistical analyses.

Kilman and colleagues (2014) moreover found that listeners with a high proficiency in English were better able to inhibit both native and non-native babble maskers (similarly, native listeners with a larger vocabulary were shown to adapt better to accented speech in the presence of background noise than listeners with a smaller vocabulary; Banks, Gowen, Munro, & Adank, 2015), while Cooke and colleagues (2010) observed that the detrimental effect of an informational masker (a competing talker) over an energetic masker (speech modulated noise) was larger for non-native listener groups with a lower proficiency in English compared to non-native listener groups with a higher proficiency in English. On the basis of these findings, Kilman and colleagues (2014) speculate that a higher proficiency in the non-native language improves listeners' ability to inhibit speech from a competing talker (in both the native and non-native language) in the background, and link this to simultaneous bilinguals' improved executive, and particularly inhibition, control (Bialystok, 2001;

Bialystok, Craik, Klein, & Viswanathan, 2004). Kilman and colleagues thus seem to argue that high-proficient non-native speakers might obtain the same improved inhibition ability as simultaneous bilinguals. Inhibition has been described by Sikora and colleagues (Sikora, Roelofs, Hermans, & Knoors, 2016) as the ability to reduce the activation of an unwanted dominant or automatic response (Aron, Robbins, & Poldrack, 2004). Research has shown that inhibition builds up over time (Ridderinkhof, 2002; Van den Wildenberg, Wylie, Forstmann, Burle, Hasbroucq, & Ridderinkhof, 2010). Potentially, listeners with worse inhibition ability are worse at inhibiting active candidate words during the competition process, resulting in worse performance in non-native speech perception in noise, in contrast to listeners with a better inhibition ability, who will be quicker in reducing their competitor space. We investigate this by using non-native listeners' inhibition ability, measured by the Flanker task, as a predictor of the word recognition task.

2 Method

2.1 Participants

Fifty native English students (38 females; M_{age} = 20.9 years, SD_{age} = 2.1, age range: 18-29 years) from the University of York, UK, and 61 native Dutch students (48 females; M_{age} = 21.4 years, SD_{age} = 2.1, age range: 18-26 years) from the Radboud University, Netherlands, participated in the experiment, and were paid for their participation. The Dutch participants had an average of 7.9 years of formal English education (*SD* = 2.3; range: 6-16 years). None of the participants reported to have a history of speech and/or hearing disorders. All participants signed a consent form prior to the experiment. No participant took part in more than one experiment.

2.2 Materials

The stimuli consisted of 42 triplets of English words, each consisting of, what we refer to as, a head-word (e.g., *letter*), a word which shared word-initial information with the head-word (e.g., *lettuce*; referred to as matched-onset-word), and a word which shared word-final information with the head-word (e.g., *sweater*; referred to as matched-offset-word). The complete set consisted of (42 x 3 =) 126 words in total; an overview of all stimuli including their word frequency and the overlap between target and competitors in terms of number of phonemes is provided in Appendix A. Fifteen of the triplets consisted of bisyllabic words (45 words in total) and 27 triplets consisted of monosyllabic words (81 words in total). Seven sets of bisyllabic words were taken from the study of Allopenna and colleagues (1998). The other eight sets of bisyllabic words were based on stimuli from Shock and colleagues (2010). The sets of the monosyllabic words were selected by searching through image databases (Snodgrass & Vanderwart, 1980), online rhyme dictionaries (http://www.rhymezone.com), and online scrabble dictionaries (http://www.scrabblefinder.com). All words within a triplet had the same stress pattern according to Celex (Baayen, Piepenbrock, & van Rijn, 1993). Word frequencies ranged from 14 (‘moose’) to 13,180 (‘life’) per million. The average number of neighbours was similar for the head-word (average of 21.1 neighbours), matched-onset-words (average of 20.2 neighbours), and the matched-offset-words (average of 22.5 neighbours). The mean number of word-initial neighbours (defined as having an overlap of 2 or more phonemes) for the set of head-words was 9.4 and 9.0 for the set of onset-matched-words. The mean number of word-final neighbours for the set of head-words was 9.1 and 9.0 for the set of off-set-matched words.

The stimuli were produced by a male native speaker of Southern British English. All stimuli were recorded in a sound-attenuated booth at 44.1 kHz. Subsequently, the audio files were down-sampled to 16 kHz to make them compatible with the noise file. Next, all stimuli were cut manually using Praat (Boersma & Weenink, 2013) into one-word audio files, and the average (i.e., root-mean-square) intensity was set at 60 dB SPL.

2.2.1 Adding noise

In the next step, stationary speech-shaped noise (SSN) was added to all stimuli at three different signal-to-noise ratios (SNRs). The SNRs used in the experiment were -12 dB, -6 dB, and 0 dB. These values were based on a separate pilot test in which 12 Dutch non-native listeners of English participated and chosen such as to avoid floor and/or ceiling effects. None of the participants tested in the pilot study participated in the main experiment.

The onset and offset masking was tailored to each head-word and matched-onset-word/matched-offset-word pair, such that the stretch of speech where the head-word and the matched-onset-word/matched-offset-word overlap phonemically remained unmasked. For example, the head-word *letter* and its matched-onset-word *lettuce* share [lɛt], which is consequently left unmasked, while [ə] of [lɛtə] and [əs] of [lɛtəs] are masked (this condition is referred to as the offset-masked condition). Following the same logic, the head-word *letter* and its matched-offset-word *sweater* share [ɛtə], which consequently remains unmasked, while [l] and [sw] are masked (this is referred to as the onset-masked condition). The result of this procedure is shown in the right-hand side of Figure 1 for two triplets: the shading of the letters indicates the amount of masking for these three words (note that in the actual process phonemes were used to determine the deviation point between the two words, not letters).

< Insert Figure 1 around here. >

Noise was added to the stimulus files using a semi-automatic method. First, boundaries were manually placed in the audio files at positive-going zero-crossings between the overlapping part of the word and the part that differed, e.g., *lett|er* - *lett|uce* for the offset-masked condition *and l|etter – sw|eater* for the onset-masked condition (where | denotes the manually placed boundary). Subsequently, an X was put in the tier to mark the part of the signal that had to be masked. A random stretch of speech-shaped noise was subsequently added to the sound files on the marked part of the word using a custom-made Praat script. For the onset-masked stimuli, 200 ms of leading noise was added to the stimuli; for the offset-masked stimuli, 200 ms of trailing noise was added. A Hamming window was applied to the noise, with a fade in of 10 ms for onset masking and a 10 ms fade out for offset masking. The mean overlap between head-words and their matched-onset-words was 2.5 phonemes, and the overlap between head-words and their matched-offset-words was 2.7 phonemes.

2.2.2 Lists

During the experiment, a listener would only receive one masked version of each head-word and would not receive an matched-onset or matched-offset word with its corresponding head-word. So, participants received either a head-word with word-initial masking and the matched-onset-word from the same triplet, or the head-word with word-final masking and the matched-offset-word from the same triplet.

To that end, the words from the triplets were first divided into four lists. Figure 1 visualises the procedure we followed for dividing the words over separate lists. The left hand side of Figure 1 shows two example triplets. The word in bold is the head-word (e.g., **letter**), the word in italics is its matched-onset-word (*lettuce*), and the underlined word is its matched-offset-word (<u>sweater</u>). First, head-words and their matched-onset-words were divided over the first two lists: the head-word with the highest frequency was put in the first list (**letter** in Figure 1), the head-word with the second highest frequency in the second list (**weather** in Figure 1), the word with the third highest frequency in the first list, and so on. The matched-onset-words were placed into these two lists, such that a head-word and its matched-onset-word were never in the same list (see also the left two columns of the middle part of Figure

1). Subsequently, the (same) head-words and their matched-offset-words were divided over the third and the fourth list following the same procedure. Lists 3 and 4 then only contained head-words and matched-offset-words, as illustrated in the right two columns of the middle part of Figure 1. The right-hand side of Figure 1 shows the parts that are obscured with noise, indicated by the shading. During the experiment, each participant would then only, e.g., receive lists 1and 3 or lists 2 and 4, see the grey/white-coding for the lists.

Subsequently, each list (42 stimuli) was divided into three blocks which each received a different SNR. The words were divided over the blocks on the basis of their word frequency to ensure that each block contained words with similar word frequencies. First, in each list the frequency of the words was ordered from high to low. Second, the list was divided in 14 sets, in which the first set contained the highest frequency words, the second set the second highest frequency words and the last set the lowest frequency words. Third, each word of a set was randomly put in one of the three blocks. Then, a different SNR value was assigned to each block in a list, such that each participant got all three SNRs per list (and thus got each SNR twice in the experiment). For example, one participant would get list 1 and hear the first presented block with a SNR value of -12 dB, the second presented block with a SNR value of 0 dB, and the last block with a SNR value of -6 dB. Next, in the second part of the experiment that participant would get list 3, again with the first block with a SNR value of -12, the second block with a SNR value of 0, and the last block with a SNR value of -6. Thus, the order of the SNR values per block remained the same for both parts of the experiment for each participant.

A full counterbalance of, the block orders, SNR orders, and lists resulted in 144 different experimental lists : 6 possible block orders (1-2-3, 1-3-2, etc.) * 6 possible SNR orders (-12, -6, 0; -12, 0, -6, etc.) * 4 possible list order (list 1-3, 3-1, 2-4, 4-2). For each of the 50 participants we tested, a random number between one and 144 was generated, to avoid selection bias, which corresponded to the experimental list number. The order of presentation of the block and SNR orders and the lists was fairly balanced. Nevertheless, variations occurred in how often a stimulus was presented in a specific condition. This is controlled for in the statistical analyses by including a Number-of-occurrences covariate. So, each experimental list contained both an onset-masked list (list 1 or 2) and an offset-masked list (list 3 or 4).

2.3 Procedure

Participants were tested individually in a sound-treated booth. The stimuli were presented binaurally over closed headphones. The experiment consisted of three parts: the first two parts consisted of two of the four experimental lists as described in the previous section. Part three, which always came last, consisted of all 84 words (6 blocks) that were presented in the previous two parts without masking. So, each participant heard 168 words in total. The order of the words within each block was randomised. After every block there was a self-paced pause. Before the start of the experiment, the participant was familiarised with the task through four practice trials. Each trial began with a fixation cross which stayed on the screen for 500 ms. Subsequently, an auditory stimulus was played after which participants had to respond. Participants were instructed that they would be listening to English words that were partially obscured with noise, and they were asked to type in the word they thought they had heard followed by the return key to continue to the next trial.

2.4 Proficiency and inhibition background tests

After participation in the word recognition experiment, both the native and non-native listeners were administered the Flanker test, while the non-native listeners were also administered the LexTALE test (in that order). The word recognition experiment and the tests were administered in one session. Test outcomes of three of the non-native participants (one with a missing

LexTale score and two with missing Flanker scores) and three of the native listeners were not recorded. These participants were left out of the individual differences analysis, leaving 47 native and 58 non-native listeners. Correlations between the two measures for the non-native listeners were not significant.

Inhibition ability was assessed using the computerised version of the Flanker task (Eriksen & Eriksen, 1974). During the task, participants had to indicate (by clicking either the 'z' or the '/' key on the keyboard) which direction the middle symbol (a leftward or rightward pointing arrowhead) in a row of 5 symbols points in three types of contexts, congruent (>>>>> or <<<<<), incongruent (<<><< or >><>>), and neutral (==>== or ==<==). The participants were instructed to maximise speed and accuracy. Six different stimuli were each presented 12 times in the test part (order of trial presentation was randomised for each participant) to make 72 trials. Prior to the test, six practice trials were presented. The individual Flanker cost for each participant was determined by computing the Flanker interference effect, i.e., each participant's mean logRT (of the RT in ms) in the incongruent condition was divided by that individual's mean logRT in the neutral condition. The higher a participant's Flanker cost, the poorer was their selective attention. The mean Flanker cost of the English native listeners was 1.024 (SD = .008) and for the Dutch non-native listeners was 1.032 (SD = .012).The difference between the native and non-native listeners was significant ($t(100)=3.44$, $p < .001$).

LexTALE (Lexical Test for Advance Learners of English; Lemhöfer & Broersma, 2012) was used to assess individuals' proficiency of English. LexTALE is an easy-to-use, swift vocabulary knowledge task. It consists of a visual lexical decision task of 60 trials (40 English words and 20 nonwords). Listeners were presented with the items one-by-one on a computer screen and were asked to make a word/nonword judgement. The task takes about 3 to 4 minutes to complete. The mean LexTALE score was 65.8 (SD = 13.7), which corresponds to an upper intermediate or B2 proficiency level of English (Lemhöfer & Broersma, 2012).

3 Results

Three main analyses were carried out. We first carried out a general analysis of the word recognition accuracies in the different noise conditions, preceded by an analysis of the accuracies in the clean condition versus the easiest SNR condition. This general analysis does not only provide a group comparison between the native and non-native results, it will also highlight the extent to which background noise interferes with spoken-word recognition in native and non-native listening. To that end, the accuracy scores for the two masking conditions and the three SNR conditions (thus without the data for the clean condition) were analysed for the native English and the non-native Dutch listeners experiments (first set of analyses, Section 3.1). The second and third analyses go beyond the general group comparison. The second analysis investigates the effect of noise on the size of the competitor space and the role of stimulus characteristics (Section 3.2) in native and non-native listening. The third analysis investigates individuals' proficiency and inhibition ability (Section 3.3) on (non-)native listening in noise.

All analyses were carried out using generalised linear mixed-effect models (Baayen, Davidson, & Bates, 2008) in R (R Development Core Team, 2008) containing both fixed and random effects, using the binomial (logit) and the Poisson family. The dependent variable in the first and third sets of analyses was the correctness (binomial distribution) of the answer to the auditory stimulus by the listeners. The dependent variable in the second set of analyses was the number of different erroneous responses or misperceptions (Poisson distribution). Fixed factors were SNR (-12, -6, 0, with -12 on the intercept), Position of Noise (onset-masked vs. offset-masked; the latter is the reference category), and Language (native English vs. non-native Dutch, with the former as the reference category)[1].

Obvious spelling mistakes and homophones were corrected such that they received the same orthographic transcription (e.g., *plain* and *plane* were both treated as correct). The stimulus *horn* was removed from all analyses since *horn* was recognised correctly less than 10% overall by the native and non-native listeners.

Each set of analyses started with building the most complex model, i.e., the model containing all main effects and interactions. We included, as much as possible, the maximal random slope and intercept structure for the random effects (cf. Barr, Levy, Scheepers, & Tily, 2013). By-stimulus slopes and by-subject slopes for SNR and SNR by Language, and by-subject slopes for Position of Noise were included in the first set of analyses. In the second set of analyses, only by-stimulus slopes were relevant; by-stimulus slopes for SNR by Language and SNR by Word Frequency were added. For the third analysis, by-stimulus slopes for SNR, LexTale, and Flanker were included as well as by-subject slopes for SNR. Subsequently, backward stepwise selection was applied in which the interaction and predictors which did not reach the significance level (5%) were removed from the model one-by-one (from both the fixed and random structure), starting with the least significant interaction or predictor. Each change in the fixed-effect and random-effect structure was evaluated using the Akaike information criterion (AIC; Aikake, 1973) which measures the relative adequacy of the model. The best model was retained, i.e., the model with the lowest

[1] To test whether bisyllabic and monosyllabic words gave different results, a factor indicating whether a stimulus word was bisyllabic or monosyllabic was included in the first and second set of analyses. The results showed that this factor did not explain any additional variance in the data, and the factor was consequently left out of all other analyses.

AIC value or if two models did not significantly differ from one another (i.e., had a difference in AIC of less than 2, see also Burnham & Anderson, 2004) the model with the fewest parameters (degrees of freedom). We checked whether leaving out the random structure from the analysis would have changed the outcomes. The changes in the fixed part turned out to be inconsequential.

< Insert Table 1 around here. >

3.1 The effect of noise on word recognition accuracy

Table 1 presents the mean word recognition accuracies and ranges for the four listening conditions for the two listener groups. Although there is quite a difference between the best and worst performing individuals in each listening condition (see the columns denoted ‘Range’ in Table 1), overall, all participants obtained (fairly) good accuracies, even in the worst listening conditions. Inspection of the given answers showed that basically all responses were either existing words or word-like responses, indicating that even the worst listening conditions did not result in a complete breakdown of the speech recognition process but instead that listeners were still able to guess.

In order to investigate the extent to which background noise interfered with spoken-word recognition, we first compared the accuracy scores of the clean condition to the accuracy scores for SNR = 0 with Language and SNR Condition (SNR 0 versus clean) and their interaction as fixed factors. The best model only contained random intercepts and by-subjects and by-items slopes for SNR (including more random slopes gave non-convergent outcomes). The statistical analyses showed, first, that the non-native listeners recognised significantly fewer words than the native listeners (β = -2.024, *SE* = .221, $p < .001$). Second, our manipulation of adding background noise resulted in a significant difference in the number of correctly recognised words between the clean and SNR 0 condition. Significantly more words were recognised in the clean condition than in the SNR 0 condition ($\beta = 2.099$, $SE = .337$, $p < .001$), and this difference was larger for the non-native listeners as shown by the significant interaction between SNR and Language ($\beta = -1.034$, $SE = .248$, $p < .001$). So, listening in noise is more difficult than when no background noise is present for both the native and the non-native listeners.

Figure 2 shows the proportion of correct responses for the three SNR conditions for the two noise positions and the clean condition separately. The line with squares shows the offset-masking results, while the line with bullets shows the results for the onset-masking condition. The diamonds indicate the results for the clean condition. As Figure 2 shows, the overall results of the native English (Figure 2a) and non-native Dutch (Figure 2b) listener groups are remarkably similar except for an overall downward shift of the proportion of correct responses for the Dutch non-native listeners. The statistical analysis (see Table 2 for the parameter estimates in the best-fitting model of performance, and for the random structure see Appendix B – Table A) indeed confirmed that the Dutch non-native listeners gave significantly fewer correct answers than the English native listeners (see the significant main effect of Language in Table 2).

< Insert Figure 2 around here. >

< Insert Table 2 around here. >

Regarding our manipulations, the analysis showed that significantly fewer words were recognised when the onset of a word was masked (bulleted line in Figure 2; see the significant main effect of Position of Noise in Table 2) compared to when the offset was masked (line with squares). Moreover, with increasingly better listening conditions, significantly more words were recognised by both the native English and the non-native Dutch listeners (see the significant main effects of SNR 0 and -6 dB in Table2). The four significant two-way interaction terms (see Table 2; all of them are ordinal, i.e., the interaction lines do not cross each other) support these conclusions, as they do not drastically change the general picture of similarity in the way Position of Noise and SNR have an impact on the two groups of listeners (see again Figure 2).

There is a significant interaction of SNR 0 dB and Position of Noise, which shows that the difference between the onset-masking and offset-masking conditions reduced significantly from the worst to the best listening condition in background noise. At the same time, the increase in proportion of correctly recognised words with increasingly improving SNRs was significantly smaller for the English native listeners (difference between SNR 0 and SNR -12 was 15.8%) compared to the Dutch non-native listeners (difference between SNR 0 and SNR -12 was 20.3%). That effect turns out to be strong enough to give a significant SNR 0 dB by Language interaction. In other words, the Dutch non-native listeners suffered more from deteriorating listening conditions than the native English listeners. The third interaction effect is Position of Noise by Language. This indicates that the detrimental effect of onset-masking was relatively smaller for the non-native Dutch listeners than for the native English listeners.

3.2 The effect of noise on multiple activation process and the competitor space

The second set of analyses investigated listeners' misperceptions in terms of the number of different erroneous answers and the influence of stimulus characteristics on the number of different erroneous answers. Importantly, the number of erroneous answers are an indication of how many (and which) candidate words were activated during listening to the auditory stimulus and how this differs for native and non-native listeners. The number of erroneous answers can thus be used as a proxy to characterise differences in the size of the competitor space for native and non-native listeners and as such answer the question whether the presence of background noise leads to a differential increase in number of simultaneously activated candidate words in native and non-native listening. Considering the larger competitor space of non-native listeners in clean listening conditions compared to native listeners and the typically larger effect of the presence of noise on speech processing for non-native listeners compared to native listeners, we would expect a larger increase in the competitor space for the non-native compared to the native listeners, even after accounting for how often a word was misrecognised (error rate) and the number of times a word was presented in the experiment (remember, due to the random selection of stimulus lists, the different stimulus words had unequal presentation numbers in the full experiment). More errors lead to more erroneous answers (tokens), which in turn (could) lead to more different erroneous words (types), which could fully explain the distinction between native and non-native listeners. If indeed the non-native listeners show a larger increase in competitor space than the native listeners under the influence of the presence of background noise, this would show itself as an interaction between Language and SNR.

The dependent variable in this analysis is the number of different erroneous answers given to a specific word stimulus in a specific SNR condition and a specific Position of Noise condition by the native and the non-native listeners. Since only existing words can be considered as candidate words and be in competition for recognition, answers that were non-words were left out of the analyses. Stimulus was added as a random factor. Word frequency has been suggested to play an important role in native spoken word recognition in clean listening conditions (Luce & Pisoni, 1998). The importance of word frequency in native and non-native listening during the recognition of speech masked by noise was investigated by entering Word Frequency (centred and Zipf-transformed) as a fixed effect in the analyses in addition to Language, SNR, Position of Noise, Error Rate, and Times Presented, the latter two varying for each word in each SNR x Position of Noise x Language condition.

< Insert Table 3 around here. >

Table 3 lists the frequency counts of the number of different erroneous answers (i.e., the dependent variable) for a certain word stimulus in all six Position of Noise by SNR conditions for the native English and the non-native Dutch listeners. Different answers higher than ‘0’ give the number of different erroneous answers. The Dutch listeners recognised fewer stimuli correctly than the native English listeners; moreover, their maximum number of different responses is ‘10’, whereas the maximum for the English native listeners is ‘6’. The non-native listeners thus seem to give more different erroneous answers than the native listeners. Given the count distribution of the dependent variable, a Poisson analysis was used for the subsequent analysis of the misperceptions data.

< Insert Figure 3 around here. >

Figure 3 shows the mean number of different erroneous answers for the three SNR conditions and the clean conditions plotted for the onset-masking (light grey, top line) and offset-masking (darker grey, bottom line) conditions separately. The black bullet point indicates the results for the clean condition. The results for the native English listeners are plotted in Figure 3a, those for the Dutch non-native listeners in Figure 3b.

< Insert Figure 4 around here. >

Figure 4 shows the relation between error rate and the number of different responses for the native listeners (white bars) and the non-native listeners (grey bars), respectively. The error bars indicate one standard error above and one standard error below the mean. The error rates are binned into bins of .2 proportion of errors. Figure 4 shows increasing error rates in combination with an increasing number of different erroneous answers in both listener groups. The first bin does not include the cases where no misperceptions occurred. Please note that, unlike for the calculation of the number of different erroneous answers, misrecognitions that did not result in an existing word were taken into account when calculating the error rate.

< Insert Table 4 around here. >

Table 4 shows the fixed effect estimates for the best-fitting model of performance for the native versus non-native analysis of the number of different answers (see Appendix B – Table B for the random structure). All main effects were significant, including Language, with an upward shift for the Dutch non-native listeners. The Dutch non-native listeners gave significantly more different answers than the English native listeners (see the factor Language in Table 4). For example, the word *beetle* was recognised as *beetle*, *beaten*, and *people* by the native English listeners. These responses were also given by the Dutch non-native listeners, but their responses also included *beacon*, *beat*, and *peter*.

The analysis of the misperception data showed that the number of different answers is dependent on Position of Noise and SNR and that their impact did not differ between the native and non-native listeners (no interactions with Language). Significantly more different answers were given when the onset of the word was masked compared to when the offset of the word was masked. In other words, the number of competitors was significantly higher when onset information was no longer reliably available compared to the number of competitors when offset information was no longer available. This result is in line with the results of the previous main analysis, and suggests that word-initial information is more important for successful word recognition than word-final information. Moreover, the number of different erroneous answers increased significantly (main effect: SNR) with deteriorating listening conditions.

The effect of word frequency was the same for both listener groups: overall, a higher word frequency of the word stimulus led to significantly fewer different misperceptions. The significant two-way interaction between Position of Noise and Word Frequency showed that the size and direction of the effect of word frequency on the number of different answers was dependent on the position of the noise masker. Further inspection of the interaction showed that where for the offset-masked words a higher frequency led to a decrease in the number of different misperceptions for SNR -6 and SNR -12, a higher word frequency led to an increase in the number of different erroneous answers for the SNR -6 and SNR -12 conditions when the onset of the word was masked. This suggests that when word-initial information is less reliable due to the presence of noise, listeners are less able to use word frequency to reduce the number of activated candidate words during spoken-word recognition.

Importantly, Table 5 clearly shows the pervasive effect of error rate. It has a strong main effect, and is also involved in many interaction effects. An increasing error rate (due to increasingly more difficult listening conditions) leads to an increase in the number of different answers, thus an increase in the number of competitors. The number of times presented has only a main effect and is not involved in any interactive relationship with one of the other predictors. Although both covariates modulate the number of different erroneous answers in native and non-native listening, they do not take away the language effect (as there is also a strong Language main effect). Moreover, the language effect is strengthened by the error rate by language interaction, meaning that the number of competitors increase more rapidly with an increasing number of errors in non-native than in native listeners. Thus, the non-native competitor space expands more rapidly with worse listening conditions, an effect moderated by more erroneous answers. The competitor space between native and non-native listeners thus is different (as indicated by the main effect of Language), with a larger competitor space for the non-native listeners even after accounting for the non-natives' higher error rate.

3.3 Individual differences

In order to investigate whether inhibition ability and non-natives' proficiency influence the effect of noise on (non-)native spoken word recognition, the third analysis was carried out. The analysis is similar to the analysis in the first analysis in which the data of the native English and the non-native Dutch listeners were analysed. However, here the data of the native English and non-native Dutch listeners are analysed separately. Moreover, in addition to the previously used factors Position of Noise, SNR, and Language, Flanker Cost was added to the analyses. For the non-native analysis, also LexTale was added. Both additional variables were scaled and centred to the mean. Three native and three non-native participants were excluded from the analysis due to missing data in some of the individual differences measures.

The results of the native analysis showed no significant effect for Flanker Cost. The results of the individual differences analysis of the non-native Dutch listeners are displayed in Table 5 (for the random effects structure, please see Appendix B – Table C).

<Insert Table 5 around here. >

The effects of Position of Noise and SNR remain the same as those reported in Table 2. It means that we can focus on those findings that are related to individuals' proficiency and inhibition ability. There was a simple effect of LexTale: non-native listeners with a higher proficiency in English recognised more words (see factor LexTale in Table 5). In contrast to what was found for the native listeners, inhibition ability also influenced word recognition performance for the non-native listeners. The interaction between Position of Noise and Flanker Cost showed that when word-initial information was masked, listeners with poorer inhibition ability (i.e., higher Flanker Cost) recognised more words correctly than listeners with better inhibition ability. When word-final information was masked, the role of inhibition ability is much smaller, and actually goes slightly in the opposite direction compared to that for the word-initial masking condition. This effect is shown in Figure 5 where the solid line indicates the offset-masking condition and the dashed line indicates the onset-masking condition. So, listeners who have better inhibition abilities (= lower scaled Flanker scores) have a lower proportion of correct answers than listeners with worse inhibition abilities (= higher scaled Flanker scores). The maximal random slope structure included a by-stimulus random slope for SNR and LexTale indicating that the effects of SNR and proficiency level vary over the stimuli.

4 General discussion

This paper investigates two questions: 1) does the presence of background noise lead to a differential increase in the number of simultaneously activated candidate words in native and non-native listening; 2) do individual differences in listeners' cognitive and linguistic abilities explain the differential effect of background noise on (non-)native speech recognition? We will discuss our findings regarding these two research questions in the subsections below. Although not the main goal of our work, the results of this study also inform theories on the role of word-initial and word-final information in spoken-word recognition. Existing evidence primarily focuses on native listening. Our results provide new insights into the role of word-initial and word-final information in non-native listening, which we discuss in a separate subsection below.

4.1 The effect of noise on native and non-native listening

In line with previous results (e.g., Bradlow & Alexander, 2007; Brouwer et al., 2012; Cooke et al., 2008; Ezzatian et al., 2010; Kilman et al., 2014; Mayo et al., 1997; Meador et al., 2000; Van Wijngaarden et al., 2002), we found that the recognition of spoken words in the presence of background noise was more difficult in a non-native language than in one's native language. Moreover, we observed the differential effect of noise on native and non-native listening previously observed (e.g., Cooke et al., 2008; Mayo et al., 1997; Meador et al., 2000): non-native listeners gave significantly more and more different erroneous answers than native listeners when noise was present in the background. Using the number of different responses of the listener groups as a proxy of the size of the competitor space, we can then observe that the competitor space is larger during non-native listening compared to native listening when listening in the presence of background noise. The observed differences in recognition accuracy and the size of the competitor space, however, do not seem to be caused by a differential effect of the presence of background noise on the mechanisms underlying spoken-word recognition as evidenced by the overall strong similarities in response patterns and factors that play a role in native and non-native spoken-word recognition in background noise.

The presence of noise in the speech signal decreases the phonological match between the to be recognised word and the candidate words in the native and non-native listeners' mental lexicon. This decrease in phonological match can, e.g., be due to a masking of parts of the speech signal by the noise or to the reallocation of acoustic cues from the background signal to the speech signal (see e.g., Cooke, 2009). Similar to what has been observed for the effect of imperfect sound perception on the multiple activation process in non-native listening, the decrease in phonological match between the acoustic signal and the candidate words results in an increase in the number of words with which the acoustic signal has a partial match (referred to as a *many-additional-competitor scenario* by Chan & Vitevitch, 2015) rather than an increase in the activation of one erroneous word (referred to as the *single-strong-competitor scenario* by Chan & Vitevitch, 2015). This finding persists even after controlling for error rate and the number of times a stimulus was presented in the experiment, suggesting that the phonological mismatch due to a degraded speech signal increases the competitor space in both native and non-native listening; although, it seems to be larger for non-native listeners.

There are several possible explanations, which are not necessarily mutually exclusive, for the observed differential effect of the presence of background noise on native and non-native listeners, and the difference in magnitude of the effect of the different factors influencing native and non-native spoken-word recognition in the presence of background noise. It is likely that the native speech perception system is relatively more robust against the presence of background noise due to native listeners' better ability to use low-level and high-level contextual information to compensate for a loss of information at the sound processing levels during speech recognition (see e.g., Bradlow & Alexander, 2007; Cutler, Garcia Lecumberri, & Cooke, 2008) compared to non-native listeners. Secondly, non-native listeners have been shown to activate a larger number of spurious word candidates than native listeners (e.g., Cutler et al., 2006; Spivey & Marian, 1999; Weber & Cutler, 2004). Using a visual-world eye-tracking paradigm these studies showed that due to inaccurate sound perception, non-native listeners not only look at the picture of the target word but also to distractor pictures of words with sounds that are confusable for the non-native listener (e.g., /ɛ/-/ æ/ for the Dutch non-native listeners of English in Weber & Cutler, 2004 or /l/-/r/ for the Japanese non-native listeners of English in Cutler et al., 2006), while native listeners were not found to look at those distractor pictures. A third possible explanation for the larger effect of noise during non-native listening could be that the competition process is less efficient and slower in suppressing these spurious, incorrect word candidates. For instance, due to non-native listeners simply being less certain in the different processes underlying the spoken-word recognition chain under all condition (not just noisy ones) and this uncertainty could be exacerbated by noise more for the non-native listeners than the native listeners. For instance, an increased uncertainty due to a decreased phonological match or a delay in resolving the competition might lead to an increase in the number of activated word candidates resulting in a less efficient competition phase. This explanation would be in agreement with the accumulating evidence that noise has a deteriorating effect on all speech processing levels during non-native listeners (see e.g., Bradlow & Alexander, 2007). All these explanations, though, suggest that the observed difference between native and non-native listeners in our study should then not be primarily attributed to differences in the effect of noise on spoken-word recognition, but rather that it is primarily caused by a basic difference between native and non-native listening. This conclusion ties in with findings for bilingual listeners: Early bilinguals with near-native proficiency perform equally well as monolingual listeners in quiet but they suffer more than monolingual listeners from the presence of background noise (e.g., Krizman, Bradlow, Lam, & Kraus, 2016; Shi, 2010), due to improved language-dependent processes for native listeners compared to non-native listeners (Krizman et al., 2016). It is likely this finding extends to late bilinguals and non-native listeners.

4.2 The effect of listeners' cognitive and linguistic abilities on (non-)native speech recognition in the presence of background noise

We observed differences within the group of non-native listeners regarding the effect of masking on word recognition accuracy which seem to be related to individuals' proficiency in English and, to a lesser extent, their inhibition ability. As expected, non-native Dutch listeners who had a higher proficiency in English made fewer recognition errors. There is considerable evidence showing the importance of proficiency in the non-native language for successful non-native spoken-word recognition. These studies however typically use group comparisons in which low and high proficient and/or native listener performances are compared (see, e.g., Brouwer et al., 2012; Mayo et al., 1997). As far as we are aware, our study is only the second study (Kilman et al., 2014, being the first) which investigated individuals' proficiency in a non-native language on the effect of background noise. Our results are in line with those of Kilman and colleagues (2014), (again) showing the importance of listener's proficiency in the non-native language on successful non-native spoken-word recognition. We agree with Kilman and colleagues that future investigations into the effect of background noise on non-native listening should thus always include objectively measured individuals' non-native proficiency levels. Proficiency in the non-native language however did not explain any variation related to the accuracy of spoken-word recognition in the presence of background noise (as shown by a lack of an interaction between proficiency and SNR or the position of the noise masker).

We hypothesised listeners with poorer inhibition abilities to be worse at inhibiting active candidate words during the competition process than listeners who are better inhibitors, resulting in poorer performance in speech perception in noise for poorer inhibitors in comparison to better inhibitors who will be faster in reducing their competitor space. This is not what we found. We found no effect of inhibition abilities on native spoken-word recognition in noise. For the non-native listeners, our results showed that when word-initial information was masked, listeners with poorer selective attention were found to have *higher* word recognition accuracies than listeners with better selective attention abilities. A possible explanation might be as follows. Inhibition builds up over time (Ridderinkhof, 2002, Van den Wildenberg et al., 2010). Possibly, as hypothesised, listeners who are poorer inhibitors need more time to remove competitors from the active set of candidate words compared to better inhibitors. Poorer inhibitors thus have a larger competitor set than better inhibitors. This might actually be beneficial when listening in degraded listening conditions. If the onset of a word is less reliably available, for instance due to the presence of background noise, better inhibitors might already have filtered the target word from their competitor space, while poorer inhibitors, who keep their options open for a longer period of time, might still have the target word in their competitor space. This allows poorer inhibitors to compare the later (intact) incoming input with the activated candidate words, including the target word, increasing their chance of recognising the target word. This reasoning also explains the lack of an effect of inhibition abilities for the native listeners. The analysis of the competitor space of native listeners versus non-native listeners indicated that native listeners have a smaller competitor space than non-native listeners. In line with the earlier argument explaining the difference in the effect of masking of word-initial information, possibly the correct word has already dropped out of the smaller competitor space of the native listeners. Consequently, it is less likely to find an effect of inhibition control on native word recognition performance as the correct word has already disappeared from the competitor space. Our finding ties in with that of Banks and colleagues (2015), who investigated the role of inhibition on the adaptation to accented speech in the presence of background noise by native listeners. Although they found an effect of inhibition abilities on adaptation this effect could not be linked to speech recognition in noise. In short, individuals are differently impacted by the presence of noise as observed through the role of inhibitory abilities for the non-native listeners on speech recognition in noise. This will have its implications for estimating the listening capacities of second language learners but also for the interpretation of the results of audiological tests when taken in a non-native language.

4.3 The role of lexical information on (non-)native speech recognition in the presence of background noise

The noise manipulation of the stimuli that was used in this study allows us to analyse the results in terms of the importance of word-initial and word-final information for native and non-native spoken-word recognition in the presence of background noise. In line with previous studies (e.g., Allopenna et al., 1998, Coumans, van Hout, & Scharenborg, 2014; McQueen & Huettig, 2012; Scharenborg, Coumans, Kakouros, & van Hout, 2016; Slowiaczek et al., 1987; van der Vlugt & Nooteboom, 1986), word-initial information seems to be more important for successful word recognition than word-final information. The results showed that the masking of word-initial information was more detrimental to spoken-word recognition than the masking of word-final information, leading to lower word recognition accuracies and higher numbers of different erroneous answers. This finding cannot be explained by a difference in the number of neighbours (words with a higher neighbourhood density are typically recognised slower and less accurate than words with a less dense neighbourhood, Luce & Pisoni, 1998) as the words with word-initial masking and those with word-final masking had on average a similar number of neighbours. Nor can this difference be explained by differences in the number of word-initial or word-final competitors for the stimulus words: potentially, if the word-initially masked words had on average more word-final competitors than the word-finally masked words had word-initial competitors, then masking the word-initial information would leave more possibilities for making an error than when word-final information is masked. However, the number of word-initial and word-final competitors for the stimuli in our experiment was highly similar. This explanation is therefore unlikely. The decreased intelligibility of word-initial information due to the presence of noise seems to result in the activation of more candidate words than when word-final information is less intelligible. This suggests that word-initial information is more important than word-final information in restricting the number of activated lexical candidates or word-initial information is weighted more heavily than word-final information during the later stages of spoken-word recognition in degraded listening conditions. This is in agreement with phonological theory and findings from phonetic studies. Van Son and Pols (2003) showed using a large corpus of Dutch speech that word onsets are more clearly articulated and that word-initial segments have a higher “segmental information content”, i.e., they contribute more to word-disambiguation than segments later in the word. Segments later in the word are more often reduced and carry less segmental information content, these segments are often redundant and are not needed for word recognition. Moreover, according to phonological theory, the ‘sonority profile’ states that the sonority of the syllable increases from the start of the syllable onwards (Clements, 1990; Gussenhoven & Jacobs, 2011). Consequently the onset of syllables is maximally salient. Gussenhoven and Jacobs (2011) argue that the prominence of the onset of syllables (or words) attracts the attention of the listener probably for perceptual ease. Although we did not pit word-initial and word-final information directly against one another using a visual-world eye-tracking paradigm, our results did show a change in the dynamics of the multiple activation and competition processes. Our results show a decrease in word recognition accuracy and, importantly, a clear increase in the number of different erroneous answers when listening conditions deteriorate compared to the clean condition. So, not only do we observe an increase in the number of onset competitors (in line with Ben-David et al., 2010), we also observed an increase in the number of offset competitors (in line with the older listeners in Ben-David et al., 2010; Brouwer & Bradlow, in press; McQueen & Huettig, 2012). The presence of background noise thus not only seems to increase the activation levels of competitor word candidates, it also increases the number of candidate words.

Although the native listeners outperformed the non-native listeners in all listening conditions, the size of the detrimental effect of the masking of word-initial information was found to be larger for the native English listeners than for the non-native Dutch listeners: the difference between the onset and offset masking conditions increased more with deteriorating listening conditions for the English native listeners than for the Dutch non-native listener group. In line with the reasoning above, possibly, non-native listeners keep more candidate words that match with either the word's onset or its offset alive than native listeners do when noise is present in the background. As the correct word might still be kept active and thus can still be retrieved at a later point in time, the effect of noise on word-initial information, although worse than the effect of noise on word-final information, does not further deteriorate during non-native listening.

The number of different erroneous answers was modulated by word frequency in both native and non-native listening: a higher word frequency led to significantly fewer different misperceptions, and thus words with higher frequencies were easier to recognise than words with lower word frequencies. This finding adds to ample evidence that word frequency plays an important role in native spoken word recognition in clean listening conditions (e.g., Marslen-Wilson, 1987; Luce & Pisoni, 1998) and in degraded listening conditions (Luce, 1986; Goldinger, Luce, & Pisoni, 1989). The role of word frequency has been found to be smaller in non-native listening compared to that in native listening (Bradlow & Pisoni, 1999; Imai, Walley, & Flege, 2005). Our results, however, show that the role of word frequency in non-native listening is similar to that in native listening. These differences might however be due to differences in the range of word frequencies and selection of the materials used in our study compared to the other studies.

There is however a differential effect of word frequency on the recognition of words with onset-masking versus words with offset-masking. When word-initial information is less reliably available due to the presence of noise, listeners are less able to use word frequency to reduce the number of activated candidate words during spoken-word recognition, suggesting that word-frequency information is (more strongly) tied to word onset. This finding ties in with theories on the locus of word frequency in speech processing. Dahan and colleagues (Dahan, Magnuson, & Tanenhaus, 2001) found, using a visual-world eye-tracking paradigm, that listeners were faster to fixate a picture of a high-frequency word than a picture of word with an identical onset but with a lower frequency. They concluded that word frequency has its effect right at the earliest moments of lexical access. Word onset thus not only contains a lot of information, it is also highly important for word frequency to affect lexical activation. If word-initial information is masked, word frequency information is not readily available to the spoken-word recognition process, which will slow down the winnowing of the competitor set. Our results seem to suggest that if word-initial information is not reliably available, the word frequency information can no longer be (properly) used as an information source to guide lexical access and restrict the number of activated candidate words. This is the case for both native and non-native listeners.

4.4 Concluding remarks

Taking all results together, a clear picture emerges of the effect of the presence of background noise on native and non-native spoken-word recognition. The presence of noise in the speech signal seems to result in a change in the dynamics of the multiple activation and the competition and recognition processes underlying spoken-word recognition. This noise effect seems to be fundamentally similar for native and non-native listeners (and probably also early-bilinguals and late-bilinguals, see e.g., Krizman et al., 2016), the difference being one of degree: the observed lower performance of non-native listeners compared to native listeners is caused by a difference in magnitude of the effect of the different factors influencing native and non-native spoken-word recognition in the presence of background noise. In addition to an effect of background noise on the multiple activation process, background noise also has an effect on the competition process. Listeners are able to flexibly adjust their reliance on word-initial and word-final information when a change in listening conditions demands it. So the extent to which candidate words are considered for recognition is dependent on the listening conditions. This finding adds to earlier, similar findings by, e.g., Brouwer, Mitterer, & Huettig (2011), Farris-Trimble, McMurray, Cigrand, & Tomblin (2014), and McQueen & Huettig (2012). These flexible adaptations to the processes underlying spoken-word recognition enable native and non-native listeners to partially overcome the degrading effect of background noise at various levels of the spoken-word recognition process.

It is expected that current (computational) models of spoken-word recognition (such as Cohort (Marslen-Wilson & Tyler, 1980), TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994), and Fine-Tracker (Scharenborg, 2010)) are able to model the reduced phonological matching due to the presence of noise. The effect of background noise on the competition process could be modelled by a slowing down of the build-up of the word activations of the candidate words considered for recognition over time, and a less strict elimination process for the removal of candidate words from the set of activated words (e.g., by a less strict mismatch parameter as in Shortlist). Follow-up visual-world eye-tracking experiments and computational modelling studies can shed light onto the question what the mechanisms are underlying the enhanced phonological competition, e.g., whether the observed enhanced phonological competition can be explained by a slowing down of the build-up of word activation over time, a less efficient elimination of candidate words that no longer have a good enough fit with the speech signal or that other mechanisms are at play.

To summarise, native listeners outperformed non-native listeners in all listening conditions. Nevertheless, the effect of background noise on the multiple activation process is remarkably similar in native and non-native listening. The presence of noise reduces the phonological match between the input and the stored candidate words and consequently increases the set of candidate words considered for recognition during spoken-word recognition in both native and non-native listening. The results indicate that the observed performance differences between the English and Dutch listeners should not be primarily attributed to a differential effect of noise, but rather to the difference between native and non-native listening. Possibly, during non-native listening not only more candidate words are activated compared to native listening, but also non-native listeners might have more difficulty eliminating possible word candidates early during the recognition process. Proficiency and inhibition ability have been found to influence non-native spoken-word recognition in noise with a higher proficiency in the non-native language and worse inhibition ability leading to improved non-native spoken-word recognition in noise.

Acknowledgements

This research is supported by a Vidi-grant from the Netherlands Organization for Scientific Research (NWO; grant number 276-89-003) awarded to Odette Scharenborg. Juul Coumans is now at the Brain Center Rudolf Magnus, University Medical Center Utrecht. We would like to express our gratitude to Sven Mattys for generously giving us access to his testing facilities at the University of York, UK, Polina Drozdova for running the native English experiments, Alastair Smith for recording the stimuli, and Joop Kerkhoff for creating the Praat script.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language, 38(4), 419-439. doi:10.1006/jmla.1997.2558.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B.N.; Csáki, F., 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971, Budapest: Akadémiai Kiadó, p. 267-281.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. Trends in Cognitive Sciences, 8(4), 170–177.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items, Journal of Memory and Language, 59(4), 390-412. doi 10.1016/j.jml.2007.12.005.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX Lexical Database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015) Cognitive predictors of perceptual adaptation to accented speech. Journal of the Acoustical Society of America, 137(4), 2015-2024.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68, 255-278.

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, K. M., Reingold, E. M., & Schneider, B. (2011). Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements. Journal of Speech, Language and Hearing Research, 54, 243-262.

Bialystok, E. (2001). Bilingualism in Development: Language, Literacy and Cognition. Cambridge:Cambridge University Press. doi:10.1017/CBO9780511605963.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. Psychology and Aging, 19, 290–303. doi:10.1037/0882-7974.19.2.290.

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(4), 2339-2349.

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener- and item-related factors. Journal of the Acoustical Society of America, 106(4), 2074-2085.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/.

Bohn, O.-S., & Munro, M. J. (Eds.). (2007). Language experience in second language speech learning: In honor of James Emil Flege. Amsterdam: John Benjamins.

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. Language and Cognitive Processes, 27, 1205-1224.

Broersma, M., & Scharenborg, O. (2010). Native and non-native listeners’ perception of English consonants in different types of noise. Speech Communication, 52(11-12), 980-995. doi:10.1016/j.specom.2010.08.010.

Brouwer, S., & Bradlow, A.R. (in press). The temporal dynamics of spoken word recognition in adverse listening conditions. Journal of Psycholinguistic Research.

Brouwer, S., van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners:

language familiarity and semantic content. Journal of the Acoustical Society of America, 131(2), 1449–1464.

Brouwer, S., Mitterer, H., & Huettig, F. (2011). Speech reductions change the dynamics of competition during spoken word recognition. Language and Cognitive Processes, doi:10.1080/01690965.2011.555268.

Burnham, K.P. & Anderson, D.R. (2004). Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods & Research, 33, 261. Doi: 10.1177/0049124104268644.

Chan, K.Y., & Vitevitch, M. S. (2015). The Influence of Neighborhood Density on the Recognition of Spanish-Accented Words. Journal of Experimental Psychology: Human Perception and Performance, 41(1), 69–85. doi:10.1037/a0038347.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston and M. E. Beckman (eds), Papers in laboratory phonology I. Between the grammar and physics of speech. Cambridge: Cambridge University Press, 283–333.

Cooke, M. (2009). Discovering consistent word confusions in noise. Proceedings of Interspeech, Brighton, UK.

Cooke, M., García Lecumberri, M. L., & Barker, J. (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(1), 414-427, 2008. doi:10.1121/1.2804952.

Cooke, M., García 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. doi:10.1016/j.specom.2010.04.004.

Coumans, J., van Hout, R., & Scharenborg, O. (2014). Non-native word recognition in noise: The role of word-initial and word-final information. Proceedings of Interspeech, Singapore, pp. 519-523.

Cutler, A. (2012). Native listening: The flexibility dimension. Dutch Journal of Applied Linguistics, 1(2), 169-187. doi:10.1075/dujal.1.2.02cut.

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(2), 1264-1268. doi:10.1121/1.2946707

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. Journal of Phonetics, 34(2), 269-284. doi:10.1016/j.wocn.2005.06.002.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. Cognitive Psychology, 42, 317–367.

Drozdova, P., van Hout, R., & Scharenborg, O. (2015). The effect of non-nativeness and background noise on lexical retuning. Proceedings of the International Congress of Phonetic Sciences, Glasgow, UK.

Eriksen, B. A., & Eriksen, C. W. (1974). Effect of noise letters on upon the target identification in a non-search task. Perception and Psychophysics, 16, 143-149.

Ezzatian, P., Avivi-Reich, M., & Schneider, B. A. (2010). Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking? Speech Communication, 52(11), 919-929.

Farris-Trimble, A., McMurray, B., Cigrand, N., & Tomblin, J. B. (2014). The process of spoken word recognition in the face of signal degradation. Journal of Experimental Psychology: Human Perception and Performance, 40(1), 308-327.

Garcia Lecumberri, M. L., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. Speech Communication, 52, 864-886. doi:10.1016/j.specom.2010.08.014.

Gow, D. W., & Gordon, P.C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. Journal of Experimental Psychology: Human Perception and Performance, 21(2), 344-359. doi:0.1037//0096-1523.21.2.344.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. Journal of Memory and Language, 28**,** 501–518.

Gussenhoven, C., & Jacobs, H. (2011). Understanding phonology, 3rd edition. Routledge.

Hintz, F., Scharenborg, O. (2016). The effect of background noise on the activation of phonological and semantic information during spoken-word recognition. Interspeech 2016, San Francisco, CA, 2816-2820.

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish accented words by native English and Spanish listeners. The Journal of the Acoustical Society of America, 117, 896–907. http://dx.doi.org/10.1121/1.1823291.

Kilman, L., Zekveld, A. A., Hällgren, M., & Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. Frontiers in Psychology, 5, 651. doi: 10.3389/fpsyg.2014.00651.

Krizman, J., Bradlow, A. R., Lam, S. S.-Y., & Kraus, N., (2016). How bilinguals listen in noise: linguistic and non-linguistic factors. Bilingualism: Language and Cognition, Available on CJO 2016 doi:10.1017/S1366728916000444 R.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. Behavioral Research Methods, 44, 325-343. doi:10.3758/s13428-011-0146-0.J.

Luce, P. A. (1986). Neighborhood of words in the mental lexicon. Research on speech perception, Technical report No. 6. Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

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

Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. Cognition,25**,** 71–102.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. Cognition, 8, 1–71.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: the importance of word onsets. Journal of Experimental Psychology: Human Perception and Performance, 15(3), 576-585.

Mayo, L. H., Florentine, M., & Buss, S. (1997). Age of second-language acquisition and perception of speech in noise. Journal of Speech, Language and Hearing Research, 40, 686–693.

McClelland, J. L., & Elman, J. L. (196). The TRACE model of speech perception. Cognitive Psychology, 18, 1-86.

McQueen (2005). Speech perception. In: The Handbook of Cognition (Eds. Lamberts, Goldstone), Sage, London, 255-275.

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognised. Journal of the Acoustical Society of America, 131(1), 509-517. doi:10.1121/1.3664087.

Meador, D., Flege, J. E., & Mackay, I. R. A. (2000). Factors affecting the recognition of words in a second language. Bilingualism: Language & Cognition, 3(1), 55-67.

Nábêlek, A. K., Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. Journal of the Acoustical Society of America, 75, 632-634.

Norris D. (1994). Shortlist: a connectionist model of continuous speech recognition. Cognition, 52, 189-234.

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. Psychological Science, 12, 445-449.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Retrieved from http://www.R-project.org.

Ridderinkhof, K. R. (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. In W. Prinz & B. Hommel (Eds.), Attention and performance XIX: Common mechanisms in perception and action (pp. 494–519). Oxford: Oxford University Press.

Scharenborg, O., Hintz, F., & McQueen, J. M. (in preparation). Investigating the relative importance of word-initial and word-final information in adverse listening conditions.

Scharenborg, O., Kolkman, E., Kakouros, S., Post, B. (2016). The effect of sentence accent on non-native speech perception in noise. Interspeech 2016, San Francisco, CA, 863-867.

Schock, J. Cortese, M. J., Khanna, M. M., & Toppi, S. (2012). Age of acquisition estimates for 3,000 disyllabic words. Behavior Research Methods, 44(4), 971–977. doi:10.3758/s13428-012-0209-x.

Shi, L-F., (2010). Perception of acoustically degraded sentences in bilingual listeners who differ in age of English acquisition. Journal of Speech, Language, and Hearing Research, August 2010, Vol. 53, 821-835. doi:10.1044/1092-4388(2010/09-0081)

Shimoda, D. (2010). Scrabblefinder. Retrieved from http://www.scrabblefinder.com.

Sikora, K., Roelofs, A., Hermans, D., & Knoors, H. (2016). Executive control in spoken noun-phrase production: Contributions of updating, inhibiting, and shifting. Quarterly Journal of Experimental Psychology.

Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Phonological priming in auditory word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13(1), 64-75. doi:10.1037//0278-7393.13.1.64.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 6(2), 174-215. doi:10.1037/0278-7393.6.2.174.

Spivey, M., & Marian, V. (1999). Cross talk between native and second languages: partial activation of an irrelevant lexicon. Psychological Science, 10, 281–284.

Van den Wildenberg, W. P. M., Wylie, S. A., Forstmann, B. U., Burle, B., Hasbroucq, T., & Ridderinkhof, R. K. (2010). To head or to heed? Beyond the surface of selective action inhibition: A review. Frontiers in Human Neuroscience, 4, 1–13.

van der Vlugt, M. J., & Nooteboom, S. G. (1987). Auditory word recognition is not more sensitive to word-initial than to word-final stimulus information. Journal of the Acoustical Society of America, 81(S2). doi:10.1121/1.2024178.

Van Son, R.J.J.H., & Pols, L.C.W. (2003). How efficient is speech? Proceedings of the Institute of Phonetic Sciences, 25, 171-184.

Van Wijngaarden, S. J., Steeneken, H. J., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. Journal of the Acoustical Society of America, 111(4), 1906-16.

Weber A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. Journal of Memory and Language, 50(1), 1-25. doi:0.1016/S0749-596X(03)00105-0.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. Cognition, 32(1), 25-64. doi:10.1016/0010-0277(89)90013-9.

Table 1. Word recognition accuracies for the two listener groups in clean and in the three SNR condition. Range indicates the worst and best individual's scores.

	Clean		SNR 0		SNR -6		SNR -12	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Native listeners	99.3 (8.2)	94.0-100.0	95.7 (20.2)	81.5-100.0	87.8 (32.7)	64.3-100.0	79.9 (40.1)	63.0-96.4
Non-native listeners	92.2 (26.8)	75.9-98.8	83.8 (36.8)	64.3-100.0	74.1 (43.8)	48.1-96.4	63.5 (48.2)	37.0-82.1

Table 2. Fixed effect estimates for the best-fitting models of performance for the native English (on the intercept) and non-native Dutch listeners on the stimuli presented in noise, n = 9211 observations.

Fixed factor	**B**	***SE***	***P***
Intercept	2.665	.206	< .001
Position of Noise	-1.219	.155	< .001
SNR 0 dB	2.149	.275	< .001
SNR -6 dB	.826	.195	< .001
Language	-1.448	.154	< .001
Position of Noise × SNR 0 dB	.455	.209	.029
Position of Noise × SNR -6 dB	.275	.169	.104
Position of Noise × Language	.389	.144	.007
SNR 0 dB × Language	-.734	.201	< .001
SNR -6 dB × Language	-.167	.161	.300

Table 3. Frequency counts of the number of different responses for all distinct stimuli (83) in all six Position of Noise by SNR conditions for the native English and the non-native Dutch listeners.

#different erroneous answers	Native English	Non-native Dutch
0	326	197
1	116	146
2	30	79
3	16	39
4	7	15
5	1	15
6	2	3
8	0	2
10	0	2

Table 4. Fixed effect estimates for the best-fitting model of performance for the analysis of the number of different answers for the English and Dutch listeners, with error rate and the number of times the stimulus was presented as covariates, n = 996 observations.

Fixed factor	**B**	***SE***	***p***
Intercept	-1.884	.184	< .001
Position of Noise	.342	.126	.007
SNR 0 dB	-1.017	.157	< .001
SNR -6 dB	-.274	.133	.040
Error Rate	2.924	.293	< .001
Language	.623	.132	< .001

Word Frequency	-.390	.127	.002
Times Presented	.067	.012	< .001
Error Rate × Position of Noise	-.557	.268	.038
Error Rate × SNR 0 dB	1.280	.299	< .001
Error Rate × SNR -6 dB	.246	.227	.278
Error Rate × Word Frequency	.554	.191	.004
Error Rate × Language	.763	.216	< .001
Position of Noise × Word Frequency	.315	.127	.013

Table 5. Fixed effect estimates for the best-fitting model of performance for the individual differences analysis of the non-native listeners, n = 4812 observations.

Fixed factor	**β**	***SE***	***P***
Intercept	1.165	.179	< .001
Position of Noise	-.728	.113	< .001
SNR 0 dB	1.790	.196	< .001
SNR -6 dB	.851	.125	< .001
Flanker Cost	-.009	.074	.904
LexTale	.194	.066	.004
Position of Noise × Flanker Cost	.170	.082	.037

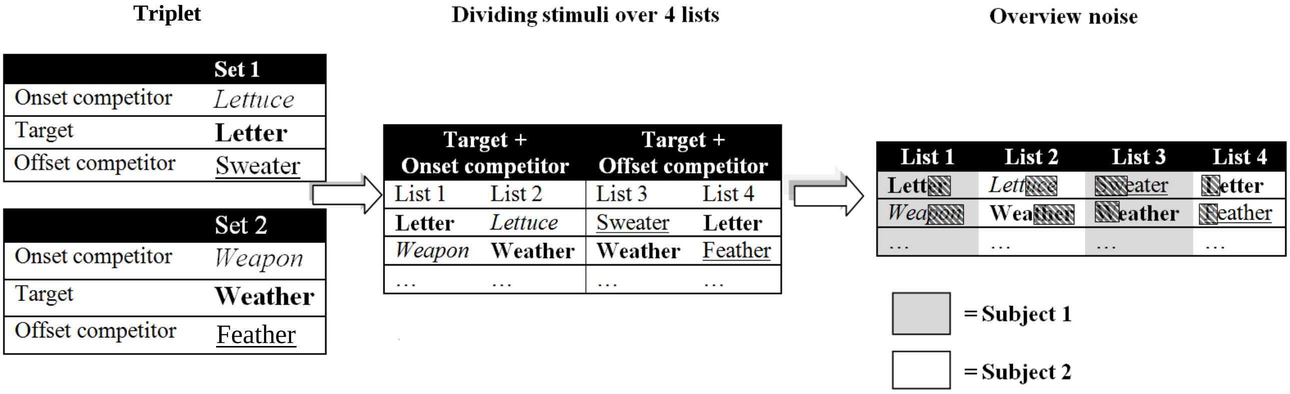

Figure 1. Noise-masking procedure and division of the stimuli over the lists.

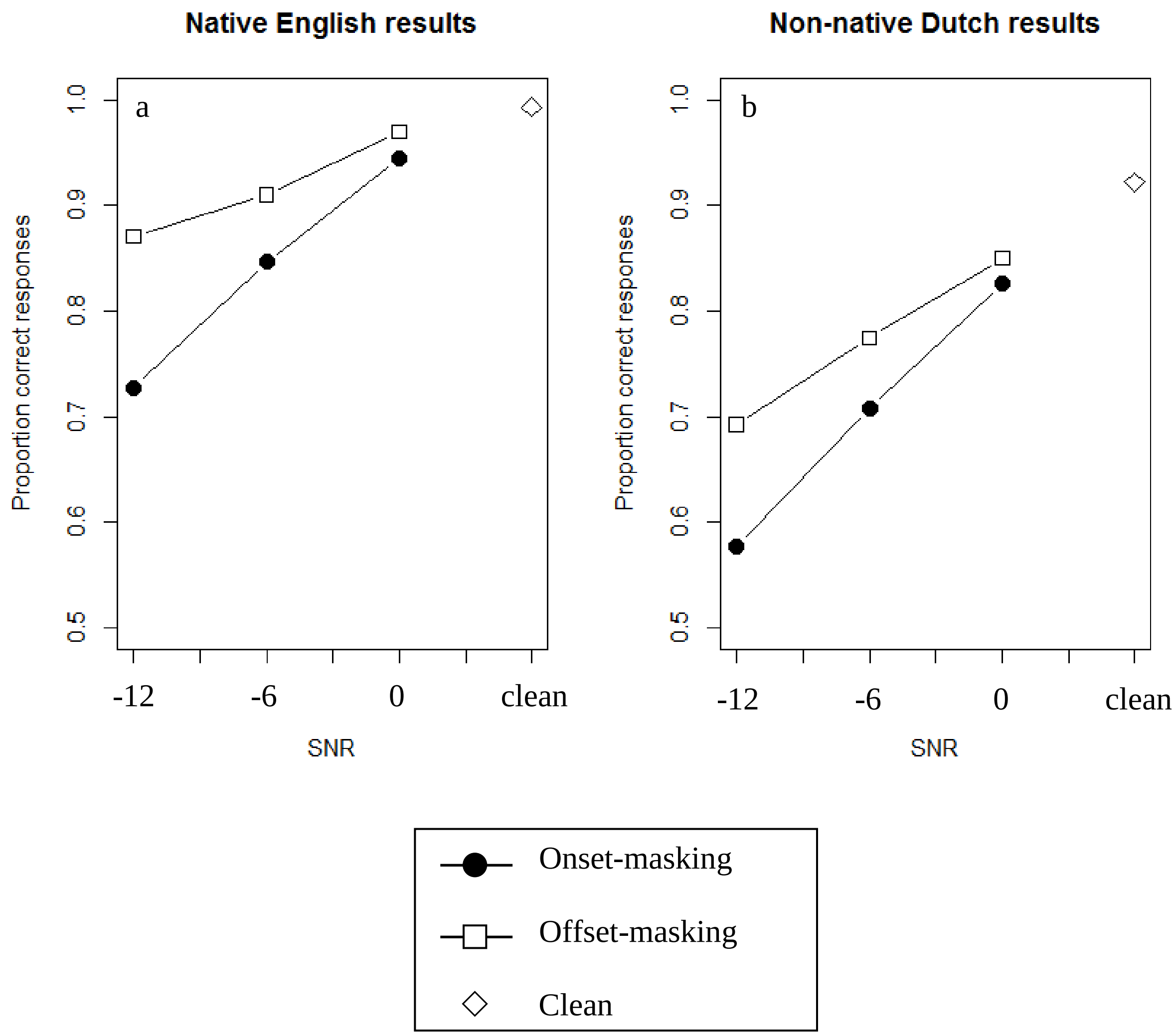

Figure 2. Proportion of correct responses for the SNR conditions and the clean condition, for the two masking conditions separately. Figure 2a shows the native English results; Figure 2b shows the Dutch non-native results.

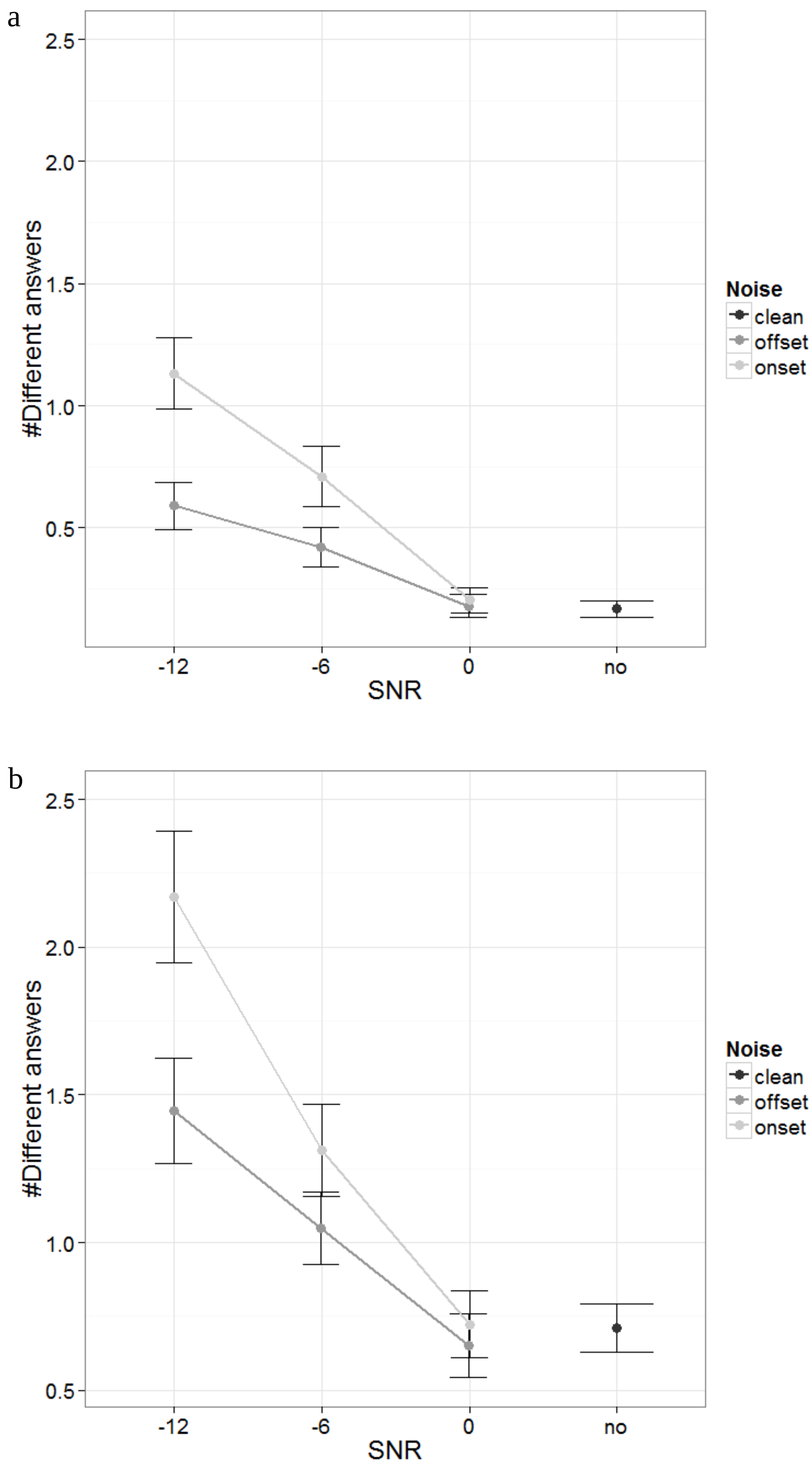

Figure 3. The number of different erroneous responses for the SNR conditions and the clean condition plotted for the onset- and offset-masking separately. Figure 3a: native English data. Figure 3b: non-native Dutch data. Error bars indicate one standard error above and one standard error below the mean.

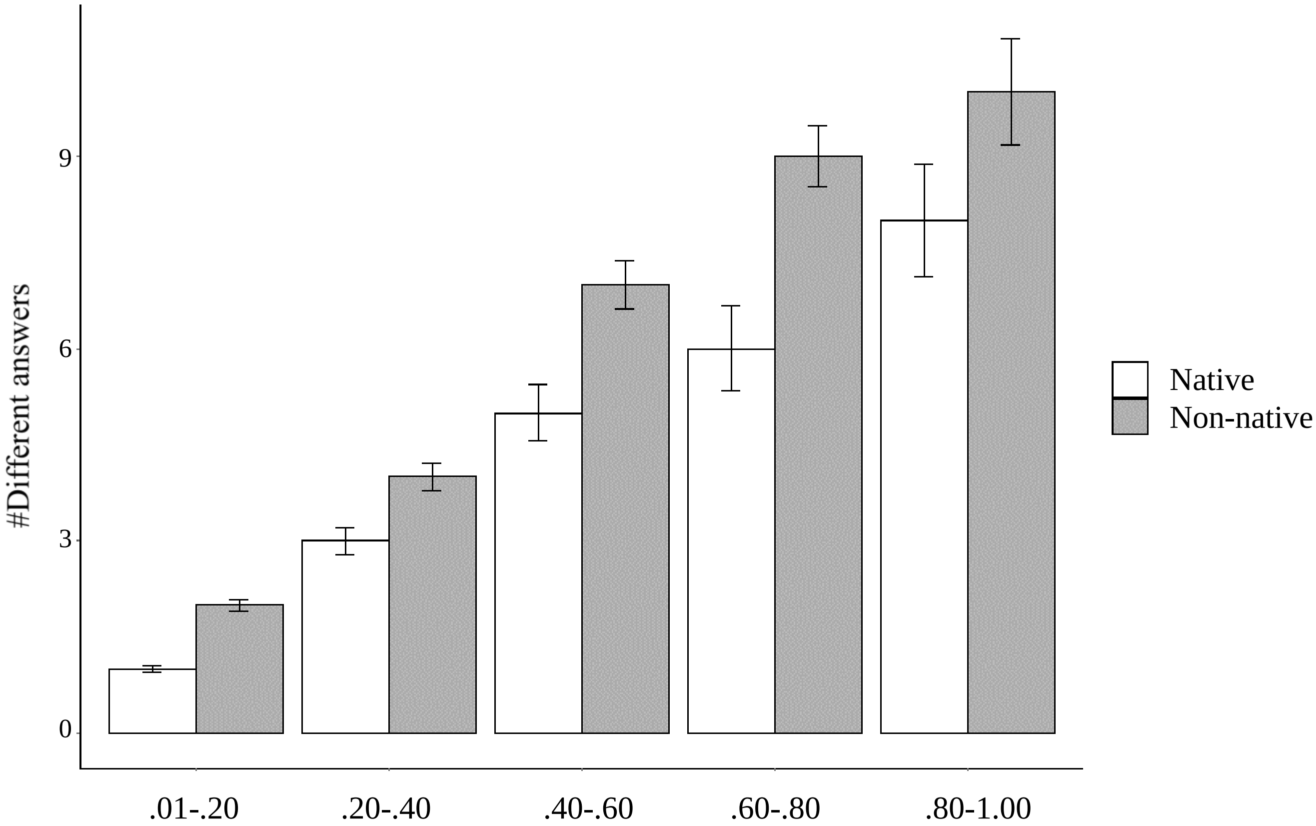

Figure 4. Bar plot of the relation between error rate and the number of different responses by the native listeners (grey bars) and the non-native listeners (yellow bars). Error bars indicate one standard error above and one standard error below the mean. The bins represent proportions error rate.

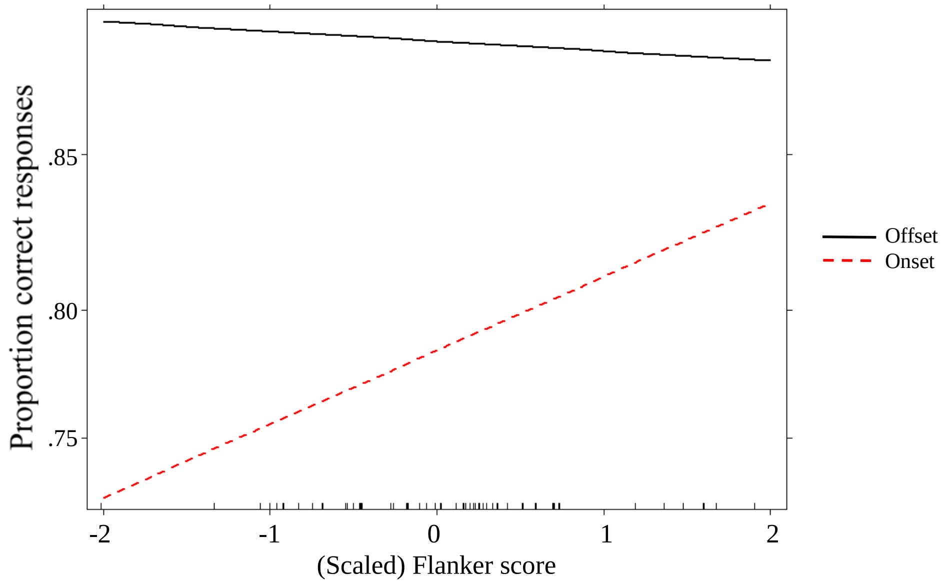

Figure 5. Interaction between (scaled) Flanker and Position of Noise for the non-native listeners. The solid line indicates the offset-masking condition; the red dashed line indicates the onset-masking condition.

Appendix A

The target, matched-onset-word, and matched-offset-word sets used in this study, with their word frequencies. The columns *#ph overlap* denote the number of phonemes overlapping between the competitor and head-word.

Head-word		Matched-onset-word			Matched-offset-word		
	Word freq		Word freq	#ph overlap		Word freq	#ph overlap
Letter	2166	lettuce	115	3	sweater	196	4
Weather	1185	weapon	431	2	feather	99	4
measure	872	melon	38	2	treasure	186	4
Dollar	263	dolphin	24	3	collar	327	4
pardon	148	party	6669	2	garden	1998	3
Pillow	247	pillar	171	3	willow	65	3
soccer	67	socket	87	3	locker	90	4
Bunny	19	bucket	237	2	money	7226	3
Pickle	46	picture	1905	3	nickel	37	3
Carrot	45	carriage	231	3	parrot	45	4
Poster	105	postman	42	4	toaster	11	5
Casket	35	castle	424	3	basket	320	5
Paddle	31	padlock	24	3	saddle	160	3
Beaker	21	beetle	84	2	speaker	308	4
Sandal	19	sandwich	191	3	candle	140	4
Life	13180	line	3920	2	knife	663	2
Lane	635	lake	718	2	chain	585	2

Nose	1307	note	1479	2	rose	1356	2
dog	1285	doll	312	2	frog	74	2
Shape	1204	shade	415	2	cape	278	2
gun	1138	gum	92	2	sun	2689	2
Bag	1122	bat	183	2	flag	356	2
Bridge	1034	brick	498	3	fridge	70	3
store	1015	storm	502	3	door	5891	2
Boat	1000	bone	478	2	coat	936	2
Plane	815	plate	656	3	crane	39	2
Cat	739	cap	489	2	rat	156	2
Wing	585	wind	1945	2	ring	1188	2
Shell	518	shed	399	2	bell	507	2
cave	491	cake	383	2	wave	811	2
Grain	466	grail	24	3	brain	1214	3
Rail	413	rain	1279	2	tail	571	2
Ghost	351	goat	209	2	coast	861	3
Seal	216	seat	1396	2	wheel	514	2
Throne	175	throat	770	3	stone	1542	2
Horn	166	horse	1518	2	corn	445	2
Robe	156	road	3791	2	globe	184	2
Clown	58	cloud	560	3	down	21923	2
Clip	57	cliff	264	3	ship	793	2
Peach	47	peak	412	2	beach	1060	2
Snail	46	snake	263	3	whale	104	2

Yarn	24	yard	637	2	barn	186	2
Moose	14	moon	951	2	goose	107	2

Appendix B

Table A. Random effects structure of the best-fitting models of performance for the native English (on the intercept) and non-native Dutch listeners on the stimuli presented in noise, n = 9211 observations.

Random factor	***SD***
Stimulus (intercept)	1.650
Stimulus × SNR 0 dB	1.376
Stimulus × SNR -6 dB	.750
Subject (intercept)	.400

Table B. Random effects structure for the best-fitting models of performance for the analysis of the number of different answers for the English and Dutch listeners, with error rate and the number of times the stimulus was presented as covariates, n = 996 observations.

Random factor	***SD***
Stimulus (intercept)	.034
Stimulus × Language	.077

Table C. Random effects structure for the best-fitting models of performance for the individual differences analysis of the non-native listeners, n = 4812 observations.

Random factor	***SD***
Stimulus (intercept)	1.639
Stimulus × SNR 0 dB	1.429
Stimulus × SNR -6 dB	.668

Stimulus × LexTale	.203
Subject (intercept)	.339