Original Article

Trends in Hearing
2015, Vol. 19: 1–12
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216515579127
tia.sagepub.com
SAGE

Native and Non-native Speech Perception by Hearing-Impaired Listeners in Noise- and Speech Maskers

Lisa Kilman[1,2], Adriana Zekveld[1,2,3], Mathias Hällgren[2,4], and Jerker Rönnberg[1,2]

Abstract

This study evaluated how hearing-impaired listeners perceive native (Swedish) and nonnative (English) speech in the presence of noise- and speech maskers. Speech reception thresholds were measured for four different masker types for each target language. The maskers consisted of stationary and fluctuating noise and two-talker babble in Swedish and English. Twenty-three hearing-impaired native Swedish listeners participated, aged between 28 and 65 years. The participants also performed cognitive tests of working memory capacity in Swedish and English, nonverbal reasoning, and an English proficiency test. Results indicated that the speech maskers were more interfering than the noise maskers in both target languages. The larger need for phonetic and semantic cues in a nonnative language makes a stationary masker relatively more challenging than a fluctuating-noise masker. Better hearing acuity (pure tone average) was associated with better perception of the target speech in Swedish, and better English proficiency was associated with better speech perception in English. Larger working memory and better pure tone averages were related to the better perception of speech masked with fluctuating noise in the nonnative language. This suggests that both are relevant in highly taxing conditions. A large variance in performance between the listeners was observed, especially for speech perception in the nonnative language.

Keywords

speech perception, native and nonnative, noise- and speech maskers, nonnative language proficiency, cognitive abilities

Introduction

It is often difficult to understand native speech in noisy environments, and the difficulty even increases when the speech is in a nonnative language. This is a well-known phenomenon and has been extensively studied over the years (e.g., Mayo, Florentine, & Buus, 1997; Van Engen, 2010; van Wijngaarden, Steeneken, & Houtgast, 2002). If an individual faces the additional challenge of hearing impairment, this entails another dimension to speech perception difficulties. Hearing loss can impair the ability to perceive a native conversation even in quiet, and the problem generally increases in noisy conditions (Ng, Rudner, Lunner, & Rönnberg, 2015). How language, interfering maskers, and hearing loss interact in speech perception has, to our knowledge, not been examined in a single study before. In such complex listening conditions, a number of factors may affect speech perception. These include individual variables like age, cognitive abilities, and pure tone average (PTA) along with different target speech (native and nonnative language) and masker characteristics, involving *energetic* or *informational* masking effects. The term *energetic masking* refers to the distorting effects caused by spectro-temporal overlap between the target speech and masker signal (Brungart, 2001). *Informational masking* (see Kidd, Mason, Richards, Gallun, & Durlach, 2007, for a review) refers to any masking effects that go beyond what can be explained by energetic masking.

[1]Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden
[2]Linnaeus Centre HEAD, The Swedish Institute for Disability Research, Linköping University, Linköping, Sweden
[3]Audiology/ENT & EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands
[4]Department of Otorhinolaryngology/Section of Audiology, Linköping University Hospital, Linköping, Sweden

Corresponding author:
Lisa Kilman, Department of Behavioral Sciences and Learning, Linköping University, S-581 83 Linköping, Sweden.
Email: lisa.kilman@liu.se

Creative Commons CC-BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 License (http://www.creativecommons.org/licenses/by-nc/3.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (http://www.uk.sagepub.com/aboutus/openaccess.htm).

While energetic maskers typically affect the audibility of speech, informational maskers increase the difficulty of perceptually and cognitively segregating the target from the masker. Generally, stationary and fluctuating maskers can produce energetic (Stone & Moore, 2014; Stone, Fullgrabe, & Moore, 2012) and modulation masking (Oxenham & Kreft, 2014; Stone et al., 2012), and competing babble can produce additional informational masking (Freyman, Helfer, Mc Call, & Clifton, 1999). Several studies have explored the relationships between informational masking and hearing impairment. Most of these studies indicated either a reduced or absent impact of informational masking effects in hearing-impaired (HI) listeners (Alexander & Lufti, 2004; Arbogast, Mason, & Kidd, 2005; Micheyl, Arthaud, Reinhart, & Collet, 2000). However, other studies have suggested that informational maskers distort speech perception especially in older HI listeners (Helfer & Freyman, 2008). The latter seems to be true for older, normal-hearing (NH) listeners as well. Helfer and Staub (2013) assessed eye-movement patterns and showed that older adults (60–81 years of age) were more affected by competing speech than younger listeners (20–25 years of age). However, even middle-aged (51–63 years) NH listeners can have more difficulties than young (19–26 years) listeners, as shown in the study of Baskent, van Engelshoven, and Galvin (2014). The speech reception threshold (SRT) was 2.1 dB worse for the middle-aged listeners than for young listeners in that study. Baskent et al. (2014) suggested that age-related effects on speech perception are already manifest in middle-aged adults (see also Ruggles, Bharadwaj, & Shinn-Cunningham, 2012).

The primary predictor of speech perception in noise by HI listeners is the degree of hearing loss (Akeroyd, 2008). However, the perception of masked speech is a complex process; besides auditory functioning, speech perception involves cognitive functions (Akeroyd, 2008; Desjardin & Doherty, 2013). The relevance of cognitive functions in speech perception, such as selective attention, processing speed, and working memory capacity, has been recognized in many studies (Akeroyd, 2008; Gatehouse, Naylor, & Elberling, 2003; George et al., 2007; Rönnberg, 2003). Working memory is important in language comprehension as it allows listeners to simultaneously maintain and process information (Baddeley, 2012; Daneman & Carpenter, 1980; Daneman & Merikle, 1996). It is also relevant for tasks like problem solving and reasoning (Engle, 2002).

When it is difficult to perceive the speech (e.g., for HI listeners in noisy surroundings), explicit cognitive processes such as working memory become increasingly relevant. The Ease of Language Understanding (ELU) model defines the relationship between *implicit* and *explicit* functions when language is processed (Rönnberg, 2003; Rönnberg, Rudner, Foo, & Lunner, 2008; Rönnberg et al., 2013). If the signal is clear, then a match occurs between input and stored representations in semantic long-term memory and lexical access proceeds easily (*implicitly*). However, under challenging conditions, it is more likely that a mismatch between input and stored representation follows, and *explicit*, cognitive processes must be engaged in an attempt to decode the signal. The probability of a mismatch defines the relation between implicit (bottom-up) and explicit (top-down) activity during speech understanding (Foo, Rudner, Rönnberg, & Lunner, 2007). Older HI listeners generally have greater decoding (or phonological) difficulties in understanding speech in noisy conditions than younger HI listeners (Dubno & Ahlstrom, 1997; Schneider, Li, & Daneman, 2007). Therefore, the probability that older HI listeners will depend on explicit processing resources is greater than for younger listeners. It has been suggested that deterioration of the peripheral auditory system in combination with age-related changes in cognitive processing are important factors underlying speech perception difficulties in older adults (Rudner, Rönnberg, & Lunner, 2011; Schneider et al., 2007).

An additional factor relevant for the perception of speech in interfering maskers is the listeners' ability to segregate auditory streams. For example, when the target and the babble masker are in the same language, the two simultaneous speech streams may be difficult to separate. The difficulties experienced in such conditions may reflect problems with perceptually isolating one stream from the other, or because attention switches back and forth between the target and the masker signal (Schneider et al., 2007). When there is information in the masker that can be linked to the information in the target, as when both target and the masker are in the same language, the semantic and linguistic interference is relatively large. In fact, the more similar the target and the masker are, the more difficult it is to keep apart the two streams efficiently (Brouwer, Van Engen, Calandruccio, & Bradlow, 2012). Therefore, when the masker speech is unfamiliar or is in another language than the target signal, this usually improves performance relative to familiar speech maskers (Brouwer et al., 2012; Van Engen, 2010). Furthermore, when the competing speech is spoken in a language from the same rhythmic class as the native language (e.g., German for English-speaking listeners), it is more difficult to ignore than if it is from a different rhythmic class (Spanish or French for English-speaking listeners; Reel & Hicks, 2012).

The general aim of the current study was to assess how HI listeners perceive native (Swedish) and nonnative (English) speech in the presence of energetic and informational maskers.

We estimated SRTs in HI listeners who perceived native and nonnative speech in noise- and babble maskers. The SRT is a sensitive measure, and the adaptive procedure avoids the risk of ceiling or floor effects in the listeners' performance (Plomp & Mimpen, 1979). Two types of maskers were applied that mainly result in energetic and modulation masking effects: stationary- and fluctuating noise maskers. We additionally applied speech maskers: two-talker babble in Swedish and two-talker babble in English. The reason for using two-talker babble was that it produces strong informational masking effects (Brungart, Simpson, Ericson, & Scott, 2001; Calandruccio, Dhar, & Bradlow, 2010; Van Engen & Bradlow, 2007).

Our second objective in this study was to assess which individual variables are associated with native and nonnative speech perception in noise by HI listeners.

In a previous study, we evaluated nonnative speech perception performance in participants with NH (Kilman, Zekveld, Hällgren, & Rönnberg, 2014). The main result of that study was that nonnative language proficiency was strongly associated with nonnative speech intelligibility in noise. Therefore, in the current study, we also assessed nonnative (English) language proficiency of the HI listeners using a standardized proficiency test. In line with the previous study, we created two subgroups based on the HI listeners' performance on this proficiency test.

In addition to nonnative language proficiency, we assessed how other individual variables, such as pure tone hearing thresholds (PTA), working memory capacity, and nonverbal intelligence, were related to native and nonnative speech perception. We included both a Swedish (native) and an English (nonnative) version of the Reading Span (working memory) test in order to measure whether the performance on the two tests was related to SRTs in the two target languages. For the assessment of nonverbal intelligence, we included Raven standard progressive matrices (Raven, Raven, & Court, 2000).

We expected that HI listeners would have more difficulties with the perception of the nonnative English target speech than the native Swedish target speech. We also expected speech maskers to interfere more than energetic maskers with speech perception in both target languages, as this was the case for the NH listeners in Kilman et al. (2014). We expected that higher English proficiency and better performance on the cognitive tests would be associated with better speech perception. In our previous study (Kilman et al., 2014), there was an interaction effect between language and English-proficiency subgroup. We therefore hypothesized an interaction effect for the HI listeners as well.

Methods

Participants

Twenty-three participants (14 females, 9 males) with an average age of 50.1 years ($SD = 10.2$, range = 28–65) took part in the experiment. The participants were recruited from the audiology clinic at Linköping University Hospital, Sweden. Inclusion criteria for all participants were that they had an acquired bilateral, sensorineural hearing impairment with no severe tinnitus complaints, that they were native Swedish speakers and had learned English as NH children in elementary school. All participants filled out a questionnaire in which they answered questions about their knowledge and training in English and how frequently they were using English in daily life (see Table 1). The project was approved by the Ethics Committee in Linköping, and all the listeners provided written informed consent. The listeners received a small gift for taking part.

Linguistic profile, education, and PTA. To be able to define the participants' linguistic (nonnative) profile, the following information about the participants was assessed: language history, language status, language stability, language competence, and demand for language use (for more details, see von Hapsburg & Peña, 2002).

The Swedish school system consists of compulsory primary school (7–16 years of age) and optional secondary school (16–19 years of age). The participants' number of years of education varied between 8 years and 21.5 years ($M = 13.7$ years). The participants started learning English between 9 and 12 years of age in grade 3 to 6: grade 3 ($n = 4$), grade 4 ($n = 16$), grade 5 ($n = 2$), grade 6 ($n = 1$). English has existed as an educational subject in the Swedish school system since the 1880s, but became the first foreign language in 1939 (German before). From the 1950s, English was compulsory from grade 5. English has since then been compulsory. The participant with 8 years of education had completed school in a period when compulsory education took only 8 years. The participants with an educational level of 21.5 years had obtained a PhD degree. During the past decades, the teaching time, as well as when students begin to learn English, has varied only marginally. The most common situation is to begin with English in the fourth grade and continue through the end of secondary school. All participants considered English as their primary secondary language.

Nonnative language competence was tested in a standardized comprehension test (see below). According to the test results, participants were divided into a high- and a low-proficiency subgroup. The frequency with which the participants used English in daily life varied between *never* and *daily*. The "H" and "L" within the brackets stand for High or Low proficiency as determined with the proficiency test: daily (3H, 1 L), every week (3H, 2 L), every month (1 L), every year (2H, 5 L), holidays (1H), never (2H, 3 L).

Table 1. Participant Information.

Part	Age	PTA_4	Educ/y	BegEng/gr	Educ Eng	Use of Eng	Profic Eng
1	65	36.8	9	5	PrimSch	Yearly	Low
2	28	55.6	19	4	SecSch	Daily	High
3	38	54.8	14	4	SecSch	Never	High
4	50	43.1	16,5	4	SecSch	Weekly	Low
5	65	38.8	8	6	PrimSch	Never	Low
6	55	59.4	14.5	4	SecSch	Weekly	High
7	60	44.4	14	3	SecSch	Never	High
8	59	52.5	14	4	PrimSch	Never	Low
9	51	39.8	11	4	SecSch	Yearly	Low
10	54	50.6	11	4	SecSch	Yearly	Low
11	30	55.6	17	3	SecSch	Daily	Low
12	57	45.6	12	4	Courses	Yearly	High
13	48	71.3	11	3	FCE	Weekly	High
14	60	60.0	14	4	Courses	Weekly	Low
15	54	47.5	14.5	4	SecSch	Monthly	Median
16	47	38.1	21.5	4	SecSch	Yearly	High
17	45	25.0	16	4	FCE + Univ	Daily	High
18	43	35.0	21.5	4	Univ	Daily	High
19	42	28.1	16.5	4	Univ	Holid	High
20	56	50.0	14	4	SecSch	Never	Low
21	61	52.5	11	5	SecSch	Yearly	Low
22	44	48.8	11	3	SecSch	Yearly	Low
23	55	41.8	15	4	SecSch + Courses	Weekly	High

Note. Part = participant number; PTA_M = 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz; Educ = education; Eng/gr = English, grade; Profic = proficiency; PrimSch = primary school; SecSch = optional secondary school; FCE = the first certificate in English (Cambridge English language assessment); Univ = English at University level; Holid = holidays.

PTA thresholds for both ears at the frequencies 500, 1000, 2000, 4000 Hz are shown for each participant in Table 1. The average across frequencies (PTA_4) was 46.7 dB HL ($SD = 10.7$). The PTA_4 ranged from 25.0 dB HL to 71.3 dB HL. The average degree of hearing loss varied from slight (16–25 dB; $n = 1$) through mild (26–40 dB; $n = 6$), moderate (41–55 dB; $n = 13$), moderately severe (56–70 dB; $n = 2$) to severe (71–90 dB; $n = 1$; Clark, 1981).

Table 1 provides information about each participants' age, mean PTA_4, number of years of (total) education, the grade when participants began to learn English, the level of the education in English, English language use (frequency), and if the participant was high or low proficient in English as indicated by the proficiency test.

Two participants reported never using English but still appeared to be highly proficient in the language. These participants reported skill and interest in the English language since childhood, although they did not develop this interest. Their current professions did not require using English on a regular basis.

Experimental Test and Stimuli

The SRT test was developed by Plomp and Mimpen (1979) and was applied to measure sentence intelligibility. SRTs were measured using either American English Hearing In Noise Test (HINT; Nilsson et al., 1994) or Swedish HINT (Hällgren, Larsby, & Arlinger, 2006) target sentences. Each HINT sentence set consists of phonemically balanced sentences grouped in 25 lists with 10 sentences in each. The material comprises short everyday sentences, judged to be natural by native speakers. The sentences were recorded by male native speakers. The listeners performed eight test conditions: English or Swedish target language, either one combined with one of four types of masker: a stationary masker, a fluctuating masker, two-talker babble Swedish, and two-talker babble English (see description below). Each condition contained 20 sentences, and every new condition started with several practice sentences: The first condition had 10, and the following had five practice sentences each. The sequence of conditions was counterbalanced across listeners, and each sentence was used only once per listener.

The presentation levels of the target and masker signals were individually adjusted offline according to the Cambridge prescription formula (Moore & Glasberg, 1998), based on the pure tone thresholds of the best ear. The masker level was changed in a stepwise two-up-two-down adaptive procedure (Plomp & Mimpen, 1979) targeting 50% sentence intelligibility. Masker onset was 3 s before speech onset, and masker offset was 1 s after speech offset. The participants were asked to repeat each sentence aloud, and the experimenter scored whether all words in the sentence were reproduced correctly.

The stationary masker was similar to the speech-shaped maskers that were developed by Nilsson et al. (1994) and Hällgren et al. (2006). The spectrum of the masker was shaped according to the long-term average spectrum of the speech material of the corresponding set (same procedure for Swedish and English).

The fluctuating masker was created from the speech-shaped noise of the target language and had the same envelope fluctuations as the two-talker babble in Swedish or English. These envelopes were extracted by applying the Hilbert transform and a low-pass filter with cut-off frequency of 32 Hz (first order Butterworth filter, 6 dB/octave), in line with Agus, Akeroyd, Gatehouse, and Warden (2009). The two fluctuating maskers that were used were spectrally matched to the target language in Swedish or English respectively, and also matched temporally to the babble in Swedish or English, respectively.

The two-talker babble in Swedish included one native Swedish male and one native Swedish female reading from Swedish newspapers. The two-talker babble in English included one male native British English speaker and one female American English speaker reading from articles in English/American newspapers. The two-talker babbles were created by mixing the sound-tracks from the female and male speakers in Swedish and American/English, respectively. The speech maskers were spectrally matched to the long-term spectrum of the target speech presented (Swedish or English).

Cognitive and Language Tests

Reading Span. The Reading Span test is a measure of working memory capacity (Daneman & Carpenter, 1980; Rönnberg, Lyxell, Arlinger, & Kinnefors, 1989). In the test, short sentences were presented word-by-word on a computer screen. Half of the sentences made sense (The pupil arrived late) and the other half did not (The pear went out). Immediately after each sentence, a question appeared on the screen asking whether the sentence made sense or not. The participants answered by button presses, yes or no. Sentences were presented in sets. The set sizes progressively increased from three sentences in the sets to five sentences in each set. After each set, the participants were asked to orally recall either the first or the last word in the sentences. The participants did not know in advance which words (the first or the last) they had to report. We scored the number of words correctly recalled, regardless of order (max score = 23). The Reading Span test was assessed in Swedish and English. The order of the Reading Span tests in both languages was counterbalanced across participants.

Nonverbal reasoning ability. The Raven standard progressive matrices (Raven et al., 2000) assess nonverbal reasoning ability (fluid intelligence). The test is a multiple-choice measure of 60 matrices divided into five sets, A to E. The task is to identify what missing piece, to be selected from given alternatives, best completes a larger pattern. The participants performed sets B to D. Every set is progressively more difficult than the previous set, and the difficulty also increases within each set (max score = 36).

English-proficiency test. The test assesses English language comprehension and is a standardized, national test, essentially developed for the optional Secondary School level (www.skolverket.se/prov-och-bedom, http://www.nafs.gu.se/digitalAssets/1193/1193558_last_exp.pdf). The English test consists of a text and two sets of tasks, one set with questions to answer in the participants' own words and one set with sentences in which the bold printed words should be explained with only one final word in the open end (e.g., *If you* ***brush off*** *criticism, it means that you don't want to* _______) (listen) (max score = 12).

English proficiency groups. Two subgroups were created based on the English-proficiency test results. The median was 7 so participants with scores < 7 formed *low-proficiency group*, and participants with scores > 7 formed *high-proficiency group*.

Procedure

Test administration took place in one session of approximately 3.5 hr. The test session started with an audiometric test, carried out by an experienced audiologist, followed by the SRT tests, and was finished after the cognitive test battery. The order of the cognitive tests was counterbalanced. The listeners received oral instructions prior to each test. The auditory tests took place in a sound-treated room, and the cognitive tests were performed in a quiet nearby room. The auditory stimuli were presented over headphones to both ears (Sennheiser HD600).

Statistical Analyses

First, the descriptive statistics of SRTs in Swedish and English in the four noise conditions were calculated.

Then, we performed three separate repeated-measures analyses of variance (ANOVA). The first included SRTs for *Swedish as target language* as the dependent variables with the masker type (stationary noise, fluctuating noise, Swedish babble, English babble) as the within-subject independent factor. The second ANOVA was similar, but now included SRTs for *English as target language* as the dependent variable. The third ANOVA included all eight SRTs (i.e., for both Swedish and English as target language) as the dependent variables. Again, *masker type* was included as a within-subject factor, and *language* and *English-proficiency group* were between-subject factors. In each of the three analyses, we also included PTA_4 as a covariate to examine whether PTA_4 significantly influenced the relationship between the independent and dependent factors. The three ANOVA's allowed us to test the impact of the different masker types on the native and nonnative languages separately, as well as possible interactions between the two languages, the masker types, and proficiency group.

Additionally, we assessed the descriptive statistics of performance on the English-proficiency test, Raven progressive matrices, and the Reading Span tests in Swedish and English. Finally, we assessed the associations between age, number of years of education received, PTA_4, English proficiency, performance on the Raven matrices, Swedish Reading Span, English Reading Span on the one hand, and the SRTs in the 2×4 conditions on the other hand.

Results

SRT in Noise

The means and standard deviations of SRTs in the two target languages are shown in Table 2 for each of the masker types. As can be seen, SRTs for the English target language are consistently higher than SRTs for the Swedish target language. This indicates more difficulties with nonnative as compared with native target language perception.

The first ANOVA, with SRTs for *Swedish as target language* as the dependent variable demonstrated a main effect of *masker type* (stationary noise, fluctuating noise, Swedish babble, English babble): $F(3, 66) = 75.25$, $p < .001$. The pairwise comparisons (Bonferroni adjusted for multiple comparison at the 0.05 level) showed that the Swedish babble and the English babble did not differ significantly (t $(22) = 2.84$, $p = .06$). The Swedish babble was more interfering than the fluctuating noise ($t(22) = -9.44$, $p < .001$), and the English babble was also more interfering than the fluctuating noise ($t(22) = -5.25$, $p < .001$). Finally, the stationary noise was the least interfering masker, as compared with the fluctuating noise ($t(22) = -6.30$, $p < .001$). When PTA_4 was included as a covariate, the effect of masker type became insignificant: $F(3, 63) = 0.62$, $p = .61$.

A separate ANOVA was conducted to assess the impact of *masker type* when the *target language was English*. The results demonstrated a main effect of masker type: $F(3, 66) = 28.20$, $p < .001$. The pairwise comparisons (Bonferroni adjusted for multiple comparison at the 0.05 level) again showed that the Swedish babble and the English babble did not differ significantly ($t(22) = 2.10$, $p = .3$). The Swedish babble was more interfering than the fluctuating noise ($t(22) = -10.57$, $p < .001$), and the English babble was also more interfering than the fluctuating noise ($t(22) = -4.6\, p < .001$). Fluctuating and stationary noise were equally interfering ($t(22) = -.58$, $p > .05$). When PTA_4 was included as a covariate, the effect of masker type was still significant: $F(3, 63) = 3.13$, $p = .032$.

The third ANOVA included SRTs for both target languages as the dependent variables, and target language, *masker types* (stationary noise, fluctuating noise, Swedish babble, and English babble), and *language proficiency-subgroup* as independent factors. The results demonstrated a main effect of *target language*: $F(1, 20) = 110.6$; $p < .001$. This effect indicated that speech recognition was better when the target language was Swedish as compared with English. This ANOVA also showed an interaction effect between *target language* and *masker type*: $F(3, 60) = 3{,}4$, GG-corr $p < .005$, $\varepsilon = 0.64$. We performed post hoc Bonferroni-corrected *t* tests to examine the origin of this interaction effect. The effect of *target language* (Swedish vs. English) was larger for the stationary masker as compared with the fluctuating masker ($t(22) = 4.2$, $p < .05$). Between the other masker types, the effect of *target language* did not statistically differ ($p > .05$). Finally, the ANOVA showed an interaction effect between *target language* and *English-proficiency group,* $F(1, 20) = 22.1$; $p < .001$, indicating that the effect of English proficiency depended on the *target language*, with larger differences between the groups for the English *target language*. No main effects of *English-proficiency group* and *masker type* were observed. When PTA_4 was included as a covariate, there was still an interaction effect between target language and English-proficiency group: $F(1, 19) = 15.68$;

Table 2. Means and SDs (Between Parentheses) of SRTs in Swedish and English.

	SRT-stat	SRT-fluc	SRT-BS	SRT-BE
Swedish target language	−1.9 (2.2)	0.1 (1.9)	2.8 (2.2)	1.6 (2.3)
English target language	3.9 (4.7)	4.1 (4.4)	7.3 (4.4)	6.3 (3.9)

Note. SRT = speech reception threshold; stat = stationary masker; fluc = fluctuating masker; BS = babble Swedish; BE = babble English.

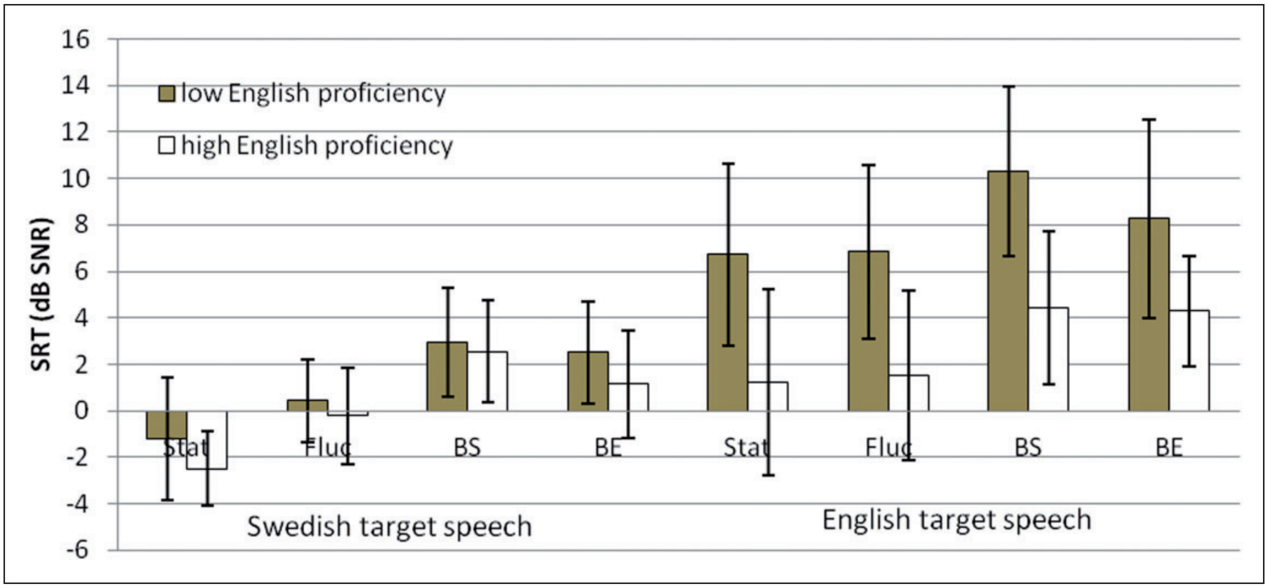

Figure 1. The mean SRTs for the high and low English-proficiency group in each of the eight conditions. Error bars reflect +/− 1 standard deviation. SRT = speech reception threshold; Stat = stationary masker; Fluc = fluctuating masker; BS = babble Swedish; BE = babble English.

Table 3. Means and SDs in Cognitive Performance (SDs Within Parentheses).

English proficiency	7.3 (3.4)
Raven	30.0 (5.5)
Swedish RSpan	13.1 (3.4)
English RSpan	10.3 (4.2)

Note. RSpan = reading span.

$p < .001$. No other significant results were observed. Figure 1 shows the performance in native and nonnative target speech of the high- and low-proficiency groups in the four SRT conditions.

Cognitive Performances

Table 3 shows the descriptive statistics of the performance on the English proficiency, Raven, Swedish Reading Span, and English Reading Span tests.

The relationships between age, number of years of education, PTA_4, English proficiency, Raven, Swedish Reading Span, English Reading Span, and SRTs in the eight conditions were investigated. The Spearman correlation coefficients (see Table 4) show that English proficiency was associated with each of the four SRTs (stationary noise, fluctuating noise, babble Swedish, babble English) in the English target language. English proficiency was also associated with the other cognitive tests (Raven, Swedish Reading Span, and English Reading Span). To gain more insight into what extent English proficiency influenced the relationships among the other variables, we assessed whether the relations among the other variables remained significant when English proficiency was controlled for. To this end, we performed a partial correlation analysis. The correlation coefficients in Table 4, marked with ■ remained significant when English proficiency was controlled for. Swedish Reading Span was still related with SRT for English target speech masked with fluctuating noise, suggesting that larger working memory capacity is related to better perception of English target speech in fluctuating noise. Swedish Reading Span was also associated with Raven, indicating a relation between better nonverbal cognitive capacity and larger working memory capacity when assessed in the native language (Conway, Kane, & Engle, 2003). Higher educational level was related to better SRTs when the English target speech was masked by Swedish babble. Higher age and lower education were also related to each other when controlling for PTA_4, suggesting that the older listeners had a lower education in general.

When English proficiency was controlled for, each of the associations between SRTs, the other cognitive measures, and English Reading Span disappeared. This suggests that these factors did not explain any variance in SRTs over and above the variance explained by English proficiency.

Discussion

The current study assessed how HI listeners perceived native (Swedish) and nonnative (English) speech in the

Table 4. Spearman Correlation Coefficients Between Age, Education, PTA_4, the Cognitive Tests, and the Speech Reception Threshold (SRT) Tests.

	Age	Educ	PTA_4	En prof	Raven	SwRsp	EnRsp
Educ	−.62** ■			.55**			
En prof	−.52*				.71**		.76**
Raven						.66**■	.63**
SwRSp				.59**			.64**
Sw SRT stat		−.44*					
Sw SRT fluc			.59**■				
Sw SRT BS			.62**■				
Sw SRT BE			.72**■				
En SRT stat	.59**	−.59**		−.72**	−.55**		
En SRT fluc	.58**	−.69**	.59**■	−.71**	−.51*	.49*■	−.42*
En SRT BS	.56*	−.68**■	.62**	−.76**	−.64**		−.52*
En SRT BE	.51*	−.62**		−.65**			

Note. The cells marked with ■ reflect the remaining significant variables after controlling for English proficiency. Educ = education; PTA = pure tone average; En prof = English proficiency; SwRSp = Swedish reading span; EnRSp = English reading span; SRT = speech reception threshold; stat = stationary masker; fluc = fluctuating masker; BS = babble Swedish; BE = babble English.
*$p < .05$. **$p < .01$.

presence of energetic and informational maskers. To better understand this complex relationship, we estimated SRTs for the two target languages in different masker types, including babble maskers in each of the two languages. We also acquired a variety of cognitive measures, including objectively tested English language proficiency. Additionally, we assessed the relationship between PTA_4, English proficiency and SRTs to evaluate if these relations differed between the two target languages.

Not surprisingly and as predicted, there was a main effect of language, indicating more difficulty for English as compared with Swedish as target language. This is consistent with the results from NH listeners, as described in Kilman et al. (2014). This main effect of target language could also include effects of, for example, sentence difficulty, specific difficulties associated with the particular speaker of the target speech, and differences in the quality of the recording between the languages. However, these factors probably had minor effects on the observed results.

There was also a main effect of masker type for both English and Swedish target speech perception. Performance was poorer when the masker consisted of Swedish or English babble. Similar performance for both speech maskers were also observed for the English target language in NH listeners (Kilman et al. 2014). The Swedish babble was the most interfering masker for the Swedish target speech in NH listeners. The interference from speech maskers is consistent with previous studies (Calandruccio et al., 2010; Van Engen & Bradlow, 2007). The difficulties may be due to (a) the linguistic similarity between the target and the masker, (b) potentially intelligible words in the language maskers, or (c) both. However, it is not possible to say if the difficulties with the speech maskers for the HI listeners in the current study and the NH listeners in Kilman et al. (2014) derive from the same source. The NH listeners can probably distinguish more words from the interfering two-talker speech maskers than the HI listeners. Still, the HI listeners in this study had the presentation level of the target and the masker adjusted individually, according to the Cambridge formula (Moore & Glasberg, 1998). Previously described results suggest that the use of hearing aids could increase the informational masking effects, as informational masking increases when the sensation level increases (Alexander & Lufti, 2004). The similar linguistic streams represent a highly difficult condition, taxing attention as well as perceptional processes. Attention is required to focus on one auditory stream, while perceptual processes are needed for separating the streams and interpreting the meaning from the target stream. Nevertheless, when the masker speech is unknown or in another language, the listeners usually show a release from masking (Calandruccio et al., 2010). This was the case for the NH listeners in Kilman et al. (2014) when the target was Swedish and the babble was English. In the current study, the HI listeners performed equally poorly in the two babble maskers for both target languages. The fact that Swedish and English belong to the same rhythmic class (Reel & Hicks, 2012) might contribute to this result. It is also likely that impaired spectral and temporal resolution (Moore, 1985) provides an additional explanation that the hearing aid cannot compensate for, as it involves a reduced ability to distinguish between different sounds.

In general, performance was better for the HI listeners in the noise maskers than in the speech maskers in both target languages. However, in the Swedish target language, the fluctuating noise was more interfering than the stationary noise, while in the English target language, the noise maskers did not differ significantly. The fact that there is no benefit from fluctuations in the noise in the native language is consistent with previous studies in HI listeners (Festen & Plomp, 1990; Versfeld & Dreschler, 2002). The similar difficulty experienced for the two noise-maskers for English as target language indicated that the effect of target language (Swedish vs. English) was larger for the stationary masker than for the fluctuating masker. The difference in SRTs between Swedish and English target languages was approximately 6 dB for the stationary masker and around 4 dB for the fluctuating masker. One explanation for the larger masking effect of the stationary masker might be the lower proficiency in the nonnative target in combination with the absence of dips in stationary noise. When the language is nonnative, the importance of phonetic and semantic speech cues is larger than for a native language, probably due to the listeners' imperfect knowledge in the nonnative language (Calandruccio, Buss, & Hall, 2014). However, the attempt to find such cues in the presence of the steady-state masker appeared to be less successful. The other maskers included in the experiment had larger potential to provide such speech cues to the listener.

As expected, and consistent with the study by Kilman et al. (2014), there was an interaction effect between target language and English-proficiency group. This interaction effect was based on the different effect of English proficiency on the perception of English and Swedish target languages. In English as the target language, the high-proficiency group indeed performed better than the low-proficiency group, but this difference was not expected, nor observed for Swedish as target language. The current study again shows the importance of taking into account the proficiency level when assessing the influence of target language so this holds for both NH and HI listeners. We expected that high English proficiency and good cognitive performance would be associated with better speech perception. We also hypothesized that worse hearing acuity would be related to lower speech perception performance.

As expected, English proficiency was related to SRTs in English. For Swedish target language, PTA_4 was associated with three out of four SRTs suggesting that hearing acuity is a strong factor here. This is consistent with Akeroyd (2008). It is unclear why we did not observe a relation between PTA_4 and SRTs for Swedish target language and stationary noise. A possible explanation may be that the Cambridge formula used to adapt the auditory signal to the individuals' hearing loss was sufficiently compensating for hearing problems for stationary maskers but not for the other masker types. For Swedish target language, the effect of masker type disappeared when PTA_4 was included as covariate in the ANOVA. This may indicate that the differences between the masker types were (partly) reflecting differences in PTA_4. As shown in Table 4, higher PTA_4 was associated with poorer SRTs, except for the stationary noise masker. The absent relation between PTA_4 and the babble in English may indicate that factors like language proficiency or other cognitive skills better explain the differences between the listeners. Apparently, hearing impairment is not driving those differences.

Consistent with Kilman et al. (2014), working memory (English and Swedish Reading Span), nonverbal intelligence (Raven), and English proficiency were significantly related to each other. This may indicate that they are all measuring partly different, but related cognitive factors. Higher age and lower education were associated with poorer English proficiency. It is possible that some of the older HI listeners had a poorer schooling in English in combination with limited practice, which may have resulted in this lower degree of English ability. The quality of education in English was probably lower in the mid−50 s of the previous century, and pupils only received 4 years of English education at that time.

We observed more associations with cognitive measures for the perception of English as compared with Swedish speech. From Kilman et al. (2014), we know that English proficiency is an important factor for speech perception in English. Therefore, we performed a partial correlation analysis to investigate which of the associations between the cognitive measures and SRTs would remain significant when controlling for English proficiency. The analyses showed that only a few correlations remained significant after controlling for English proficiency. This again suggests that English proficiency is an important factor when perceiving English speech and that it is partly related to other cognitive abilities as well.

Note that the associations between PTA_4 and the Swedish SRTs did not change when controlling for English proficiency. This suggests that English proficiency and PTA_4 are relatively independent, and hence, both factors are important. However, PTA_4 was also associated with the English SRTs in the fluctuating masker. The fact that Swedish Reading Span and PTA_4 were both related to the English SRTs in the fluctuating masker and remained significant after controlling for English proficiency may indicate that a combination of top-down (working memory) and bottom-up (PTA) factors is associated to speech perception in the fluctuating masker. The finding that a fluctuating masker is an estimate for auditory performance and also for nonauditory factors is in line with the findings reported by George et al. (2007). Cognitive abilities becomes more important when the background masker is fluctuating than when it is stationary, as it is cognitively more demanding to ignore a rapidly changing background than it is to ignore a steady background (Pichora-Fuller, 2009). This provides support for the ELU framework which states that in challenging conditions, working memory and cognitive abilities are particularly relevant (Rönnberg, 2003; Rönnberg et al., 2008, 2013).

In this study, there was a large variance in SRTs between the listeners. The performance was influenced by the listener-related factors such as hearing-impairment, age, cognitive abilities, and proficiency in the nonnative language. Also, external factors played a role such as the target language (native vs. nonnative) and the different masker types. The interactions between these factors affected the complexity of the listening conditions as experienced by the listeners. Each of the variables assessed in this study is relevant for nonnative speech perception in noisy conditions by individuals with hearing impairment. However, how the variables interact with each other, or, presumably, compensate or exaggerate the effects of each other was not possible to define in the present study.

The current study implies that, for a clinician, it may be complicated to predict the speech perception challenges encountered by an individual, based upon pure tone audiometry alone. For example, two listeners with hearing loss may have the same pure tone thresholds. One could be a young individual, not so highly educated and low proficient in the nonnative language, while the other could be an elderly individual, highly educated and highly proficient in the nonnative language. The experiences and the performances of these two listeners will probably differ, especially for nonnative speech perception; but on what basis and to what extent? Although the current study provides part of the answer to this question, more research is required into this topic to fully understand the impact of the various factors on speech perception.

Conclusions

- Results indicate that the speech maskers produced more interference than the noise maskers in the native as well as the nonnative languages.
- The interaction effect between language and masker type suggests a relatively larger interference of the stationary masker compared with the fluctuating masker in the nonnative language than in the native language. This result indicates that the imperfect knowledge in the nonnative language generates a larger need for phonetic and semantic cues, which is not obtainable in the stationary noise as there are no dips in the masker. Stationary maskers in a nonnative language may be relatively more disturbing for the HI listener.
- English proficiency is an important factor associated with interindividual differences in English speech perception in noise. PTA is the primary predictor for native speech perception in noise.
- A large variance in performance between the listeners was observed, especially for speech perception in the nonnative language.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The research was supported by the grant to the Linnaeus Centre HEAD, Linköping University, Sweden, from the Swedish Research Council (349-2007-8654).

References

Agus, T. R., Akeroyd, M. A., Gatehouse, S., & Warden, D. (2009). Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise. *The Journal of the Acoustical Society of America*, *126*, 1926–1940.

Akeroyd, M. A. (2008). Are individual differences in speech reception threshold related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47*(Suppl 2): 53–71.

Alexander, J. M., & Lufti, R. A. (2004). Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights. *The Journal of the Acoustical Society of America*, *116*(4), 2234–2247.

Arbogast, T. L., Mason, C. R., & Kidd, G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *117*(4 Pt 1), 2169–2180.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. (doi: 10.1146/annurev-psych-120710-100422.

Baskent, D., van Engelshoven, S., & Galvin, J. J. (2014). Susceptibility to interference by music and speech maskers in middle-aged adults. *The Journal of the Acoustical Society of America*, *135*(3), 147–153.

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. *The Journal of the Acoustical Society of America*, *131*, 1449–1459.

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

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5), 2527–2538.

Calandruccio, L., Buss, E., & Hall, J. W., III (2014). Effects of linguistic experience on the ability to benefit from temporal and spectral masker modulation. *The Journal of the Acoustical Society of America*, *135*(3), 1335–1343.

Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, *128*, 860–869.

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *ASHA*, *23*, 493–500.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7.12*, 547–552.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *9*, 561–584.

Daneman, M., & Merikel, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422–433.

Desjardin, J. L., & Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear and Hearing*, *34*, 261–272.

Dubno, J., & Ahlstrom, J. (1997). Additivity of multiple maskers of speech. In W. Jesteadt (Ed.), *Modeling sensorineural hearing loss* (pp. 253–272). Mahwah, NJ: Erlbaum.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. The Journal of the Acoustical Society of *America*, *88*(4), 1725–1736.

Foo, C., Rudner, M., Rönnberg, J., & Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *Journal of the American Academy of Audiology*, *18*, 553–566.

Freyman, R. L., Helfer, K. S., Mc Call, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, *106*, 3578–3588.

Gatehouse, S., Naylor, G., & Eberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *International Journal of Audiology*, *42*, 77–85.

George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M.Houtgast, T. (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2362–2375.

Hällgren, M., Larsby, B., & Arlinger, S. (2006). A Swedish version of the Hearing In Noise Test (HINT) for measurement of speech recognition. *International Journal of Audiology*, *45*, 227–237.

Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing*, *29*, 87–98.

Helfer, K. S., & Staub, A. (2013). Competing speech perception in older and younger adults: Behavioral and eye-movement evidence. *Ear and Hearing*, *35*, 161–170.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. L. (2007). Informational masking. In W. Yost (Ed.), *Springer handbook of auditory research 29: Auditory perception of sound sources* (pp. 143–190). New York, NY: Springer.

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

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

Micheyl, C., Arthaud, P., Reinhart, C., & Collet, L. (2000). Informational masking in normal-hearing and hearing-impaired listeners. *Acta Otolaryngologica*, *120*(2), 242–246.

Moore, B. C. J. (1985). Frequency selectivity and temporal resolution in normal and hearing-impaired listeners. *British Journal of Audiology*, *19*, 189–201.

Moore, B. C. J., & Glasberg, B. R. (1998). Use of a loudness model for hearing-aid fitting. I. Linear hearing aids. *British Journal of Audiology*, *32*(5), 317–335.

Ng, E. H. N., Rudner, M., Lunner, T., & Rönnberg, J. (2015). Noise reduction improves memory for target language speech in competing native but not foreign language speech. *Ear and Hearing*. Advance online publication. doi: 10.1097/AUD.0000000000000080.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*, 1085–1099.

Oxenham, A. J., & Kreft, H. J. (2014). Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends in Hearing*, *18*, 1–14.

Pichora-Fuller, K. M. (2009). How cognition might influence hearing aid design, fitting, and outcomes. *The Hearing Journal*, *62*, 1.

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, *18*, 43–52.

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for Raven's progressive matrices and vocabulary scales. Section 3: The standard progressive matrices*. England: Oxford Psychologists Press.

Reel, L. A., & Hicks, C. B. (2012). Selective auditory attention in adults: Effects of rhythmic structure of the competing language. *Journal of Speech Language and Hearing Research*, *55*, 89–104.

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology*, *42*, 68–76.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., . . . ,Rudner, M. (2013). The Ease of Language Understanding (ELU) model:

Theoretical, empirical and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 31doi:10.3389/fnsys.2013.00031.

Rönnberg, J., Lyxell, B., Arlinger, S., & Kinnefors, C. (1989). Visual evoked potentials: Relation to adult speechreading and cognitive function. *Journal of Speech and Hearing Research*, *32*, 725–735.

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*, 99–105.

Rudner, M., Rönnberg, J., & Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. *Journal of the American Academy of Audiology*, *22*, 1–12.

Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2012). Why middle-aged listeners have trouble hearing in everyday settings. *Current Biology*, *22*, 1417–1422.

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, *18*, 559–572.

Stone, M. A., Fullgrabe, C., & Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America*, *132*(1), 317–326.

Stone, M. A., & Moore, B. C. J. (2014). On the near non-existence of "pure" energetic masking release for speech. *The Journal of the Acoustical Society of America*, *135*(4), 1967–1977.

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

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native and foreign language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*, 519–526.

van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America*, *111*, 1906–1916.

Versfeld, N. J., & Dreschler, W. A. (2002). The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. *The Journal of the Acoustical Society of America*, *111*, 401–408.

von Hapsburg, D., & Peña, E. D. (2002). Understanding bilingualism and its impact on speech audiometry. *Journal of Speech Language and Hearing Research*, *45*, 202–213.