Why listening in background noise is harder in a non-native language than in a native language
A review

Scharenborg, Odette; van Os, Marjolein

DOI
10.1016/j.specom.2019.03.001

Publication date
2019

Document Version
Accepted author manuscript

Published in
Speech Communication

Citation (APA)
Scharenborg, O., & van Os, M. (2019). Why listening in background noise is harder in a non-native language than in a native language: A review. *Speech Communication*, *108*, 53-64. https://doi.org/10.1016/j.specom.2019.03.001

Why listening in background noise is harder in a non-native language than in a native language: A review

Odette Scharenborg[1,2] and Marjolein van Os[3]

[1] Multimedia Computing Group, Delft University of Technology, The Netherlands

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

[3] Language Science and Technology, Saarland University Saarbrücken, Germany

O.E.Scharenborg@tudelft.nl, vanos@coli.uni-saarland.de

Corresponding author:

Odette Scharenborg

Delft University of Technology

Faculty of Electrical Engineering, Mathematics, and Computer Science

Van Mourik Broekmanweg 6

2628 XE Delft

The Netherlands

Abstract

There is ample evidence that recognising words in a non-native language is more difficult than in a native language, even for those with a high proficiency in the non-native language involved, and particularly in the presence of background noise. Why is this the case? To answer this question, this paper provides a systematic review of the literature on non-native spoken-word recognition in the presence of background noise, and posits an updated theory on the effect of background noise on native and non-native spoken-word recognition. The picture that arises is that although spoken-word recognition in the presence of background noise is harder in a non-native language than in one's native language, this difference is not due to a differential effect of background noise on native and non-native listening. Rather, it can be explained by differences in language exposure, which influences the uptake and use of phonetic and contextual information in the speech signal for spoken-word recognition.

Keywords: non-native, spoken-word recognition, background noise, cognitive processes

1 Introduction

Successful speech recognition is a key factor for social integration and communication. At the same time, it is thought that people who speak more than one language outnumber monolingual speakers. Most people who learned another language some years after having first acquired their first language (referred to as non-native listeners), even those with a high proficiency in the non-native language involved, will have noticed that communication, and especially understanding the specific words that have been spoken, is more difficult in a non-native than in the native language, particularly in the presence of background noise (see for experimental evidence, e.g., Borghini & Hazan, 2018; Bradlow & Alexander, 2007; Mayo et al., 1997; Meador et al., 2000; Scharenborg et al., 2018a; note, also early or simultaneous bilinguals, i.e., people who learned two or more languages (nearly) simultaneously from an early age, have been found to suffer more from the presence of background noise than monolingual listeners; e.g., Mayo et al., 1997). The main reason for this problem seems obvious: Imperfect knowledge of the language and the presence of background noise (together referred to as adverse listening conditions) interact strongly to our disadvantage (e.g., Bradlow & Pisoni, 1999; Garcia Lecumberri, Cooke, & Cutler, 2010; Mayo, Florentine, & Buus, 1997).

Research on non-native listening in noise has, so far, mostly focussed on phoneme perception, showing that phoneme perception in noise is worse for non-native listeners than for native listeners (e.g., Broersma & Scharenborg, 2010; Cooke et al., 2010; see for a review, Garcia Lecumberri et al., 2006), an effect that is now fairly well understood (see Section 1.2). In recent years, an increasing number of studies have been published focussing on the effect of the presence of background noise on word recognition. The results of these studies again show a native advantage. The obvious question is, why is word recognition in the presence of background noise harder in a non-native language than in one's native language? Since different studies used different research methodologies, with different stimuli, tasks, noise levels, and types of noise, and tested different groups of participants with various language backgrounds and proficiency levels, and a systematic comparison of these studies is lacking, this question is not easily answered.

The aim of this paper is to answer the above question. In order to do so, this paper first provides a review of the literature on non-native spoken-word recognition in the presence of background noise, in which the studies (see Appendix A for an overview of the papers on non-native spoken-word recognition in background noise discussed in this review; we will refer to these papers as the 'sample'[1]) on non-native spoken-word recognition in background noise are for the first time systematically compared in order to understand the size of the native advantage (Section 2), the effect of different types of noise on spoken-word recognition (Section 3), the role of semantic and prosodic context (Section 4), and the role of individual differences in proficiency in the non-native language and cognitive abilities (Section 5). As such, this paper provides the first review of the literature on non-native spoken-word recognition in background noise. In the final section of this paper, the reviewed research is synthesised into an updated theoretical account of the effect of background noise on native and non-native listening (which thus abstracts away from the methodological differences between studies). In the remainder of

[1] The sample of papers on non-native spoken-word recognition reviewed in this paper aims to be exhaustive. The papers were found by browsing through the bibliographies of papers on non-native spoken-word recognition in background noise, and by typing relevant keywords into a search engine. Studies were included in the sample when they investigated word recognition and not for example phoneme recognition, and when they included a group of non-native listeners tested in some form of background noise. We adhere to the terminology used in the original papers, i.e., some papers talk about late-bilinguals, which according to our definition as outlined above would be non-native listeners.

this section we briefly summarise the processes underlying native and non-native spoken-word recognition and the effect of noise on these processes before turning to the question why word recognition in the presence of background noise is more difficult in a non-native language compared to a native language in Sections 2-5.

1.1 Native and non-native spoken-word recognition in quiet

The spoken-word recognition process can be viewed as the search for the optimal mapping of the acoustic speech signal onto a word. Several (computational) models of (native) word recognition have been proposed, such as TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994), and PARSYN (Luce, Goldinger, Auer, & Vitevitch, 2000; see for reviews: McQueen, 2005; Weber & Scharenborg, 2012). Most influential models agree on the following. As the auditory information unfolds over time, it is mapped onto stored representations of the words in the mental lexicon. This process is generally viewed as consisting of three underlying cognitive processes. First, all words that partly overlap with the input, irrespective of their onsets, are activated simultaneously (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Gow & Gordon, 1995; Luce & Pisoni, 1998; Slowiaczek, Nusbaum, & Pisoni, 1987; Zwitserlood, 1989). This is referred to as the multiple activation process. As each language only has a limited set of phonemes from which all the words in that language are built (Maddieson, 1984), words are often highly similar (e.g., *tall*, *ball*, *mall*, *call* only differ in their first consonant), and shorter words are often embedded in longer words (e.g., *sun*, *I*, *rye*, *rise*, *rises* in *sunrises*). So upon hearing the word *tall*, all other words which resemble it, e.g., *ball*, *mall*, *call*, *tell*, *toll*, etc., will also be activated and compete for recognition (for a review, see McQueen, 2005). The number and nature of the words (the 'neighbourhood') that are activated have been shown to affect the speed and accuracy of word recognition: Words that have a dense and/or high-frequency neighbourhood tend to be processed slower and less accurately, thus requiring more cognitive effort (Luce & Pisoni, 1998; but see Vitevitch & Rodríguez, 2005, for results that challenge this canonical view). During the competition process, active candidates that fail to match the acoustic input and/or the semantic context are inhibited, leaving the optimal word candidate given the acoustic input and semantic context to be recognised (Marslen-Wilson, 1993). In the final step the semantic information related to the selected words is integrated into the ongoing sentence, which is known as the integration process (Marslen-Wilson & Tyler, 1980), and the word is recognised.

Non-native spoken-word recognition happens in much the same way as native spoken-word recognition. However, languages differ in their phoneme inventories (e.g., Dutch does not have the /æ/ as in English *bad*, while English does not have the /y/ as in Dutch *vuur*, English translation: *fire*). Non-native listeners will have to learn the non-native sound categories which might consequently be less well specified or even absent. This leads to a decrease in the phonological match between the speech signal and the non-native listener's sound categories (compared to that during native listening), which has been shown to lead to a decrease in phoneme perception accuracy (see for an overview, Bohn & Munro, 2007). There is ample evidence showing that the misperception of speech sounds leads to an increase of activated words due to an increase of words that partially match the (mis)recognised speech sounds (Broersma, 2012; Cutler, Weber, & Otake, 2006; Pallier, Colomé, & Sebastián-Gallés, 2001), not only from the non-native language but also the native language (Spivey & Marian, 1999; Weber & Cutler, 2004). Using an eye-tracking study, Weber and Cutler, for instance, showed that Dutch non-native listeners of English upon hearing the word English word *panda* would not only look at a picture of a *panda* but also at a picture of a *pen*, while English listeners would only look at the picture of the *panda*. The Dutch listeners confused English /æ/ with Dutch /ɛ/ which led to the spurious activation of *pen*. These spurious competitors are difficult to suppress, resulting in more competition for non-native than native listening (Broersma & Cutler, 2008, 2011), decreasing word recognition accuracy (Scharenborg et al., 2018a).

1.2 Speech processing in background noise

By far most research on the effect of background noise during non-native spoken-word recognition uses additive noise, thus mostly leaving aside distortions of the acoustic signal due to reverberation or transmission channels (but see Rogers et al., 2006; and, e.g., Nabelek, 1988; Nabelek & Donahue, 1984 for the effect of reverberation on phoneme perception). An often used distinction to describe the type of masking by additive noise is energetic masking and informational masking (Shinn-Cunningham, 2008). In the case of energetic masking, both the target speech and competing noise contain energy in the same critical frequency bands at the same time (Brungart, 2001). Because of this, listeners cannot effectively identify and use the acoustic cues needed to identify sounds. Put differently, energetic masking occurs due to the direct interaction of the background noise with the speech signal outside the listener (Pollack, 1975). Informational masking is 'noise' that interferes with speech perception inside the listener (Lidestam et al., 2014; Pollack, 1975). Informational masking is a container concept for all types of interferences after the effect of energetic masking has been taken into account (e.g., Cooke et al., 2008; Garcia Lecumberri et al., 2010, Mattys et al., 2009). For example, imagine another person speaking in the background. The speech of that talker will mask the speech of the target talker you are attending to. This is energetic masking. At the same time, if the background talker speaks in a language you understand, the linguistic message in the background talker's speech will also interfere with recognition of the target talker's speech. Note, however, that an informational masker does not necessarily also provide energetic masking, for instance, carrying out a second task takes away cognitive resources from the speech recognition task and interferes with intelligibility of the speech signal (Mattys et al., 2009). This interference of the second task is also considered to be an informational masker. Since this review is concerned with speech processing in background noise, we will not focus on this type of informational masking.

Reverberation is not a background noise in the same vein as an energetic or informational masker, rather in the case of reverberation, the masking energy comes from the target speech itself. The sounds are reflected from surfaces and this results in a smeared signal (Garcia Lecumberri et al., 2010). Specifically, offsets of sounds are obscured, phoneme durations are prolonged, and bursts are smoothed. In the context of this review, reverberation is considered a background noise and consequently studies on reverberation are included.

In addition to the type of background noise, another important factor to consider is the level of the background noise. The level of the noise is measured in terms of Signal to Noise Ratio (SNR), which is a measure of the relative amplitude of the speech signal compared to the background noise, where a positive number means that the speech signal is stronger than the background noise, and a negative number the reverse. At SNR is 0 dB, both sound sources are equally loud. The severity of the masking effect, and thus the reduction in intelligibility of the speech signal, is dependent on the number and size of “glimpses” still available to the listener (Cooke, 2006). “Glimpses” are defined as those time-frequency regions where the energy of the speech exceeds the energy of the background noise by at least 3 dB.

The presence of background noise can obscure acoustic cues of the (target) speech, or acoustic cues from the background speech might ‘attach’ themselves to the target speech (Cooke, 2009). Both situations will lead to the listener perceiving incorrect acoustic cues and subsequently the listener is likely to hear a different sound than was intended by the talker (Cooke, 2009; Garcia Lecumberri et al., 2010). The presence of background noise thus decreases the phonological match between the speech signal and the listener’s sound categories, which results in a decrease in phoneme identification 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 deteriorated phonemic perception on the spoken-word recognition process is however less clear. This is the question we aim to answer and the topic of this review.

2 The native advantage

Studies generally report little or no differences between word recognition scores for native listeners and high-proficiency non-native listeners in quiet (see the papers in Appendix A; non-native listeners with a lower proficiency do perform worse than native listeners; e.g., Cooke, Garcia Lecumberri, & Barker, 2008). The difference in word recognition performance between native and (high-proficiency) non-native listeners occurs primarily in the presence of background noise (see the papers in Appendix A). A majority of the research on non-native spoken-word recognition in background noise focusses on quantising the size of the performance gap between native and non-native listeners, i.e., the native advantage. Different strategies and methodologies are used to investigate the size of the native advantage (Section 2.1). Section 2.2. discusses an important open question, i.e., whether this native advantage is constant or differs in varying levels of noise. Section 2.3 discusses the role of the selection of the native and non-native listener groups and the effect of the native language background of the non-native listener.

2.1 Measuring the native advantage

Different studies in our sample use different methods and stimuli to determine word recognition in noise performance. Stimuli are often taken or adapted from standardized tests, such as the Speech Perception in Noise (SPIN) test (Mayo, Florentine, & Buus, 1997; Tabri et al., 2010), the (Revised) Bamford-Kowal-Bench (BKB(-R)) Standard Sentence Test (Bradlow & Bent, 2002; Brouwer, Van Engen, Calandruccio, & Bradlow, 2012; Van Engen, 2010), the Hearing

In Noise Test (Jin & Liu, 2012; Kilman, Zekveld, Hällgren, & Rönnberg, 2014; Zhang, Xie, Li, Chatterjee, & Ding, 2014), a standardized Mandarin Chinese speech perception test (Zhang et al, 2014), the OLSA test (Warzybok, Brand, Wagener, & Kollmeier, 2015), the GÖSA test (Warzybok et al., 2015), the speech reception threshold test (SRT; Van Wijngaarden et al., 2002), and the VU98 test (Kaandorp et al., 2015). All these tests contain full sentences with keywords. Other used tests only contain brief carrier phrases or individual keywords (CID W22, Rogers et al., 2006; Shimizu, Makishima, Yoshida, & Yamagishi, 2001; Digit Triple Test, Warzybok et al., 2015). See Section 4.1 for an overview of the effect of semantic context on word recognition in noise.

The most-often used approach to determine word recognition in noise performance, used by 16 out of 23 studies in our sample, is to select a number of different SNRs, present the target speech signal at a fixed level of intensity at these different SNRs, and calculate the number or percentage of correctly identified words (see Appendix A for a listing). The target speech is typically presented at an intensity level between 60 and 70 dB SPL, while a wide range of SNRs is used: ranging from -13 dB to +20 dB in our sample. Word recognition accuracy deteriorates at more difficult SNR conditions, and all studies report better performance for native listeners than for non-native listeners, particularly in more difficult listening conditions.

A second approach to determine the size of the native advantage in spoken-word recognition in noise is to determine the SNR levels for which both listener groups would yield similar performance. Based on previous research and a small pilot study, Bradlow and Alexander (2007) found that a +4 dB SNR adjustment was enough to overcome the native advantage for a group of non-native listeners with various language backgrounds. Arguing that their participants had a higher proficiency level than the participants in the study by Bradlow and Alexander, Brouwer et al. (2012) believed that a +2 dB SNR adjustment for the non-native listeners would allow them to reach a similar level of baseline performance as the native listeners. However, results showed that this was not the case, but rather that also their relatively high-proficiency participants needed a +4 dB SNR adjustment to perform similarly to the group of native listeners. Cooke et al. (2008) used a +6 dB SNR adjustment, as previous research (comparing native word recognition in an informational masker condition with non-native word recognition in an energetic masking condition; Garcia Lecumberri & Cooke, 2007) suggested that a positive difference of 6-8 dB SNR would be enough to overcome the native advantage. While the authors do not give statistical evidence for this, inspection of their results suggest that at least for the energetic masking condition, non-native listeners indeed performed similarly to native listeners with a +6 dB SNR adjustment. Van Engen (2010) found that non-native listeners need an SNR adjustment of approximately +8 dB to reach similar levels of word recognition as native listeners.

A third approach of determining word recognition in noise performance is to use an adaptive procedure to determine for each participant individually the speech reception threshold, i.e., the SNR to obtain an accuracy score of 50% (Kaandorp et al., 2016; Kilman et al., 2014; Mayo et al., 1997; Takayanagi et al, 2002; Van Wijngaarden et al., 2002; Warzybok et al., 2015). In this procedure, first an individual starting level is determined or a fixed value is used, and consequently the SNR level is lowered in steps of 5, 3, or 2 dB (different for different studies in our sample) until the listener made a recognition error. Then the SNR is raised in steps of 2 dB until a correct response is given. This up-down adaptive procedure is continued until the participant reaches 50% accuracy scores. Results show that native listeners can reach 50% accuracy in conditions with louder noise than non-native listeners. The difference between native listeners and late L2 learners is approximately 6 dB SPL (although this seems to be dependent on the non-native language in question, see, e.g., van Wijngaarden et al., 2002), similar to the value mentioned by Cooke et al. (2008) and Garcia Lecumberri and

Cooke (2006). One advantage of the adaptive procedure compared to using fixed SNR levels is that it is easier to capture individual differences and more precise levels of SNRs to correctly recognize 50% of words. Nevertheless, all procedures yield the same result: Native listeners outperform non-native listeners when noise is present in the background, and the native advantage can be overcome with SNR adjustments of around +4 to +8 dB SPL.

2.2 Is the size of the native advantage constant across different noise levels?

Results in agreement with both the native advantage being constant and being different across different noise levels have been reported. For instance, Bradlow and Bent (2002) found that non-native listeners were not more adversely affected by increasing levels of noise than native listeners, though they speculate that this might be due to floor effects for the non-native group in the difficult listening condition. Scharenborg et al. (2018a) also found similar effects of background noise for native and non-native listeners. In a word identification task in different levels of speech-shaped-noise, they observed an overall downward shift of correct responses for the non-native group relative to the native group. These results suggest that the native advantage is constant across listening conditions, meaning that while non-native listeners perform worse than native listeners in adverse listening conditions, this difference is the same irrespective of the difficulty of the listening conditions. On the other hand, Jin and Liu (2012) showed that the native advantage increased as listening conditions became more challenging. Cooke et al. (2008) also found that non-native listeners suffered more from the effects of increasing levels of noise in a task where they identified keywords in simple sentences, though not in a more difficult task. Similarly, Rogers et al. (2006) found that bilingual listeners are more adversely affected than monolingual listeners in increasing noise conditions, though this finding is only present in one of the two tests they used.

2.3 Selection of the native and non-native listener groups

Listeners are typically tested in their non-native language and then compared to a control, native listener group (e.g., Ezzatian et al., 2010; Scharenborg et al., 2017; Shimizu et al., 2001; Zhang et al., 2014). In some cases though, the non-native listeners are also tested in their native language (Golestani et al., 2009; Van Wijngaarden et al., 2002). Importantly, the language background of the listeners in the non-native listener group differs between studies.

In the selection of the non-native listeners, feasibility plays a large role. In multilingual countries or countries where the general population has a high level of proficiency in a particular non-native language, the general (student) population can participate in studies investigating word recognition in that non-native language. This allows for groups of participants with a homogeneous background (Brouwer et al., 2012; Cooke et al., 2008; Coumans et al., 2014; Ezzatian et al., 2010; Golestani et al., 2009; Jin & Liu, 2012; Scharenborg et al., 2016a, 2018a; Shimizu et al, 2001; Van Engen, 2010; Van Wijngaarden et al., 2002; Zhang et al., 2014). Also studies that specifically investigated early bilinguals tended to have a group of participants with homogeneous language backgrounds (Ezzatian et al., 2010; Mayo et al., 1997; Meador et al., 2000, Rogers et al., 2006). Testing participants with a similar language background results in less variation caused by other factors than the manipulated factor, as the group is more homogeneous. However, results from a very specific population might not apply to populations with different language backgrounds. The availability of a large group of (exchange) students with various language backgrounds learning the same non-native language, allows one to investigate non-native word recognition without effects being caused by one common native language. These exchange students generally have different proficiency levels and exposure to the non-native language, ranging from several months to many years (Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Ezzatian et al., 2010; Kaandorp et al., 2015; Tabri et al, 2010; Takayanagi et al, 2002; Warzybok et al, 2015). Findings from groups of participants with varying language backgrounds might therefore be more robust and more generalisable than those from homogeneous groups of participants.

3 The effect of different types of noise on spoken-word recognition

The effect of background noise on spoken-word recognition is dependent on the type of noise that is present. Section 3.1 discusses research focussing on the effect of energetic and informational maskers on spoken-word recognition in background noise, while Section 3.2 discusses research focussing on the effect of reverberation.

3.1 Energetic and informational masking

Almost all studies in our sample investigate the effect of an energetic masker on native and non-native spoken-word recognition. Studying energetic masking is a good starting point for understanding the effect of noise on spoken-word recognition, as energetic masking is primarily a masker of the acoustic information, while informational masking is the masking that remains after the effect of energetic masking has been taken into account. Investigating pure energetic maskers and comparing these to informational maskers allows for the assessment of the “extra” effects, which then may be attributed to informational masking. Understanding the effect of energetic masking thus makes it easier to determine the effect of the informational aspect of the informational masker.

Most often speech-shaped noise (SSN) is used (see Appendix A for a listing). SSN has a spectrum that approximates the average long term spectrum of the speech of an adult male speaker. It simulates the noise found in real life situations, for example at a noisy party. Another frequently used energetic masker is multi-speaker babble. Multi-speaker babble noise has a similar masking effect as SSN (Golestani et al., 2009), though results obtained by Jin and Lui (2012) suggest a possible interaction of native language and type of energetic masker. The number of talkers used in multi-speaker babble ranges from two (Brouwer et al., 2012; Van Engen, 2010) to 12 (Mayo et al., 1997; Tabri et al., 2010). A few studies used white (Bradlow & Bent, 2002; Shimizu et al. 2001) or pink noise (Meador et al., 2000; Shimizu et al. 2001). Shimizu et al. additionally used aircraft noise, which was similar to pink noise, and which was set at an SNR level comparable to that during flights. They found significantly worse recognition performance in white noise conditions than in pink and aircraft noise conditions at the same SNR levels when listening conditions were difficult. This suggests that white noise has a larger negative effect on non-native word recognition than pink and aircraft noise.

Only five studies in the sample used informational masking in their experiment (Brouwer et al., 2012; Cooke et al., 2008; Ezzatian et al., 2010; Kilman et al., 2014; Van Engen, 2010). For instance, Cooke et al. (2008) used one-talker competing speech with similar target and masking sentences for the target speech and the speech in the background. This study aimed to investigate the effects of energetic and informational masking on native and non-native word recognition. Their stimuli consisted of short sentences from the Grid corpus (Cooke, Barker, Cunningham, & Shao, 2006), which have a fixed structure and contain colour, letter, and number keywords, thus reducing contextual information. Results showed that both in energetic masking and informational masking conditions, the native listeners outperformed the non-native listeners, with an increasing difference between the groups in the energetic masking condition when the noise levels increased. The study does not, however, directly compare the energetic and informational masking conditions, which is also made difficult by the different measures used for the word recognition performances: For the experiment on energetic masking the scores were converted to rationalized arcsine units (RAU), while the scores for the experiment on informational masking were presented in percentages. Nevertheless, the results suggest that informational masking is more disruptive to spoken-word recognition than energetic masking for both native listeners and non-native listeners.

Kilman et al. (2014) did directly compare energetic and informational masking. They used four different types of noise: two energetic maskers, i.e., stationary speech-shaped noise and fluctuating noise, and two informational maskers, i.e., two-talker babble noise in Swedish and English. Both types of babble noise contained speech from a male and female speaker. Word recognition results consistently showed that listeners performed better in their native language than the non-native language. Moreover, they also found the informational maskers to result in worse performance than energetic maskers, both for native and non-native language target speech. A possible reason for this finding is that informational maskers mask the speech signal similarly to energetic maskers, but have added interference from the intelligible speech in the background noise.

A factor influencing the strength of the masker is whether the language of the masking speech is known to the listener: a known language gives a larger masking effect than an unknown language. Brouwer et al. (2012) investigated the effect of different masking noise languages on word recognition in the listeners' native or second language. Monolingual English speakers and Dutch non-native speakers of English were presented with English sentences embedded either in two-talker babble noise with English competing speakers or with Dutch competing speakers. Results showed that word recognition performance was higher when the background language mismatched the target language, both for participants listening to their native language and the participants listening to a non-native language, though this effect was smaller for the Dutch group of listeners. In the Dutch background speech condition, the effect of the masker was thus less strong for the English group (who did not know Dutch) than for the Dutch non-native listeners of English (who could understand the Dutch in the background speech). These findings led the authors to conclude that when target and masker are more similar, they have a stronger negative effect on word recognition performance. Similar results have also been found by Van Engen (2010), who found that when listening to English target speech embedded in either an English masker or a Mandarin Chinese masker, native English listeners showed a greater release from the Mandarin Chinese masker than Mandarin Chinese native listeners. Additionally, in this experiment, the group of native Mandarin Chinese listeners performed better at the word recognition task when the target speech was embedded in English competing speech than Mandarin Chinese competing speech. Together, these results suggest that both the similarity between the target speech and competing speech, and whether this competing speech is in the listener's native language, play a role in word recognition performance.

Taken together, these studies show that native listeners outperform non-native listeners on word recognition tasks, both in tasks where energetic maskers and where informational maskers are used. Both groups of listeners seem to be more adversely affected by informational masking than energetic masking. This finding can be explained by the fact that while energetic masking obscures parts of the target speech signal, informational masking has the same effect, i.e., the informational masker obscures the acoustics, but additionally has the linguistic message in the background noise interfering with speech recognition of the target speaker. The number of studies investigating the effect of informational masking and the number of studies doing controlled comparisons between the effect of energetic and informational masking is however still rather low. More research is needed to investigate the specific effects of different noise types.

3.2 Reverberation

Only one study in the sample investigated the effect of reverberation on word recognition in noise performance (Rogers et al., 2006). In this study, English monolinguals and early Spanish-English bilinguals (age of acquisition (AOA) < 6 years) participated in a word recognition task, where monosyllabic words were presented in quiet; in noise only; or in noise plus reverberation.

Speech-shaped noise was used in both noise conditions, and the reverberation level was similar to that naturally occurring in public meeting rooms. The results showed that both monolingual listeners and bilingual listeners obtained identical, perfect scores in the quiet control condition, and that both groups performed worse in the condition with noise plus reverberation than in the condition with noise only. Importantly, bilingual listeners performed significantly poorer than monolinguals in both the condition with noise and the condition with noise plus reverberation. These differences between the groups were similar between the conditions, suggesting that the presence of reverberation did not affect early bilinguals more than it did monolinguals.

4 The role of context in spoken-word recognition in background noise

Two types of contextual information have been found to play an important role in explaining the native advantage in spoken-word recognition in background noise. Section 4.1 discusses the role of semantic context, while Section 4.2 discusses the role of prosodic information.

4.1 Semantic context

An important factor that has been shown to affect word recognition is semantic context. Target words in predictable sentences are easier to recognise than the same words in a less predictable sentence. For example, the target word television is recognised more easily in the high-predictability sentence *a* than in the low-predictability sentence *b*:

a. *the man is watching the show on his television*

b. *the man just bought a new television*

The facilitatory effect of semantic context on the activation of words in native listening is however highly sensitive to the intelligibility of the speech, i.e., the semantic context produces less activation of semantically related items in degraded listening conditions (Aydelott & Bates, 2004). Nevertheless, semantic information facilitates word recognition in degraded listening conditions for native listeners (Aydelott & Bates, 2004; Aydelott et al., 2012).

The results for non-native listeners are somewhat mixed, with some studies indicating that non-native listeners are able to use semantic contextual information when listening in noise, while others found no such facilitating effect for the non-native listeners. Mayo et al. (1997) showed that late bilinguals were less able to take advantage of semantic contextual information than monolinguals or early bilinguals. Mayo and colleagues used both low-predictability sentences and high-predictability sentences, which allowed them to directly assess the effect of context on word recognition scores in different groups of participants. The late bilinguals showed no difference in word recognition scores in low- and high-predictability sentences, indicating that semantic contextual information did not help recognition in these participants. Both monolinguals and early bilinguals did show a difference in word recognition in low- and high predictability sentences. Bradlow and Alexander (2007) investigated whether the non-native disadvantage could be overcome when acoustic (clear speech) and semantic (high-predictability sentences) information were available. Their results showed that non-native listeners are able to use semantic information in the sentence to improve word recognition in clear listening conditions, but cannot use this information in noise. Warzybok et al. (2015) compared three different German hearing tests, which are generally used in clinical audiology. These tests differed in the amount of semantic contextual information available in the sentence. When three digits were presented in noise in a carrier phrase with no semantic contextual information, speech recognition scores were equally good for the native and the non-native listeners with intermediate L2 proficiency or higher. Comparing a hearing test with low-predictability sentences and a hearing test with high-predictability sentences, results showed a larger difference in scores between the native and non-native listeners in the high-predictability test, with non-natives performing worse regardless of their proficiency levels. These lower recognition scores show that intermediate to high proficiency non-native listeners are less able to use semantic contextual information than native listeners.

Golestani et al. (2009) isolated the semantic level of speech using a retroactive priming paradigm to investigate whether this high-level component contributes to the native language context benefit. In this paradigm, participants listened to two words, the first of which (the prime) was embedded in noise, while the second (the target) was clearly audible. These two words were either semantically related or unrelated. After target word offset, two words were presented visually on a screen, one identical to the prime and one distractor that was semantically related to the prime. Participants had to select the word they had heard as the prime. This was done both in the participant's native language and the non-native language. Results showed more accurate scores for semantically related prime-target pairs when listening in the native language than when listening in the non-native language, which led the authors to conclude that semantic information contributes to the non-native disadvantage for word recognition in noise.

The effect of semantic context on spoken-word recognition can be attributed to two types of semantic information, which each have a distinct effect on spoken-word recognition. Part of the facilitating effect of sentence context can be attributed to semantic (or associative) priming (e.g., Meyer & Schvaneveldt, 1971), i.e., earlier recognised words (so-called primes) in the sentence activate subsequent, related words (often referred to as target words) leading to faster recognition of those later words. Primes thus have a direct effect on the multiple activation process, and on speed of recognition. This priming effect is smaller in non-native compared to native listening in noise (Aydelott et al., 2012; Golestani et al., 2009). Second, there is an effect of predictability (or semantic congruency) above and beyond priming which has been related to the fit of the activated word in the sentence context (referred to as semantic integration) (FitzPatrick & Indefrey, 2010; Zwitserlood, 1989). In non-native listening, semantic integration has been found to modulate the multiple activation process, constraining the number of activated word candidates from the native language (Li & Yip, 1998) to the point where they receive little to no activation (FitzPatrick & Indefrey, 2010). Moreover, ERP studies have shown that semantic integration is slower in non-native than in native listening (e.g., FitzPatrick & Indefrey, 2010), thus suggesting an effect of semantic context on the integration process. The reduced exploitation of semantic information by non-native listeners in noisy listening conditions is argued to reflect a need for more contextual information in order to build up semantic structures (Oliver et al., 2012).

The role of context on word recognition, although highly important, is still not fully understood. The above reviewed literature indicates that semantic context can have an effect on all three cognitive processes underlying the spoken-word recognition process. Open questions concern the effect of background noise on the accessibility of semantic features of words and the integration of this information for word recognition during both native and non-native listening.

In everyday conversations, talkers typically produce meaningful sentences, and the listener's task is to recognise these words from the meaningful context. In contrast to recognising words in isolation, a task often used in research on non-native spoken-word recognition in noise, these meaningful sentences provide contextual information which facilitates word recognition in quiet and noisy backgrounds. This suggests that semantic information from the meaningful sentences can be used to compensate for loss of information at lower (acoustic and sound) processing levels due to noise masking. If one is interested in the effect of background noise on pure word recognition, presenting words in meaningful, high-predictable sentences is thus not advisable because listeners can use the contextual information to help recognise the target words, and since non-native listeners are less able to use this information they are at a disadvantage compared to the native listeners.

Thus, when interpreting results of experiments which investigate word recognition using meaningful sentences the difference in natives' and non-natives' ability to use semantic, contextual information for word recognition needs to be taken into account. Alternatively, the influence of contextual information needs to be minimised. This can be done by using isolated words (Coumans, van Hout, & Scharenborg, 2014; Golestani et al., 2009; Kaandorp, De Groot, Festen, Smits, & Overts, 2015; Scharenborg, Coumans, Kakouros, & Van Hout, 2016a; Scharenborg et al., 2018a; Takayanagi, Dirks, & Moshfegh, 2002), by presenting words in a fixed carrier phrase (Rogers et al., 2006; Shimizu et al., 2001; Warzybok et al., 2015), by using sentences that are grammatically correct but semantically meaningless (Ezzatian et al., 2010) or using a keyword spotting task in which sentences contain only a limited number of common words (Bradlow & Alexander, 2007; Cooke et al., 2008; Meador, Flege, & MacKay, 2000). For example, Golestani et al. (2009) used isolated words in a priming paradigm, as this allowed them to separate the semantic level of speech from the syntactic and pragmatic levels present in sentences. Scharenborg et al. (2016a, 2018a) specifically investigated whether word-initial or word-final information was more important in word recognition, and thus used isolated words where either the word's onset or offset was masked by noise.

4.2 Prosodic context

Sentence accent (or sentential stress; Kahnemuyipour, 2009) plays an important role in speech comprehension (Akker & Cutler, 2003; Cutler, Dahan, & van Donselaar, 1997). For instance, compare the following two sentences, which consist of the same words but have different sentence accent (denoted by upper case), and consequently have a different meaning:

a. *The GRANDFATHER was playing chess*

b. *The grandfather was playing CHESS*

Where in sentence *a* it is emphasised that it was the *grandfather*, rather than, e.g., the grandmother, who was playing chess, in one reading of sentence *b* it is emphasised that the game of chess was played, and not some other game. Sentence accent thus expresses semantic focus. Focussed syllables are processed faster (Shields et al., 1974); moreover, multiple meanings of homophones are activated when a word is in focussed position, while none are activated when the word is in unfocussed position (Blutner & Sommer, 1988). Focussed syllables can thus considered to be processed in more detail, or more deeply, resulting in the activation of all meanings of the homophone. Rapid and effective processing of accent placement in an utterance is thus highly important in efficient comprehension of meaning (see for a review, Cutler et al., 1997) as it is pivotal in understanding the important parts of a speaker's message.

In optimal listening conditions, native listeners are able to exploit prosodic cues in the speech signal signalling upcoming sentence accent to actively focus their attention to those parts of the sentence where accent will fall (Akker & Cutler, 1997). Different languages however have different instantiations of sentence accent due to differences in syntactic structure of the language (Kahnemuyipour, 2009), which causes difficulty for listeners to exploit prosodic cues in a non-native language. Nevertheless, non-native listeners, at least those with a high proficiency in the non-native language, have been shown to be able to detect sentence prominence (Rosenberg, Hirschberg, & Manis, 2010; Scharenborg, Kolkman, Kakouros, & Post, 2016b; Scharenborg, Kakouros, Meunier, & Post, 2018b; Wagner, 2005) and to use acoustic, prosodic cues for prominence detection that are similar to those used by native listeners (Akker & Cutler, 1997; Wagner, 2005). Nevertheless, non-native listeners are slower to detect sentence accent (Scharenborg et al., 2016b), and display a reduced efficiency in using prosodic information signalling sentence accent to build semantic frameworks, i.e., they are less able to integrate these information sources for spoken-word recognition (Akker & Cutler, 1997). The reduced efficiency in building semantic frameworks could be due to a slower semantic integration in non-native listening compared to native listening (Akker & Cutler, 1997). This hypothesis is consistent with ERP findings of slower semantic integration in non-native than in native listening (FitzPatrick & Indefrey, 2010). Moreover, non-native listeners have been found to have more difficulty exploiting prosodic cues in the presence of background noise than native listeners (Scharenborg et al., 2016b, 2018b). This impaired uptake and exploitation of prosodic cues might reduce the efficiency of the spoken-word recognition process, resulting in, e.g., less fast and deep processing of focussed syllables, which might have consequences for the activation of related words. The effect of impaired prosodic cue uptake on non-native spoken-word recognition is largely unknown, but is likely to be a factor in explaining the native advantage during spoken-word recognition in noise.

5 The role of proficiency and cognitive abilities on spoken-word recognition in noise

A common observation is that some people have more difficulty in listening in the presence of background noise than others. However, despite these individual differences, psycholinguistic theories of spoken-word recognition are often based on behavioural results that are averages over multiple subjects, thus removing between subject variation. There are however a couple of studies which investigate the role of individual differences on non-native speech recognition in background noise. Section 5.1 discusses the role of proficiency in the non-native language, while Section 5.2 discusses the role of cognitive abilities on spoken-word recognition in the presence of background noise.

5.1 The role of proficiency

A number of studies used objective measures to assess participants' proficiency in the non-native language. These objective measures include standardized tests such as TOEFL (Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Jin & Liu, 2012; Van Engen, 2010), Cambridge Advanced Examination (Cooke et al., 2008), the Common European Framework of Reference (CEFR, Verhelst et al, 2009; Kaandorp et al, 2016; Warzybok et al., 2015), the Mill Hill vocabulary score (Ezzatian et al., 2010), and the Lexical Test for Advanced Learners of English (LexTALE, Lemhöfer & Broersma, 2012; Coumans et al., 2014; Scharenborg et al., 2016a,b, 2018a). However, other studies used subjective measures such as self-ratings of proficiency (Takayanagi et al., 2012; Van Wijngaarden et al., 2002), reported age of onset and length of residence (Meador et al., 2000), or length of formal education in the non-native language (Brouwer et al., 2012; Golestani et al., 2009; Shimizu et al., 2001). Only one study in the sample did not give any information on the non-native proficiency levels of the participants (Zhang et al., 2014).

Many studies tested participants with similar proficiency levels in the non-native language, but some studies in particular investigated the effect of language proficiency on word recognition in noise scores. Warzybok et al. (2015) tested four groups of participants with various native language backgrounds. These groups were of equal sizes (N = 10), and consisted of learners with basic, intermediate, high intermediate, or advanced skills in the non-native language. Results showed that on highly familiar and limited materials, non-natives with at least intermediate (B1-B2 on the CEFR) non-native language proficiency perform equally well as native listeners. When the materials are more complex and semantic context is available, all non-native groups perform worse than the native listeners. Participants with basic skills (A1-A2 on the CEFR) in the non-native language performed significantly worse than the advanced learners, and showed large within-group variance. The authors argued that this variance might be caused by the different strategies used by participants, who were unfamiliar with the words. Ezzatian et al. (2010) also compared four different groups of participants: one group of native English listeners, two groups of late bilinguals (one group consisting of listeners who moved to Canada between the ages of 7-14 years, and one group of listeners who moved to Canada after the age of 15 years), and one group of listeners with mixed native languages. They also found that the later in life English was acquired, the more difficulty the listeners had with spoken-word recognition in the presence of noise. In their experiment, they investigated the effect of spatial separation of the target speech and the noise maskers and found that all listener groups benefitted to the same extent from the separation of the target speech and the masker.

Kilman et al. (2014) aimed to investigate how L2 proficiency affects native and non-native word recognition in different types of background noise. Native speakers of Swedish participated in a Swedish and English word recognition task and were also presented with a standardised test used in Swedish schools to measure their English proficiency. To investigate the effect of English proficiency on word recognition in the English task, participants were divided into two groups, one of low proficiency and one of high proficiency. Results showed that the proficiency in English of the Swedish participants was the most important factor in explaining differences in non-native speech perception in noise: Participants with higher English proficiency levels outperformed participants with lower English proficiency levels in spoken-word recognition in noise. Zhang et al. (2014) investigated how noise and language proficiency affect speech recognition in native Mandarin Chinese listeners of English, and likewise found an effect of proficiency on word recognition in noise.

Scharenborg et al. (2018a) investigated the role of proficiency in the non-native language on the level of the individual listener. Rather than splitting the non-native listener group into several proficiency levels, they used Dutch listeners' individual LexTale scores as a measure of their (lexical) proficiency in English and added these to the statistical analyses. The

LexTale task is a free, fast, and easy to use, visual unspeeded lexical decision task for advanced learners of English (Lemhöfer & Broersma, 2012), in which participants are presented with 60 items (words and non-words) shown on a screen one-by-one. Participants have to indicate by button press whether the item on the screen is an existing word in English. The LexTale score is calculated as the percentage correct responses (to all stimuli). The score relates to the proficiency levels as defined in the CEFR for language proficiency levels.

An important advantage of taking proficiency into account at an individual level is that there is no need to make a priori groups, whose cut-off points can be arbitrary. Scharenborg et al. (2018a) tested native English and native Dutch participants in an English word recognition task where words were presented in clean and partially embedded in noise. Similar to Kilman et al. (2014), the results showed that, in general, native speakers performed better on the word recognition task than non-native speakers. With regard to individual proficiency levels, it was found that higher proficiency in the non-native language led to better word recognition performance; however, proficiency was not found to relate to word recognition in noise performance. Ezzatian et al. (2010), however, did observe a correlation between individuals' vocabulary scores and their speech reception thresholds, although they found no such correlation between listeners' reading ability and their speech reception thresholds. These mixed results seem to suggest that the specific proficiency measure used is important for finding a link between proficiency and spoken-word recognition in background noise, which in turns raises the question how to most effectively measure proficiency in a non-native language. This question is as yet unanswered, and goes beyond the scope of this review.

Some studies specifically investigated the effect of age of L2 acquisition (AOA) on word recognition in noise. AOA is a major factor that determines second language proficiency (Flege, Yeni-Komshian, & Lui, 1999). Mayo et al. (1997) tested three groups of participants: native monolingual, early-bilinguals (simultaneous bilinguals since birth or bilingual-sincetoddler, AOA < 6), and late-bilinguals (AOA > 14). Results showed that the first two groups score better in conditions with more noise than the late-bilingual group, and that the early-bilinguals group perform worse than monolinguals. According to Mayo et al., these findings of bilinguals (both early and late) performing worse on word recognition in noise are likely due to interference from the (other) first language. Similar results were found by Ezzatian et al. (2010), Meador et al. (2000), and Shi (2010). Additionally, Meador et al. found an effect of the use of the native language: A group of early bilinguals who used their native language relatively little outperformed a group of early bilinguals who used their native language more (8% vs. 32% daily native language use) in all but the easiest SNR level. Rogers et al. (2004) compared word recognition in quiet and noise by monolingual listeners and early bilinguals (AOA < 6), and found similar recognition scores for the two groups in quiet, but poorer scores for the group of bilinguals in noise. Investigating monolinguals, early bilinguals, and early trilinguals (AOA < 6), Tabri et al. (2010) found no significant differences between the groups in low levels of noise (10 or 15 dB SNR), but, again, lower recognition scores for bilinguals and trilinguals compared to monolinguals in higher levels of noise (5 or 0 dB SNR). These studies concluded that even early bilinguals who perform in a native-like way in quiet, are more adversely affected by noise than their monolingual peers, possibly because of interference of their other language(s). However, groups of simultaneous early bilinguals in these studies were often very small (N = 3 in both Mayo et al. and Rogers et al.), so more research is needed to investigate whether these results generalise to larger and other listener groups.

In summary, early bilinguals and high-proficiency non-native listeners perform native-like in quiet and relatively easy listening conditions with low levels of noise. However, in more difficult listening conditions, non-natives perform worse than native listeners, regardless of their proficiency levels and regardless of the method used to determine their proficiency (be it objective or subjective measures). Even early bilinguals are adversely affected by background noise during word recognition compared to monolingual native listeners. These results show the importance of the age of acquisition of the non-native language and proficiency in the non-native language on listeners' performance in word recognition in noise tasks. To account for inequivalent proficiency levels and to improve comparability of the results across studies, future investigations into the effect of background noise on non-native listening should include objectively measured individuals' non-native proficiency levels (see also Kilman et al., 2014; Scharenborg et al., 2018a).

5.2 The role of cognitive abilities

Kilman et al. (2014) and Scharenborg et al. (2018a) not only aimed to investigate how proficiency affects native and non-native word recognition in background noise, but they were also interested in the role individuals' cognitive abilities, i.e., working memory (Kilman et al.) and inhibition abilities (Scharenborg et al.), as factors that might explain individual differences in spoken-word recognition in background noise. In a speech perception in noise experiment with Swedish non-native listeners of English, Kilman and colleagues (2014) did not find a clear link between working memory abilities and word recognition performance. As mentioned above, they did find that listeners with a high proficiency in English were better at word recognition in noise, and argued that, similar to simultaneous bilinguals who have been found to have an improved executive, and particularly, inhibition control (Bialystok, 2001; Bialystok, Craik, Klein, & Viswanathan, 2004), high proficiency in the non-native language improves listeners' ability to inhibit both native and non-native babble maskers. Kilman et al. thus seem to argue that high-proficiency non-native speakers might obtain the same improved inhibition ability as simultaneous bilinguals, although they did not explicitly test the role of inhibition ability on word recognition in noise.

Scharenborg et al. (2018a) investigated the role of inhibition abilities, particularly selective attention, on English isolated word recognition in noise by Dutch non-native listeners of English. They found no effect of inhibition abilities on native spoken-word recognition in noise. However, non-native listeners with poorer selective attention were found to have higher word recognition accuracies than listeners with better selective attention abilities when the onset of a word was masked by noise. They argue that since inhibition builds up over time (Ridderinkhof, 2002; Van den Wildenberg et al., 2010), listeners who are poorer inhibitors might need more time to remove competitors from the active set of candidate words compared to better inhibitors. Consequently, poorer inhibitors have more candidate words activated for a longer period of time compared to better inhibitors. The larger competitor set for poorer inhibitors might actually be beneficial when listening in noise. If the onset of a word is masked by noise, better inhibitors might already have filtered the target word from the competitor space, while poorer inhibitors keep their options open for a longer period of time, and might still have the target word in their competitor space, increasing their chance of recognising the target word. This reasoning would also explain the lack of an effect of inhibition abilities for the native listeners, who would have a smaller competitor space, and thus are more likely to have dropped the target word from the competitor space. These findings tie in with those by Banks and colleagues (Banks, Gowen, Munro, & Adank, 2015), who investigated the role of inhibition on the adaptation to accented speech in the presence of background noise by native listeners. They found that listeners who were better inhibitors adapted faster and more to the accented speech than listeners who were worse inhibitors, which was shown by a larger difference in the mean speech reception threshold between the first three and the last three testing blocks for the better inhibitors compared to the worse inhibitors and a faster rate of adaptation. Following the reasoning by Scharenborg and colleagues, better inhibitors would have a smaller competitor space than worse inhibitors and would therefore be faster in recognising the word, thus speeding up and increasing the adaptation to the accented speech compared to worse inhibitors due to the faster availability of the necessary lexical information for adaptation to take place.

6 The effect of background noise on the cognitive processes underlying native and non-native listening

Most papers reviewed above aim to determine the size of the performance gap between native and non-native spoken-word recognition in background noise, and thus show *that* word-recognition in background noise is harder in a non-native language compared to one's native language. Here, we summarise and synthesise the reviewed research and bring together those and new findings from our lab into an updated theoretical account of the effect of background noise on native and non-native spoken-word recognition in order to explain *why* word recognition in the presence of background noise is harder in a non-native language than in one's native language.

There is accumulating evidence that noise has a detrimental effect on all speech processing levels during non-native listening (see, e.g., Bradlow & Alexander, 2007). The presence of background noise has been shown to have an effect on all cognitive processes underlying spoken-word recognition.

6.1 The effect of background noise on the multiple activation process

As reviewed in Garcia Lecumberri et al. (2010; see also Section 1.2), the presence of background noise impairs the phonological match between the acoustic signal and the sound categories. This decreased phonological match is shown to have repercussions on the subsequent multiple activation process: On the basis of their native and non-native word identification task in which words were partially masked by noise, Scharenborg and colleagues (2018a) showed that the presence of background noise not only results in more errors but also in an increase in the number of *different* answers provided by the native and non-native listeners, and that the number of different responses increased with worse SNR levels. At the group level, the non-native listeners gave more different answers than the native listeners, thus had a larger competitor space than the native listeners. Interestingly, after controlling for several factors, their analysis showed that the presence of background noise increases the competitor space in both native and non-native listening to the same extent. The number of activated candidate words has an effect on recognition speed and accuracy, with an increase in competitor space size leading to slower recognition speed and a lower accuracy (Norris et al., 1995). The difference between native and non-native spoken-word recognition performance in background noise as reviewed in this paper could thus be explained by the larger competitor space for the non-native listeners compared to the native listeners, which leads to more word recognition errors for the larger competitor space.

Scharenborg et al. (2018a) provided two, not mutually exclusive, possible explanations for the larger competitor space during non-native listening compared to native listening. In native listening, semantic and prosodic information optimise the search for the optimal word by restricting the search space, and they guide the search for the semantically most important part of the message of the speaker (Akker & Cutler, 2003). As reviewed above, native listeners are able to use this higher-level semantic (e.g., Bradlow & Alexander, 2007) and prosodic (Scharenborg et al., 2016b, 2018b) and lower-level acoustic information (e.g., Garcia Lecumberri et al., 2010) to a larger extent than non-native listeners. The native perceptual system is thus potentially more able and more efficient at removing candidate words that no longer fit the context leading to a smaller competitor space compared to non-native listeners.

Second, due to far less exposure to the non-native language compared to native listeners, non-native listeners' (abstract) sound categories are less well-defined than those of native listeners in the non-native language (e.g., Garcia Lecumberri et al., 2010). These less well-defined sound categories lead to inaccurate sound perception due to differences in the sound inventories between languages which in turn leads to an increase of spuriously activated candidate words during non-native listening compared to native listening (e.g., Cutler et al., 2006; Spivey & Marian, 1999; Weber & Cutler, 2004). This larger competitor space is then 'carried over' to noisy listening conditions. The differences in the size of the competitor space during native and non-native listening would then be explained by the difference in native and non-native listening.

Both these explanations hypothesize a larger competitor space during non-native listening compared to native listening due to differences in exposure to the non-native language. This 'exposure-to-language' hypothesis ties in with findings for early-bilingual listeners (i.e., listeners who learn two languages nearly simultaneously from birth): Early bilinguals with near-native proficiency perform equally well as monolingual listeners in quiet; nevertheless, they suffer more than monolingual listeners from the presence of background noise (e.g., Ezzatian et al., 2010; Krizman, Bradlow, Lam, & Kraus, 2016; Mayo et al., 1997; Meador et al., 2000; Shi, 2010). As a result of their more limited exposure to the non-native language, non-native listeners require a greater acoustic clarity in the speech signal for successful word recognition (Bradlow & Alexander, 2007) and are less able to extract linguistic meaning from the speech input (limitations in 'language-dependent processing', in line with Krizman et al., 2016). Moreover, recent evidence from a computational modelling study (Karaminis & Scharenborg, 2018) corroborates the role of exposure to language in explaining the differences between native and non-native spoken-word recognition and the effect of background noise on those. Karaminis and Scharenborg (2018) trained a newly developed computational model, ListenIN (Listening In Noise), in a native English and a Dutch non-native listener of English 'mode'. Apart from the ratio of the English vs. Dutch training material to train the model, the architecture of both the native and non-native models was identical. The models captured the key performance effects of three experiments on native and non-native spoken-word recognition in background noise, i.e., a higher error rate and more activated candidate words during non-native listening compared to native listening (from Scharenborg et al., 2018a), with a striking similarity between the models' and human results, and an attenuation and a slower resolving of the competition phase in the presence of background noise in native and non-native listening (Hintz & Scharenborg, 2016; Karaminis et al., in preparation). These simulations show how a different linguistic environment, more specifically differences in amounts of training material, might explain the imperfect language knowledge (Garcia Lecumberri et al., 2010) or the limitations in language-dependent processing (e.g., difficulties in extracting linguistic information; Krizman et al., 2016), which ultimately results in worse spoken-word recognition performance in non-native compared to native listening.

Moreover, differences in exposure to the language would also explain the differences between native and non-native listeners in the ability to use semantic contextual information (e.g., Aydelott & Bates, 2004; Aydelott et al., 2012; Bradow & Alexander, 2007; Golestani et al., 2009; Mayo et al., 1997), and would be in line with findings by Ezzatian et al. (2010) and Cooke et al. (2008), who found that a listener's ability (to pick up and use specific acoustic cues) to separate background noise from target speech is a more important factor in explaining why some people suffer less from the presence of background noise than other listeners than the listener's proficiency in the non-native language. We thus argue that native and non-native listening can and should thus be viewed as lying on a continuum of language experience, with a corresponding effect of noise on spoken-word recognition.

6.2 The effect of background noise on the competition and recognition process

There is accumulating evidence that background noise also has a direct effect on the competition process. Research has shown that listeners are able to flexibly adjust the competition process in the face of changing listening conditions (e.g., Brouwer, Mitterer, & Huettig, 2011; Farris-Trimble, McMurray, Cigrand, & Tomblin, 2014; McQueen & Huettig, 2012; Scharenborg et al., 2018a). During native listening, the activations of candidate words have been shown to change during competition when background noise is present (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, 2016; McQueen & Huettig, 2012) and more words are activated when the start of a word is masked by noise than when a word's ending is masked by noise (Scharenborg et al., 2018a).

Moreover, recent visual-world eye-tracking experiments have shown that the presence of background noise slows down the speed with which the competition process is resolved (Brouwer & Bradlow, 2016; Hintz & Scharenborg, 2016). Hintz and Scharenborg (2016) had native Dutch listeners listen to Dutch target words embedded in semantically neutral, non-predictable carrier sentences (e.g., "*He immediately thought of a 'zwaan'/swan when Bob started talking about geese*") in a quiet listening condition and with SSN at two SNRs. Eye-fixations to a visual display were monitored. Each display consisted of four pictures: either a picture of the target word (e.g., *'zwaan'/swan*) or a picture of a phonological onset competitor, overlapping with the target word in the first few phonemes (e.g., *'zwaard'/sword*), and three pictures of unrelated objects (so-called 'distractors'; e.g., *'toilet'/toilet, 'komkommer'/cucumber, 'fles'/bottle*). The key findings were that the presence of background noise attenuated and delayed looking preferences to the target words and attenuated looking preferences for onset competitors and elongated the time window in which these occurred. Subsequent testing of a group of German non-native listeners of Dutch on the same task found the same attenuation and slowing down of the competition process in the noisy listening condition compared to the quiet listening condition (Karaminis, Hintz, & Scharenborg, in preparation). In other words, the presence of noise results in an attenuation and a slowing down of the competition process in both native and non-native listening.

We hypothesize that this slowing down and attenuation could be explained by a slowing down of the build-up of word activation over time. Possibly, an increase in the number of activated candidate words leads to a slower build-up of the word activation of a specific candidate word compared to when fewer candidate words are activated. Additionally, the word activation level reached by the activated candidate words might be lower in case more candidate words are activated than when fewer candidate words are activated, as if there is a limit on the amount of available word activation. Alternatively, the slowing down of the build-up of word activation over time could be due to a less efficient elimination of candidate words that no longer have a good enough fit with the speech signal, or other mechanisms are at play. Future research will have to shed light on the mechanisms underlying this slowing down and attenuation of the competition process.

7 Concluding remarks

The picture that seems to arise from the current literature on non-native spoken-word recognition in background noise is that 1) indeed, spoken-word recognition in the presence of background noise is harder in a non-native language than in one's native language; 2) that, as suggested by both experimental and modelling studies, the difference is not due to a differential effect of background noise on native and non-native listening, but rather that the difference between native and non-native speech recognition performance is constant and 3) can be explained by differences in language exposure; and 4) that the native and non-native speech processing systems have an identical cognitive architecture. A reduced exposure to the non-native language leads to less well-defined sound categories, which lead to an increase in the number of activated words during non-native listening compared to native listening, and this difference is carried over from quiet to noisy listening conditions. We should note, however, that since most studies have focussed on the effect of energetic masking, the question about the role of informational masking on the cognitive processes underlying native and non-native speech recognition is largely unanswered and remains a topic for future research.

The presence of noise in the speech signal seems to have an effect throughout the speech recognition process: It leads to a degraded speech signal and a subsequent degraded phonological match between the speech signal and the abstract sound representations in the listener's brain, it increases the number of activated words, and the competition process is attenuated and slowed down. Importantly, the presence of background noise also changes the dynamics of the multiple activation and the competition and recognition processes underlying spoken-word recognition. 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.

Acknowledgements

This research was supported by a Vidi-grant from the Netherlands Organization for Scientific Research (NWO; grant number 276-89-003) awarded to Odette Scharenborg. Part of this work was carried out by the second author under the supervision of the first author while both authors were at the Centre for Language Studies at Radboud University.

References

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and non-native listening. Bilingualism: Language and Cognition, 6, 81–96. https://doi.org/10.1017/S1366728903001056

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.

Aydelott, J. & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. Langage and Cognitive Processes, 19, 29-56.

Aydelott, J., Baer-Henney, D., Trzaskowski, M., Leech, R., & Dick, F. (2012). Sentence comprehension in competing speech: Dichotic sentence word priming reveals hemispheric differences in auditory semantic processing. Language and Cognitive Processes, 27(7-8), 1108-1144.

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.

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.

Blutner, R., & Sommer, R. (1988). Sentence processing and lexical access: The influence of the focus-identifying task. Journal of Memory & Language, 27, 359-367.

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

Borghini, G., & Hazan, V. (2018). Listening effort during sentence processing is increased for non-native listeners: A pupillometry study. Frontiers in Neuroscience, 12, 152. Doi:10.3389/fnins.2018.00152

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., & Bent, T. (2002). The clear speech effect for non-native listeners. Journal of the Acoustical Society of America, 112(1), 272-284.

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.

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

Broersma, M., & Cutler, A. (2008). Phantom word recognition in L2. System: An International Journal of Educational Technology and Applied Linguistics, 36, 22–34.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. Quarterly Journal of Experimental Psychology, 64, 74-95.

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

Brouwer, S., & Bradlow, A.R. (2016). 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.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. Journal of the Acoustical Society of America, 109(3), 1101-1109.

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.

Cooke, M. (2006). A glimpsing model of speech perception in noise. Journal of the Acoustical Society of America, 119, 1562-1573.

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

Cooke, M., Barker, J., Cunningham, S., & Shao, X. (2006). An audio-visual corpus for speech perception and automatic speech recognition. Journal of the Acoustical Society of America, 120(5), 2421-2424.

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.

Cooke, M., García Lecumberri, M. L., Scharenborg, O., & van Dommelen, W. (2010). Language-independent processing in speech perception: A study of English consonant identification in 8 European languages. Speech Communication, 52, 954-967.

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, 519-523.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. Language & Speech, 40, 141–201. https://doi.org/10.1177/002383099704000203

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. Journal of Phonetics, 34, 269–284.

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.

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.

FitzPatrick, I., & Indefrey, P. (2010). Lexical competition in nonnative speech comprehension. Journal of Cognitive Neuroscience, 22(6), 1165-1178.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. Journal of Memory and Language, 41(1), 78-104.

García Lecumberri, M. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. Journal of the Acoustical Society of America, 119(4), 2445-2454.

García Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. Speech communication, 52(11-12), 864-886.

Golestani, N., Rosen, S., & Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: The contribution of semantics. Bilingualism: Language and Cognition, 12(3), 385-392.

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.

Hintz, F. & Scharenborg, O. (2016). The effect of background noise on the activation of phonological and semantic information during spoken-word recognition. In Proceedings of Interspeech 2016, pp. 2816–20. San Francisco, CA.

Jin, S. H., & Liu, C. (2012). English sentence recognition in speech-shaped noise and multi-talker babble for English-, Chinese-, and Korean-native listeners. The Journal of the Acoustical Society of America, 132(5), 391-397.

Kaandorp, M. W., De Groot, A. M., Festen, J. M., Smits, C., & Goverts, S. T. (2016). The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise. International Journal of Audiology, 55(3), 157-167.

Kahnemuyipour, A. (2009). The syntax of sentential stress (No. 25). Oxford University Press.

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

Karaminis, T., Hintz, F., & Scharenborg, O. (in preparation). Computational modelling of native and non-native spoken-word recognition in noise.

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

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. Behavior Research Methods, 44(2), 325-343.

Li, P., & Yip, M. C. (1998). Context effects and the processing of spoken homophones. Reading & Writing, 10(3-5), 223-243.

Lidestam, B., Holgersson, J., & Moradi, S. (2014). Comparison of informational vs. energetic masking effects on speechreading performance. Frontiers in Psychology, 5, 639. http://doi.org/10.3389/fpsyg.2014.00639

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. Perception & Psychophysics, 62, 615–25.

Luce, P.A., Pisoni, D.B. (1998). Recognizing spoken words: The Neighborhood Activation Model. Ear & Hearing, 19, 1-36.

Maddieson, I. (1984). Patterns of Sounds. Cambridge University Press, Cambridge.

Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In Altmann & Shillcock et al. (Eds.), Cognitive models of speech processing: The Second Sperlonga Meeting (pp. 187-210). Hove, UK: Lawrence Erlbaum.

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

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. Cognitive Psychology, 59(3), 203-243.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. Journal of Speech, Language, and Hearing Research, 40(3), 686-693.

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

McQueen, J. M. (2005). Speech perception. The handbook of cognition, 255-275.

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

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

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 90(2), 227-234.

Nabelek, A. K. (1988). Identification of vowels in quiet, noise, and reverberation: relationships with age and hearing loss. Journal of the Acoustical Society of America, 84, 476–484.

Nabelek, 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.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition", Journal of Experimental Psychology: Learning, Memory and Cognition, 21(1):1209-1228.

Oliver, G., Gullberg, M., Hellwig, F., Mitterer, H. (2012). Acquiring L2 sentence comprehension: A longitudinal study of word monitoring in noise. Bilingualism: Language & Cognition, 15, 841-857.

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.

Pollack I. (1975). Auditory informational masking. J. Acoust. Soc. Am. 57, S5 10.1121/1.1995329

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.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. Applied Psycholinguistics, 27(3), 465-485.

Rosenberg, A., Hirschberg, J., & Manis, K. (2010). Perception of English prominence by native Mandarin Chinese speakers. Fifth International Conference on Speech Prosody, 2010.

Scharenborg, O., Coumans, J., Kakouros, S., & Van Hout, R. (2016a). Does the importance of word-initial and word-final information differ in native versus non-native spoken-word recognition? Proceedings of Interspeech, 858-862.

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

Scharenborg, O., Coumans, J., & Van Hout, R. (2018a). The effect of background noise on the word activation process in non-native spoken-word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 44(2), 233-249.

Scharenborg, O., Meunier, F., Kakouros, S., & Post, B. (2018b). Sentence accent perception in noise by French non-Native listeners of English. Proceedings of Speech Prosody.

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, 53, 821-835. doi:10.1044/1092-4388(2010/09-0081)

Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. Journal of Experimental Psychology, 102(2), 250-255.

Shimizu, T., Makishima, K., Yoshida, M., & Yamagishi, H. (2002). Effect of background noise on perception of English speech for Japanese listeners. Auris Nasus Larynx, 29(2), 121-125.

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

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.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. Psychological Science, 10(3), 281-284.

Tabri, D., Chacra, K. M., & Pring, T. (2015). Speech perception in noise by monolingual, bilingual and trilingual listeners. International Journal of Language & Communication Disorders, 1-12.

Takayanagi, S., Dirks, D. D., & Moshfegh, A. (2002). Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing. Journal of Speech, Language, and Hearing Research, 45(3), 585-597.

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first-and second-language multi-talker babble. Speech Communication, 52(11-12), 943-953.

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 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.

Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N., & North, B. (2009). Common European Framework of Reference for Languages: learning, teaching, assessment. Cambridge University Press.

Vitevitch, M. S., Rodríguez, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. Journal of Multilingual Communication Disorders, 3(1), 64-73.

Wagner, P. (2005). Great expectations – Introspective vs. perceptual prominence ratings and their acoustic correlates. Proceedings of Interspeech, 2381-2384.

Warzybok, A., Brand, T., Wagener, K. C., & Kollmeier, B. (2015). How much does language proficiency by non-native listeners influence speech audiometric tests in noise? International Journal of Audiology, 54, 88-99.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. Journal of Memory and Language, 50(1), 1-25.

Weber, A., & Scharenborg, O. (2012). Models of processing: lexicon. WIREs Cognitive Science, 3:387-401, doi: 10.1002/wcs.1178.

Zhang, J., Xie, L., Li, Y., Chatterjee, M., & Ding, N. (2014). How noise and language proficiency influence speech recognition by individual non-native listeners. PloS one, 9(11), e113386.

Zwitserlood (1989). The locus of the effects of sentential-semantic context in spoken-word processing. Cognition 32, 25-64.

Appendix A. Overview of all the studies investigating word recognition in noise by non-native listeners discussed in this review.

Study	Target language	Listener L1	Type of noise	Noise levels	Task	Proficiency test	Research question
Bradlow & Alexander (2007)	English	English, various L1s	SSN	-2 dB (Ns); +2 dB (NNs)	High- + low-predictability sentences; clear + plain speaking style	SPEAK	Can the non-native disadvantage be overcome if enhancements at various levels of processing (semantic, acoustic) are presented in combination?
Bradlow & Bent (2002)	English	English, various L1s	White noise	-8, -4 dB	BKB-R	TOEFL, SPEAK	To what extent is naturally produced clear speech an effective intelligibility enhancement strategy for non-native listeners?
Brouwer et al. (2012)	Dutch, English	Dutch, English	2-speaker babble	-5, -3 dB (Ns); -3, -1 dB (NNs)	BKB-R	/	Does speech-in-speech recognition vary across different background speech languages, across variation in the semantic content of the background speech, and across variation in listener status regarding the languages)?
Cooke et al. (2008)	English	English, Spanish	SSN, 1-speaker babble	-6, 0, +6 dB (SSN); -9, -6, -3,	Keywords in short sentences (GRID corpus)	/	What are the differential effects of energetic and informational masking on native and non-native speech perception?

				0, +3, +6, +9 dB (babble)			
Coumans et al. (2014)	English	Dutch, English	SSN	-12, -6, 0 dB	Isolated words	LexTALE	Are non-native listeners able to use both word-initial and word-final information during spoken-word recognition, and are these equally important when listening conditions deteriorate?
Ezzatian et al. (2010)	English	English, various L1s	SSN	NS: -12, -8, -4, 0 dB; NNS: -12, -6, 0, +6 dB	Semantically unpredictable sentences	Mill Hill; Nelson-Denny	Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking?
Golestani et al. (2009)	English, French	French	SSN	-7. -6, -5, -4 dB	Sentences with semantically related / unrelated primes	/	Does the semantic component of language contribute to the native language advantage when listening to degraded speech?
Kaandorp et al. (2015)	Dutch	Dutch, various L1	SSN fluctuating noise	Variable (SRT)	VU98 meaningful sentences; digits-in-noise	CEFR	What is the effect of linguistic abilities (lexical access ability and vocabulary size) in different measures of speech-in -noise recognition in normal-hearing listeners with various levels of language proficiency?

Kilman et al. (2014)	English, Swedish	Swedish	SSN, fluctuating noise, 2-speaker babble	Variable (SRT)	HINT	Standardized English proficiency test	What is the influence of non-native language proficiency on speech perception performance?
Jin & Liu (2012)	English	Chinese, English, Korean	SSN, multispeaker babble	-10, -5, 0, +5 dB (SSN); -15, -10, -5, 0 dB (babble)	HINT	TOEFL	How do speech-shaped noise and multi-talker babble noise affect English sentence recognition for English-, Chinese-, and Korean-native listeners?
Mayo et al. (1997)	English	English, Spanish-English bilinguals	12-speaker babble	Variable (SRT)	SPIN, high + low predictability	/	How does age of L2 acquisition influence speech perception in noise?
Meador et al. (2000)	English	English, Italian-English bilinguals	Pink noise	-6, 0, +6 dB	Semantically unpredictable sentences	/	How do AOA and amount of L1 use affect recognition of English words by groups of native speakers of Italian?
Rogers et al. (2006)	English	English, Spanish-English bilinguals	SSN, SSN + reverberation	-6, -2, 0 dB (SSN); 0, +2, +4 dB (SSN + reverb.)	CID W22	/	What are the effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing?
Scharenborg et al. (2016)	English	Dutch, English, Finnish	SSN	-12, -6, 0 dB	Isolated words	LexTALE	Does the importance of word-initial and word-final information differ in native versus non-native spoken-word recognition?

Scharenborg et al. (2017)	English	Dutch, English	SSN	-12, -6, 0 dB	Isolated words	LexTALE	What is the effect of background noise on the word activation process in non-native spoken-word recognition?
Shi (2010)	English	English, various L1s	12-speaker babble + reverberation	SNR: 0, +6 dB; RT: 1.2, 3.6 s	SPIN	/	What are the effects of background noise, reverberation and context on sentence perception in listeners differing in age of English acquisition?
Shimizu et al. (2001)	English	Japanese	White noise, pink noise, aircraft noise	-4, +1, +6 dB	CID W22	/	What is the effect of background noise on the discrimination of English speech for Japanese listeners?
Tabri et al. (2010)	English	English, Arabic-English bilingals, Arabic-French-English trilinguals	12-speaker babble	0, +5, +10, +15, +20 dB	SPIN	/	What is the performance of monolingual, bilingual, and trilingual speakers on an English speech perception task in both quiet and noisy conditions?
Takayanagi et al. (2002)	English	English, various L1s	9-speaker babble	Variable (SRT)	Isolated CVC words	/	What are the effects of talker variability and lexical difficulty on spoken-word recognition among native and non-native listeners with normal and impaired hearing?

Van Engen (2010)	English	English, Mandarin Chinese	2-speaker babble	Variable (SRT -6, -3, +0, +3 dB)	BKB-R	TOEFL, LEAP-Q	What are the effects of first- and second-language talker babble on second language sentence recognition?
Van Wijngaarden et al. (2002)	Dutch, English, German	Dutch	SSN	Variable (SRT)	SRT	/	What is the intelligibility of speech in noise for non-native listeners?
Warzybok et al. (2015)	German	German, various L1s	SSN	Variable (SRT)	DTT, OLSA, GÖSA	CEFR	How much does language proficiency by non-native listeners influence speech audiometric tests in noise?
Zhang et al. (2014)	English, Mandarin Chinese	Mandarin Chinese	SSN	-13, -10, -4 dB (Ns); -6, -2, +2 dB (NNs)	HINT, MSP	/	How do noise and language proficiency influence speech recognition by individual non-native listeners?