Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 52 (2010) 919–929
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking?

Payam Ezzatian, Meital Avivi, Bruce A. Schneider *

Department of Psychology, Centre for Research on Biological Communication Systems, University of Toronto Mississauga, Mississauga, Ontario, Canada L5L 1C6

Received 31 July 2009; received in revised form 12 February 2010; accepted 8 April 2010

Abstract

Since most everyday communication takes place in less than optimal acoustic settings, it is important to understand how such environments affect nonnative listeners. In this study we compare the speech reception abilities of native and nonnative English speakers when they are asked to repeat semantically anomalous sentences masked by steady-state noise or two other talkers in two conditions: when the target and masker appear to be colocated; and when the target and masker appear to emanate from different loci. We found that the later the age of language acquisition, the higher the threshold for speech reception under all conditions, suggesting that the ability to extract speech information from masking sounds in complex acoustic situations depends on language competency. Interestingly, however, native and nonnative listeners benefited equally from perceived spatial separation (an acoustic cue that releases speech from masking) independent of whether the speech target was masked by speech or noise, suggesting that the acoustic factors that release speech from masking are not affected by linguistic competence. In addition speech reception thresholds were correlated with vocabulary scores in all individuals, both native and nonnative. The implications of these findings for nonnative listeners in acoustically complex environments are discussed.

Keywords: Second language; Bilingualism; Speech comprehension; Speech perception; Informational masking; Energetic masking; Stream segregation

1. Introduction

Because most everyday communication takes place in less than optimal acoustic settings, it is important to understand how nonnative speech perception is influenced by the presence of interfering sound sources, and importantly, whether nonnative speakers of a language are able to gain the same benefit as natives from the bottom-up (e.g. spectral separation of competing sounds, asynchronous fluctuations in the sounds' amplitude) and top-down (e.g. familiarity with a talker's voice or a priori knowledge of target location) cues that might aid speech perception in noisy environments. Since language processing depends on the interaction of low-level acoustic–phonetic, and higher-level linguistic–semantic processing, it is possible that in acoustically cluttered environments (e.g. environments with many different sound sources) where phonetic cues are likely to be obscured or masked by other sound sources, the reduced ability of nonnative listeners to extract phonetic information from target sentences might preclude them from taking full advantage of cues that aid the perception of target speech. And, in turn, a reduced ability to segregate the target speech from the background may impede optimal linguistic–semantic processing of the incoming target. In this study we examine whether the age at which a listener acquires a language affects their ability to benefit from one of the cues that allow listeners to isolate and focus their attention on a speech target, namely perceptually separating the target speech from competing sounds.

* Corresponding author. Tel.: +1 905 828 3963; fax: +1 905 569 4850.
E-mail address: bruce.schneider@utoronto.ca (B.A. Schneider).

0167-6393/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2010.04.001

1.1. Speech reception in nonnative listeners

Nonnative speakers of a language tend to have lower scores than native speakers on a number of speech-reception measures (Bradlow and Bent, 2002; Bradlow and Pisoni, 1999; Cooke et al., 2008; Mayo et al., 1997; Meador et al., 2000; Rogers et al., 2006; Rogers and Lopez, 2008; van Wijngaarden et al., 2002). This difference in performance is influenced by several factors, such as duration of exposure to the nonnative language, degree of similarity between the native and nonnative languages, knowledge of the nonnative language vocabulary and grammatical structure, frequency and extent of nonnative language use, etc. One of the most important factors influencing nonnative language proficiency, however, seems to be the age at which a nonnative language is acquired. If the age at the onset of immersion in a nonnative language environment falls beyond a 'critical period', the acoustic–phonetic characteristics of this language may not be fully acquired (e.g. Florentine, 1985; Mayo et al., 1997), thus resulting in a reduced ability to discriminate fine phonemic information, such as phonetic contrasts and phonemic categories (Bradlow and Pisoni, 1999; Flege, 1995; Meador et al., 2000). Indeed, there is some evidence to suggest that when a language is acquired beyond the critical period, native-like performance with respect to phoneme identification and discrimination may be near impossible to achieve (Abrahamsson and Hyltenstam, 2009). Moreover, studies have shown that even highly proficient, early bilinguals, who have no noticeable accents in their second language, perform worse than natives when tested in acoustically taxing listening environments, despite performing equivalently to monolinguals in quiet or optimal environments (e.g. Rogers et al., 2006; Heinrich et al., 2008).

1.2. Stream segregation

The processing of a target signal in acoustically complex environments depends on the extent to which the auditory system succeeds in perceptually segregating the target stream from the competing streams. Therefore, any cue that allows the listener to distinguish a speech target from competing sound sources can enhance the processing of the speech signal (Bregman, 1990; Schneider et al., 2007). Several features of the acoustic environment have been shown to provide such cues that facilitate stream segregation, thereby providing a significant release from masking (e.g. Brungart et al., 2006). Some of these cues are based on the inherent differences between the spectral–temporal features of the acoustic signals, while some others are based on how the signals are affected by the acoustic characteristics of the environment, and/or on the spatial arrangement of the sound sources within that environment. The spectral–temporal features of the target speech that can enhance its perceptual segregation from competing sources include differences in onset time, voice pitch (F0), accent, formant structure, prosody and intonation, and other spectral–temporal fluctuations (e.g. Akeroyd and Summerfield, 2000; Brungart et al., 2001; Darwin et al., 2003; Hawley et al. 2004). The environmental features that can influence segregation of the target stream from background sources include factors such as the degree of reverberation in the environment and the spatial arrangement of sound sources, both of which can affect interaural time and intensity differences between targets and maskers (e.g. Kidd et al.,1998; Freyman et al., 2001; Noble and Perrett, 2002; Li et al., 2004; Litovsky, 2005; Marrone et al., 2008). Up until the point at which perceptual segregation is fully achieved, the more two sources differ with respect to their spectral–temporal characteristics when they reach the ears, the easier it will be for the auditory system to segregate them into separate streams. Therefore, in any naturalistic listening environment where the perception of target speech signals may be masked by competing sound sources, successful communication will depend on both the availability of cues that can facilitate stream segregation, and the ability of listeners to take advantage of such cues when they are available.

1.3. Release from masking depends on the type of masker

As a number of studies have shown, the amount of release from masking provided by acoustic cues such as spatial separation when the target is speech depends to a large extent on the nature of the masking stimulus (Brungart et al., 2001; Helfer and Freyman, 2005, 2008; Freyman et al., 1999, 2001, 2004, 2007; Kidd et al., 2005; Li et al., 2004). For example, when the target is speech, spatially separating the target from the masker may result in less improvement in speech understanding when the masking stimulus is a steady-state noise than when the masker consists of other talkers. In the latter case, the amount of release from masking is more variable and can depend on several factors such as the number and gender of background talkers, as well as the level of the background talkers relative to the target (Brungart and Simpson, 2002; Freyman et al., 1999, 2004; Li et al., 2004). The reason for this difference in benefit due to spatial separation most likely is due to the nature of the interference with speech reception occasioned by the two types of maskers. Depending on their spectral composition, both noise and speech maskers can and often do activate regions along the basilar membrane that are needed for the processing of the speech target. In other words, the energy in the masker interferes with the encoding of the speech signal at a peripheral level. Such interference is often referred to as peripheral or energetic masking. Simultaneously presented sound sources, including competing speech, will produce energetic masking when there is spectral overlap between the competing sound sources and the target speech. However, competing speech, in addition to potentially acting as an energetic masker, is likely to initiate linguistic and semantic processing of the target speech, thereby interfering with the linguistic and semantic processing of the target speech signal. Such interference is often referred to as informational

masking because it is believed to interfere with the processing of the speech target at more central levels in the auditory pathway (Freyman et al., 2004; Li et al., 2004; Schneider et al., 2007). Because steady-state noise maskers (e.g. ventilation noise) are unlikely to initiate semantic and/or linguistic processing, any reduction in speech intelligibility that occurs in the presence of such a masker is likely to be due primarily to energetic masking. The amount of release that acoustic cues which facilitate source segregation provide in the presence of such maskers is limited. Hence, they are often used to establish the degree of energetic masking in a listening situation. Because competing speech, in addition to acting as an energetic masker, is likely to interfere with the processing of the speech target at more central auditory levels, it may produce both energetic and informational masking. A failure to perceptually segregate the target speech from competing speech therefore should result in the confusion of background streams with targets, and/or the intrusion of irrelevant informational content into working memory during the processing of target speech. Hence, any acoustic feature that facilitates source segregation when the target and maskers are speech could release the target from both energetic and informational masking. The greater degree of release of a speech target from a speech masker than from a noise masker presumably reflects reduction in the confusability of the target and maskers, and release from the interference of the speech masker with the speech target at semantic and linguistic processing levels.

1.4. Release from masking in native and nonnative listeners

Because nonnative language acquisition follows a different course than native language acquisition, it is possible that relative to native language users, nonnatives will be differentially affected by the interference of background sources, and may not reap the same benefit as natives from cues that can provide release from masking. It is thus important to understand the factors that contribute to the speech communication ability of nonnative speakers, and whether there are differences between natives and nonnatives in the ability to use cues that could aid speech perception in challenging listening environments.

However, while nonnative speech perception in noise has received a fair amount of attention, only a few studies have examined nonnative listeners' ability to take advantage of cues that could lead to improved speech perception in noisy backgrounds. One such study conducted by Bradlow and Bent (2002), examined the ability of native undergraduate and nonnative graduate students to take advantage of clear speech in the perception of simple English sentences presented against a background of white noise. They found that the performance of their nonnative participants, who had limited exposure to the sound system of the English language, improved to a significantly lesser extent than that of their native subjects when listening to sentences read using clear speech. This suggests that nonnatives have a reduced ability to take advantage of the signal enhancements provided by clear speech (e.g. increased acoustic salience, and enhancements of phonetic contrasts, Bradlow and Bent, 2002). Another study conducted by Mayo et al. (1997) used the Speech Perception in Noise test, (Kalikow et al., 1977; Bilger et al., 1984) to examine the 50%-correct thresholds of early (learned English prior to age 6) and late bilingual (learned English after age 14), and monolingual English speaking university students for low and high context sentences administered against a background of babble. Their results showed that both bilingual groups required a higher signal-to-noise ratio to attain 50%-correct thresholds relative to monolingual English speakers, and that the early bilinguals outperformed the late bilinguals on this task. More interestingly, they also found that the late bilinguals were unable to derive the same benefit from sentence context as the monolingual and early bilingual participants, indicating a reduced ability in late bilinguals to use top-down knowledge of sentence context to predict the final keywords in target sentences. Interestingly however, in a study by Bradlow and Alexander (2007), where both contextual information and use of clear speech were manipulated independently as variables, nonnative participants' perception of sentences against speech-spectrum noise was found to improve significantly, but only when sentences contained both high contextual information and were presented using clear speech. However, the signal-to-noise ratios at which these sentences were presented were more favorable than those used by Bradlow and Bent (2002) and Mayo et al. (1997).

The results of these studies indicate that the ability of nonnatives to take advantage of some cues that could enhance speech perception might not be equivalent to that of native listeners under the same acoustic conditions. The reasons why nonnative listeners seem to have reduced ability to extract the acoustic features of speech in the presence of competing sound sources needs to be further investigated. Previous studies have provided evidence of a significant release from masking due to perceived spatial separation in native listeners (e.g. Freyman et al., 1999; Li et al., 2004; Marrone et al., 2008). However, to our best knowledge, the influence of spatial separation on speech perception in nonnative listeners has never been tested.

There are at least two reasons why we might expect the degree of release from a speech masker due to spatial separation to differ between native and nonnative speakers of a language. First, the bottom-up processes involved in spatial release from masking appear to be affected by language experience. The general finding that nonnatives require higher signal-to-noise ratios than natives, or Bradlow and Bent's (2002) finding that nonnatives do not derive the same benefit from clear speech that natives do, suggests that the effectiveness of bottom-up cues is influenced by the degree of exposure to the language presented. The extent to which spatial separation releases a listener from a speech masker could depend on the effectiveness of these bottom-up cues, thereby resulting in different degrees of

release in native and nonnative listeners. Second, listening in a second language, especially when there is competing speech, presumably requires the marshalling of more cognitive resources and a greater degree of effort on the part of the listener than listening in one's own native language. Competing speech is likely to be more difficult to ignore, shared phonemes in the native and nonnative languages might activate more than one lexicon, and the presence of multiple language streams may place a greater demand on working memory. Although effective stream segregation due to spatial separation would reduce the processing load for both native and nonnative listeners, it is not clear that the degree of reduction would be the same for both. For these reasons, it would be interesting to know if the amount of release from energetic and informational masking differed between native and nonnative listeners.

1.5. Using perceived spatial separation to evaluate the degree of release from speech on speech masking

In this study, perceived rather than real spatial separation was used to investigate whether nonnatives reap the same benefit as natives from an acoustic cue which is known to facilitate stream segregation. The advantage of using perceived rather than real spatial separation is that the former allows us to control for the monaural changes in signal-to-noise ratio (SNR) that result when target and masker are spatially separated. When a speech target and a competing sound source are located to the left and right of the listener respectively, the sound shadow cast by the listener's head reduces the intensity of the masker in the ear which is contra lateral to the masker's origin, which results in an improved SNR in that ear. Hence, changing the spatial separation between target and masker will alter the SNR at each ear. To keep the monaural SNRs the same under spatially separated and colocated conditions we used the precedence effect to change the apparent location of the masker. A number of studies have shown that if the same sound is played over two loudspeakers located to the left and right of the listener, with the sound on the left loudspeaker lagging that on the right, the listener perceives the sound as emanating from the right. If, however, the time delay is reversed, the perceived location of the sound source is reversed. Because the sound is played over both loudspeakers, such a reversal has a negligible effect on the SNR at each ear.

In this study we used the precedence effect to evaluate the degree to which native and nonnative listeners benefit from perceived spatial separation of a speech signal from either a noise masker or a speech masker, and the extent to which such benefit is related to measures of language competence.

2. Method

2.1. Participants

A total of 64 university students (18–28 years old), recruited from the University of Toronto Mississauga, participated in this study and were paid for their participation. All participants had normal and balanced audiometric thresholds which were less than or equal to 20 dB HL in both ears from 250–3000 Hz (with the exception of one participant whose threshold at 250 Hz was 25 dB HL) and did not exceed 30 dB HL at 6000 and 8000 Hz.

The participants were volunteers from four different types of linguistic background. The first group consisted of 16 native English speakers who were born and raised in an English-speaking country. The second group consisted of 16 participants who arrived in Canada from a non-English speaking country between the ages 7–14 years old, at which point they were educated in English and judged themselves to be fluent in English. The third group consisted of 16 participants who were raised in a non-English speaking country for at least 15 years before emigrating to an English speaking country and were not educated in English prior to their arrival, although all had received some level of instruction in English as a foreign language in their native country. The fourth group consisted of 16 participants who were educated primarily in English but in a country where the first language is not English, and arrived in Canada when they were older than 10 years of age (except one participant who was born in a Polish community in Canada and was exposed to English only when she attended an English–French school at the age of 6 years old). The nonnative participants in this study immigrated to Canada from the following 22 countries: Brazil ($n = 1$), China ($n = 8$), Egypt ($n = 1$), Germany ($n = 4$), India ($n = 6$), Indonesia ($n = 2$), Israel ($n = 1$), Japan ($n = 1$), Korea ($n = 2$), Kuwait ($n = 1$), Lithuania ($n = 1$), Malaysia ($n = 1$), The Netherlands ($n = 2$), Nigeria ($n = 1$), Pakistan ($n = 5$), Russia ($n = 1$), Saudi Arabia ($n = 1$), Singapore ($n = 2$), Sri-Lanka ($n = 2$), Taiwan ($n = 1$), Ukraine ($n = 1$), and the United Arab Emirates ($n = 2$).

All participants completed the Mill Hill vocabulary test and the Nelson–Denny reading comprehension test (except one native participant who did not complete the Mill Hill). The average scores achieved by each of the participant groups are provided in Table 1.

2.2. Materials and apparatus

During test sessions, the listener was seated in a chair located in the center of an Industrial Acoustic Company (IAC) sound-attenuated chamber, whose internal dimensions were 283 cm in length, 274 cm in width, and 197 cm in height. Two loudspeakers were placed symmetrically in the frontal azimuthal plane at 45° angles to the left and right of the listener. The distance between the listener's head and each one of the speakers was 169 cm. The height of the loudspeakers was adjusted to match the ear level of a seated listener of average body height.

All the acoustic stimuli used for the current study were digitized at 20 kHz sampling rate using a 16-bit Tucker Davis Technologies (TDT, Gainesville, FL) System II

Table 1
Average Mill Hill and Nelson–Denny scores obtained for each group of participants. Mill Hill scores are the number of words recognised correctly out of 20. Nelson–Denny scores are the number of correct answers out of 36 multiple-choice questions. Standard deviations shown in brackets.

Group	Mill Hill scores	Nelson-Denny scores
Native	14.34	26.63
	(1.72)	(4.75)
Mixed	13.44	25
	(2.25)	(5.15)
7–14	11.38	21.5
	(3.12)	(7.26)
⩾15	10.88	20.5
	(2.58)	(7.05)

and custom software. The digital signals were converted to analog forms using Tucker-Davis Technologies (TDT) DD1 digital-to-analog converters under the control of a Dell computer with a Pentium II processor. The analog outputs were low-passed at 10 kHz with TDT FT5 filters, attenuated by two programmable attenuators (TDT PA4, for the left and right channels), and fed into a headphone buffer (TDT HB5). The output from the headphone buffers was amplified via a Harmon/Kardon power amplifier (HK3370) and delivered from the two balanced loudspeakers (Electro-Medical Instrument, 40 watts) placed in the chamber.

Target sentences consisted of 208 nonsense sentences spoken by a female talker, which were developed by Helfer (1997) and previously used in experiments by Freyman et al. (1999) and Li et al. (2004). These nonsense sentences, which are grammatically correct but semantically anomalous, each contain 3 target words in sentence frames such as "A *spider* will *drain* a *fork*", or "A *shop* can *frame* a *dog*" (target word italicized). The sentences were divided into 16 lists containing 13 sentences each. Four additional sentences were added to the beginning of each list in a random order as practice sentences. These practice sentences were excluded when marking the results. All target sentences were presented over both the right and left loudspeakers. The perceived location of the target and masking sounds was manipulated using the precedence effect (Freyman et al., 1999; Li et al., 2004). The target sentences were always perceived as emanating from the right, whereas the masker appeared to originate from the right on some blocks of trials (masker and target spatially co-located) and on the left during other blocks of trials (masker and target spatially separated).

The perceived spatial locations of the target and maskers were manipulated by introducing a delay between the signals played over both loudspeakers. When the target played over the left loudspeaker lagged the same signal played over the right loudspeaker by 3 ms, it led to the perception that the target was emanating from a source to the right of the listener. When the masker played over the left loudspeaker also lagged the same signal played over the right loudspeaker by the same amount, it was also perceived to be emanating from a source to the right. However, when the time delay was reversed the masker was perceived to be located on the left. The advantage of using perceived spatial separation rather than real spatial separation is that the SNR at each ear remains approximately the same independent of the perceived locations of the signals (see Li et al., 2004, for a more complete discussion).

Target sentences were presented with either one of two types of masking stimuli: noise or speech. The noise masker was a steady-state speech-spectrum noise recorded from an audiometer (Interacoustic [Assens, Denmark] model AC5). The speech masker was a 315 second long track created using an additional set of nonsense sentences uttered by two female talkers, and repeated in a continuous loop. The target sentences were presented at a level such that each loudspeaker, when measured separately using a Brüel & Kjær (Copenhagen, Denmark) sound-level meter (Type 1616), would produce an average sound pressure of 60 dBA at the estimated center of a listener's head. While the target's sound pressure level remained constant throughout the experiment, the sound pressure level of the masker was adjusted in order to produce six SNRs: −12,−8, −6, −4, 0, 6 dB. Each participant was tested using four ratios: native participants were tested using −12, −8, −4 and 0 dB SNR, while participants from the three nonnative groups were tested using −12, −6, 0, and 6 dB SNR. The ratios for native participants were chosen based on Li et al. (2004) who used the same stimuli with native participants and found this range to be effective in bracketing speech reception thresholds under both noise and speech masking. Previous studies with nonnative listeners (e.g. Mayo et al., 1997; Rogers et al., 2006; Heinrich et al., 2008) have found these listeners to require more favorable SNRs to achieve the same performance as native listeners. However, evidence from these and other studies suggests that the speech reception thresholds of nonnatives can vary as a function of English proficiency and number of years immersed in an English-speaking environment. We therefore employed a wider range of SNRs to allow for any significant variations in speech reception thresholds in the nonnative group.

2.3. Procedure

Each of the 16 target lists was presented at a constant SNR. Sentence lists and SNRs were counterbalanced across participants such that each list was presented at each of the 4 different SNRs an equal number of times in each group. Additionally, each sentence list was presented in each of the Separation (no perceived separation, perceived separation) and Masker (speech masker, noise masker) combinations an equal number of times. In each group, eight participants were first tested with spatial separation for the first eight lists, and without spatial separation for the remaining eight. The other eight participants were tested in a reversed order. The order in which maskers were presented was counterbalanced by having four of the eight

participants in each Separation condition tested with the speech masker first, and the noise masker second. The remaining four participants were tested in a reversed order.

Prior to the experiment, participants were required to complete a detailed questionaire regarding their linguistic background, education, occupation, and health. Each participant's audiometric hearing thresholds were measured and their English vocabulary and reading comprehension were assessed using the Mill Hill and the Nelson–Denny tests, respectively. Before beginning the first experimental session, an explanation, followed by one of the practice sentences at the easiest SNR, was given to familiarize the participant with the task. Participants were asked to repeat back the target nonsense sentence after each presentation, and were scored online for any keyword which was repeated correctly. After the participant had responded, the researcher initiated the presentation of the next trial. Each trial started with the masker sound (two-talker speech or speech spectrum noise) which was followed 1 s later by a target sentence. The masker was gated off with the target sentence. After completing eight lists, a short break was offered to participants. The total duration of each experimental session was approximately 30–45 min.

To test for any learning effects, the entire experiment was repeated for each participant following a delay of at least one week but not exceeding a month. Because an ANOVA failed to find evidence of any interaction containing repetition and group, the results were collapsed across repetitions in the data analysis reported here.

3. Results

Fig. 1 plots the percentage of correctly identified keywords, averaged over the sixteen participants in each group, as a function of the ratio of the sound pressure in the target sentence to that of the masker when the masker was speech spectrum noise (left panels) or two-talker speech (right panels) under conditions of perceived spatial separation (solid circles) or colocation of speech target and masker (unfilled circles). The smooth curves drawn through the data points are logistic functions of the form

$$y = \frac{1}{1 + e^{-\sigma(x-\mu)}},$$

which were fit by minimizing Chi Square (for a description of the fitting procedure see Yang et al., 2007). Note that the parameter μ denotes the 50% point on the psychometric function (the threshold), and that σ controls the slope of the function (higher values of σ correspond to steeper slopes). An examination of this figure suggests that the later the time of acquistion of English (native, 7–14 yrs, ⩾15 yrs), the higher the threshold estimate for both kinds of maskers and perceived locations (spatially separated or colocated). The participants in the mixed group received their first exposure to English in a non-English speaking country, some at a very early age. Nevertheless, their thresholds were not as good as those of the native listeners,

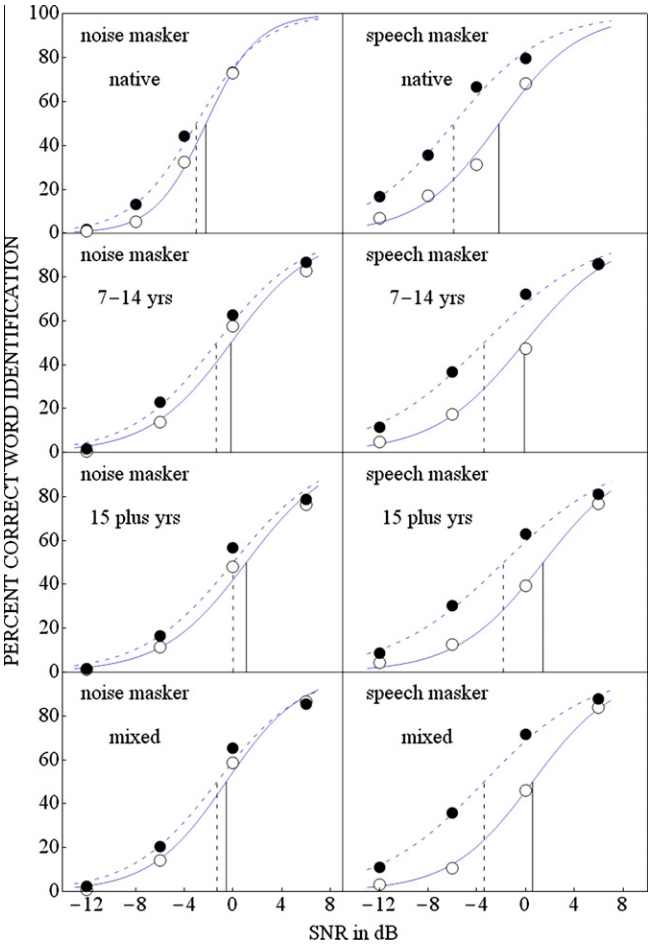

Fig. 1. Average percent correct word identification as a function of SNR in dB for four groups (native listeners: listeners who were raised in a country where English was the dominant language; 7–14 yrs: listeners who acquired English between 7 and 14 years of age; 15 plus yrs: those who were 15 years or older when they learned to speak English; mixed: those who were raised in a non-English environment but were exposed to spoken English at an early age). The open circles represent the results for the condition in which the target and masker were perceived as emanating from the same location. Solid circles represent the results for the condition in which the target and masker were perceived as emanating from different locations. Logistic functions were fit to the data from colocated (solid lines) and spatially-separate (dashed lines) conditions. Thresholds (SNRs corresponding to 50% correct on the psychometric functions) are indicated by solid vertical lines for the colocated condition, and dashed vertical lines for the spatially separate condition.

most likely because English was not generally spoken in the country in which they were raised.

This figure also indicates that the difference in thresholds between spatially separate perceived locations versus colocation (the amount of release from masking due to spatial separation) is larger when the masker is speech rather than noise. Note, however, that the amount of release appears to be the same for all four groups. There is also some indication that the slopes of the psychometric functions are steeper when the masker is noise, and somewhat shallower when the target and masker are spatially separated than when they are colocated. In addition, it appears that the

slopes of the functions when the masker is speech do not differ across the four language groups. However, there is some indication that when the masker is noise, the psychometric functions are steeper for the native listeners than for any of the other groups.

To confirm that this pattern of results also characterized the individual participants in each group, psychometric functions were fit to all individuals to obtain individual estimates of the threshold, μ, and the slope, σ. These estimates were entered into a 2 Separation (perceived separation, no perceived separation) by 2 Masker (noise masker, speech masker) by 4 Group (Native, 7–14, 15 plus, mixed) analysis of variance (ANOVA) with Group as a between-subjects factor, and Separation and Masker as within-subject factors. The ANOVA for thresholds revealed a significant main effect of Group on thresholds; $F[3, 60] = 16.24$, $p < 0.001$. Bonferroni-corrected pairwise comparisons revealed that, on average, native English speakers had significantly lower (better) thresholds than all three nonnative groups (2.03 dB lower than the 7–14 group, $p = 0.001$; 2.13 dB lower than the mixed group, $p = 0.001$; and 3.51 dB lower than the 15 plus group, $p < 0.001$). This finding replicates previous results that native language users outperform nonnatives in noisy conditions. Among the nonnative groups, the 7–14 group and the mixed group had equivalent average thresholds (mean difference = 0.099, $p > 0.99$), however both the 7–14 and mixed groups outperformed the 15 plus group (7–14 group by 1.48 dB, $p = 0.03$; mixed group by 1.38 dB, $p = 0.051$).

The main effect of Separation was also significant; $F[1, 60] = 284.71$, $p < 0.001$, as was the main effect of Masker; $F[1, 60] = 47.74$, $p < 0.001$. There was, however, a significant interaction of Separation by Masker; $F[1, 60] = 107.53$, $p < 0.001$. This interaction reflected the fact that the amount of release from masking was significantly larger when sentences were masked by the speech masker (3.57 dB) than when they were masked by the noise masker (0.98 dB). The interactions of Separation by Group and Masker by Group were not statistically significant ($F[3, 60] = 0.137$, $p = 0.937$; and $F[3, 60] = 2.078$, $p = 0.113$, respectively), indicating that the average improvement in performance in the presence of perceived spatial separation, and the average difference in performance as a function of the two different maskers was equivalent among all participant groups regardless of language background. More importantly, the analysis did not reveal a significant three-way interaction of Separation by Masker by Group ($F[3, 60] = 1.486$, $p = 0.227$), indicating that all four groups, regardless of language background, were equivalent with respect to the benefit they gained from perceived spatial separation of target and masker under both noise and speech masking conditions.

The slopes, σ, of the individual psychometric functions were also analyzed using a 2 Separation by 2 Masker by 4 Group ANOVA. This analysis revealed a significant main effect of Group on slopes ($F[3, 60] = 11.648$, $p < 0.001$), as well as significant main effects of Separation ($F[1, 60] = 50.40$, $p < 0.001$) and Masker ($F[1, 60] = 52.89$, $p < 0.001$). The interpretation of these main effects, however, is qualified by a significant three-way interaction between Separation, Masker, and Group ($F[3, 60] = 3.074$, $p = 0.034$). Fig. 1 suggests that the slopes of the psychometric functions do not vary across language groups when the masker is speech, but do when the masker is noise. To determine if this difference in the pattern of slopes for noise and speech maskers was responsible for the three-way interaction, we analysed the slopes for the speech and noise maskers separately. The results of these analyses confirmed the trends observed in Fig. 1; for the speech masker condition the main effect of Group on slopes was not statistically significant; $F[3, 60] = 1.723$, $p = 0.172$. Thus all groups had statistically equivalent slopes in the speech masking condition independently of language background. There was a significant main effect of Separation on slopes; $F[1, 60] = 27.415$, $p < 0.001$. On average, slopes decreased by a value of 0.06 when the target and speech masker were perceptually separated from each other, a result that has been reported previously (Li et al., 2004). However, the interaction of Group and Separation was not significant ($F[3, 60] = 0.730$, $p = 0.538$), indicating that the decrease in slope that occurred when targets and maskers were perceived to be spatially separate versus colocated was the same for all participants.

For the noise masker, the main effect of Group on slopes was statistically significant; $F[3, 60] = 20.539$, $p < 0.001$. Average psychometric slopes in noise were significantly steeper for the native group than the nonnative groups (native vs. 7–14 group, mean difference = 0.121, $p < 0.001$; native vs. mixed group, mean difference = 0.104, $p < 0.001$; and native vs. the 15 plus group, mean difference = 0.145, $p < 0.001$). The nonnative groups however did not differ from each other with respect to overall psychometric slopes. This finding replicates results obtained by van Wijngaarden et al. (2002) and Bradlow and Bent (2002), who used noise maskers, and Mayo et al. (1997) who used a babble masker and found that word recognition accuracy improved more slowly for nonnatives as SNR increased. The main effect of Separation on slopes was also found to be statistically significant; $F[1, 60] = 26.239$, $p < 0.001$, with the slopes being shallower in the spatially separated versus colocated conditions. There was also a significant interaction between Group and Separation; $F[3, 60] = 4.002$, $p = 0.012$. This interaction was analyzed using Bonferroni-corrected multiple comparisons which showed that while slopes for the native group were higher than those of the nonnative groups at both levels of Separation (i.e. no spatial separation and perceived spatial separation), the decrease in slope values from the no separation condition to the separation condition was significantly larger for the native group than the 7–14 and 15 plus groups (native vs. 7–14 group, mean difference = 0.081, $p = 0.012$; native vs. 15 plus group, mean difference = 0.076, $p = 0.031$) but not for the native vs. the mixed group (mean difference = 0.047, $p = 0.453$). There

were no statistically significant differences between the slopes of the nonnative groups.

To explore the contributions of vocabulary and reading comprehension skills to speech perception performance in this experiment, scores on the Mill Hill vocabulary test, and the Nelson–Denny reading comprehension test were analyzed using separate one-way between-subjects ANOVAs. For the Mill Hill vocabulary test, the ANOVA revealed a significant main effect of Group; $F[3, 60] = 6.882$, $p < 0.001$. To explore this result, Bonferroni multiple comparisons were conducted on the group means. These comparisons revealed that on average, the natives outperformed the 7–14 group by 2.95 correctly identified words, and the 15 plus group by 3.45 correctly identified words. The mixed group also had significantly better performance on the Mill Hill than the 15 plus group (mean difference = 2.56, $p = 0.029$). However the differences in performance between the natives and the mixed group, the mixed group and the 7–14 group, as well as the differences between the 7–14 group and the 15 plus group were not statistically significant. Thus, it seems that length of experience with the English language has a significant influence on vocabulary size. In this case, the native English speakers, and members of the mixed group who were educated in the English language from a young age seemed to have better vocabularies than the 7–14 and 15 plus groups who started their English education much later.

The analysis of Nelson–Denny scores also revealed a significant main effect of Group; $F[3, 60] = 3.422$, $p = 0.023$. As revealed by Bonferroni-corrected multiple comparisons, the native group had significantly higher scores on the Nelson–Denny than the 15 plus group (mean difference = 6.0, $p = 0.045$), however no other statistically significant differences were found between groups. Unlike the vocabulary test, the greatest difference in reading comprehension scores was found between the native and 15 plus group. It thus seems that length of experience with a language does not affect reading comprehension as much as vocabulary size, given that the 7–14 group performed as well as the native and mixed groups on this tasks. This result is not surprising since all participants in the above mentioned groups had at least received their high school education in a native English speaking country, whereas in case of the 15 plus group, some of the participants were international university students being educated in the English language for the first time.

To determine whether individual differences in vocabulary and reading comprehension skills could account for a significant portion of the variance in the speech perception tasks, Mill Hill scores and Nelson–Denny scores were Z-transformed within each group and entered into separate multiple regression analyses to predict Z-transformed 50%-correct thresholds in the noise and speech masking conditions. The results of the regression analyses revealed a significant linear relationship ($F[2, 60] = 5.980$, $p = 0.004$) between Mill Hill scores and performance in both the noise masker ($\beta = -0.338$, $p = 0.007$) and speech masker conditions ($\beta = -0.345$, $p = 0.006$). These results thus indicate that higher vocabulary scores are correlated with lower (better) thresholds in both the speech and noise masker conditions. However, z-transformed reading comprehension scores were not significantly correlated with thresholds for either masker. Given that the speech perception task in this experiment involved repeating back short nonsense sentences, it is not surprising that reading comprehension did not account for a significant amount of variance in performance.

To gain a better understanding of the contribution of individual differences in vocabulary size to performance in the noise and speech masker, a multiple regression was conducted using thresholds in the energetic and speech maskers to predict Mill Hill score. The results of this analysis indicated that adding speech masked thresholds to noise masked thresholds in the prediction equation did not result in a significant improvement in prediction ($F[1,60] = 2.98$, p = .089) whereas adding noise masked thresholds in a regression of z-transformed Mill Hill scores as a function of speech masked thesholds did ($F[1, 60] = 10.42$, $p = .002$). Hence, in Fig. 2 we show the z-transformed noise masked thresholds as a function of the z-transformed vocabulary scores. Fig. 2 indicates that thresholds improve (become smaller) as vocabulary improves (correlation coefficent, $r = -.39$). Subsequent analyses indicated that this improvement in thresholds with vocabulary was strongest in the 15 plus and 7–14 groups.

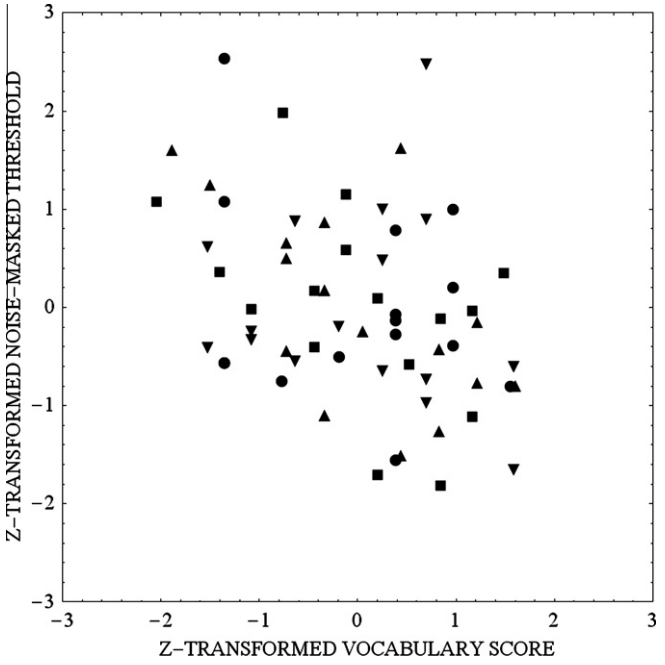

Fig. 2. Z-transformed noise-masked thresholds (colocated condition) as a function of z-transformed vocabulary scores for four different groups: native listeners (circles); listeners who first learned English between 7 and 14 years of age (squares); listeners who were 15 years old or more before learning English (triangles pointing up); and a mixed group of listeners who were raised in a non-English speaking environment but were exposed to English at an early age (triangles pointing down).

4. Discussion

4.1. The extraction of a speech signal from a noise background

As we saw in the Introduction, the extraction of a speech signal from a steady-state noise background is essentially a problem of overcoming the effects of energetic masking because it is highly unlikely that the noise masker elicits any significant activity in the semantic or linguistic processing system. Hence, one could argue that speech perception in noise assesses the listener's capacity to process speech signals that have been corrupted because of the effects of energetic masking. In cohort models of lexical access (e.g. Marslen-Wilson, 1989), it is assumed that the auditory input associated with speech activates most if not all of the cohort of words that are possible given the auditory input up until that point of time. Presumably, the corruption of the speech signal by noise could lead to a wider pattern of activation and/or slower elimination of competitors. The lower thresholds and steeper slopes in noise found for native as opposed to nonnative listeners suggest that early exposure to English in a country where English is the predominant language facilitates lexical access in the presence of steady-state noise. Specifically, equivalent increases in SNR facilitates lexical access more in native than in nonnative speakers. How might this occur? One possibility is that an increase in SNR improves performance to the same extent in native and in nonnative listeners, so that they both are able to "hear" the words better, but that equivalent increments in the ability to extract the language sounds from the noise does not translate into equivalent reductions in the pattern of activation in the lexicon. For example, a boost in the SNR that increased the number of phonemes correctly heard in a word, is more likely to reduce lexical competition in native than in nonnative listeners because the statistical properties of phoneme sequences are better represented in native than in nonnative listeners. That this native listener advantage persists even after listeners have for all intensive purposes become fluent, is consistent with the notion that there is a critical period for phonological encoding (Florentine, 1985; Mayo et al., 1997), and the acquisition of the statistical properties of a language (Saffran et al., 1996).

4.2. Speech perception in the presence of competing speech

There is reason to believe that gaining lexical access when there are speech maskers involves an additional level of complexity because the speech maskers, in addition to energetically distorting the speech signal, may also be eliciting competing activity in the lexical system. This would result in either a wider activation pattern and/or erroneous activations due to the competitors. The present results suggest that although native listeners have lower thresholds than nonnative listeners, the slopes of the functions relating percent correct word identification to SNR are the same, independent of language experience. We might expect this to occur if the ability of listeners to inhibit the processing of irrelevant alternatives or to remove irrelevant material from working memory was independent of language experience (Hasher and Zacks, 1988). A corollary of this interpretation is that any factor that would enhance stream segregation (thereby reducing activity in the semantic and linguistic systems due to speech maskers) would work equally well for native and nonnative listeners, assuming that the factor does not rely on semantic/linguistic knowledge to facilitate stream segregation. Indeed, this is what we found in this study, namely, the degree of release from a masker was the same for all groups for both maskers. Hence the degree of release from masking due to spatial separation appears to have more to do with the listener's ability to segregate the speech target from the masking background than her or his fluency in the language.

It is interesting to note that this result is consistent with that of Cooke et al. (2008) who found that native and nonnative listeners benefited to the same extent from differences in fundamental frequency between the target and competing talker. Their result also suggests that acoustic cues that are likely to facilitate stream segregation provide equivalent benefits to native and nonnative listeners. Hence, the results from this study and the present one suggest that the degree of release from masking due to acoustic level cues occurs primarily on a perceptual rather than a semantic or linguistic level as suggested by Cooke et al.

4.3. The effects of language experience on release from an informational masker

The results of this study are consistent with previous findings (e.g. Freyman et al., 1999; Freyman et al., 2004, and Li et al., 2004) that show that a) perceived spatial separation is an effective cue in providing release from masking, and b) that the release from a two-talker speech masker is significantly larger than for a continuous noise masker. If the interference caused by the two-talker masker was primarily energetic in nature, then the improvement in thresholds as a result of its perceived separation from the target location would have to be similar to the improvement observed when the noise masker was perceptually separated from target sentences. However since the release from masking due to spatial separation under the speech masking condition was significantly greater than the release in the noise masking condition, it is reasonable to assume that the major source of interference from speech masking was non-energetic, and hence informational in nature. Because the amount of release did not differ with respect to the age at which language was acquired, it is also reasonable to attribute the improvement due to spatial separation to improved stream segregation operating independent of processes involved in word identification. Hence, when only energetic masking is involved, natives benefit more than nonnatives from an increase in SNR because they can make better use of the increased audibility of the

acoustical features of the speech sound to reduce lexical competition.

4.4. The relationship between speech reception thresholds and measures of linguistic competence

In the current study we also collected two measures of linguistic competence, the Mill Hill vocabulary score, and the Nelson–Denny measure of reading ability. Not unexpectedly, both measures varied as a function of time of acquisition with native listeners outperforming those who acquired language at a later age on the vocabulary measure, and the 15 plus group on the reading measure. Because of these group differences in both language competence measures in noise and speech masked thresholds, we first z-transformed the scores within a group before looking to see whether individual differences in the language competence measures could account for variations in speech-reception thresholds. We found that speech reception thresholds improved with increases in vocabulary but not with increases in reading ability. Presumably, the significant correlation with vocabulary reflects the extent and degree of elaborateness of the individual's lexicon. We might expect lexical access to be faster and more accurate in those individual with a larger and better articulated lexicon than in those whose lexicons are not as fully developed. This is exactly what we found.

The fact that we did not observe such a correlation with the Nelson–Denny measure of reading competence is not totally surprising. Recently, Schneider et al. (2010) have argued that higher order cognitive and linguistic resources are only engaged when the task demands require them. Since the task only involved the repetition of grammatically correct but linguistically anomalous sentences, there was no need to engage processes higher than those involved in lexical access. The use of more meaningful material, and/or a change in task demands (e.g. comprehending and remembering connected discourse) would, however, require the engagement of these higher-order processes. In such cases, we would expect correlations with measures such as Nelson–Denny, which presumably assess competencies beyond those involved with lexical access. We might also expect to find differences between native and nonnative listeners with respect to their abilities to benefit from contextual effects. Indeed Mayo et al. (1997) found that context provided a greater benefit to native listeners and early bilinguals than late bilinguals when grammatically correct and semantically meaningful sentences were employed in a background of babble. At the very least, we might expect nonnative listeners to be slower at assembling and executing the higher-level cognitive processes involved in language comprehension than native listeners (Verhaeghen and Salthouse, 1997).

It would also be interesting to determine whether the effects of top-down cues that have been shown to improve stream segregation would differentially affect native and nonnative listeners. The engagement of top-down processes in word recognition and/or language comprehension will increase the processing demands in all listeners. If any of the required processes are language specific, we would expect to find differences between native and nonnative listeners.

Acknowledgments

This work was supported by the Canadian Institute of Health Research (MT15359) and Natural Sciences and Engineering Research Council of Canada (RGPIN 9952). We would like to thank James Qi for creating the program used to run our experiments and Lulu Li for help in recruiting participants.

References

Abrahamsson, N., Hyltenstam, K., 2009. Age of onset and nativelikeness in a second language: listener perception versus linguistic scrutiny. Language Learn. 59, 249–306.

Akeroyd, M.A., Summerfield, A.Q., 2000. Integration of monaural and binaural evidence of vowel formants. J. Acoust. Soc. Amer. 107, 3394–3406.

Bilger, R.C., Neutzer, J.M., Rabinowiz, W.M., Rzeczkowski, C., 1984. Standardization of a test of speech perception in noise. J. Speech Hear. Res. 27, 32–38.

Bradlow, A.R., Alexander, J.A., 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. J. Acoust. Soc. Amer. 121, 2339–2349.

Bradlow, A.R., Bent, T., 2002. The clear speech effect for non-native listeners. J. Acoust. Soc. Amer. 112, 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. J. Acoust. Soc. Amer. 106, 2074–2085.

Bregman, A.S., 1990. Auditory Scene Analysis: The Perceptual Organization of Sounds. The MIT Press, London, England.

Brungart, D.S., Chang, P.S., Simpson, B.D., Wang, D.L., 2006. Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. J. Acoust. Soc. Amer. 120, 4007–4018.

Brungart, D.S., Simpson, B.D., 2002. The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. J. Acoust. Soc. Amer. 112, 664–676.

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. J. Acoust. Soc. Amer. 110, 2527–2538.

Cooke, M., Lecumberri, M.L.G., Barker, J., 2008. The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. J. Acoust. Soc. Amer. 123, 414–427.

Darwin, C.J., Brungart, D.S., Simpson, B.D., 2003. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. J. Acoust. Soc. Amer. 114, 2913–2922.

Flege, J.E., 1995. Second language speech learning: theory, findings, and problems. In: Strange, W. (Ed.), Speech perception and linguistic experience. York Press, Timonium, MD, pp. 233–277.

Florentine, M., 1985. Speech perception in noise by fluent, non-native listeners. Proceedings of Acoust. Soc. Japan. H-85-16.

Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2001. Spatial release from informational masking in speech recognition. J. Acoust. Soc. Amer. 109, 2112–2122.

Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2004. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. J. Acoust. Soc. Amer. 115, 2246–2256.

Freyman, R.L., Helfer, K.S., Balakrishnan, U., 2007. Variability and uncertainty in masking by competing speech. J. Acoust. Soc. Amer. 121, 1040–1046.

Freyman, R.L., Helfer, K.S., McCall, D.D., Clifton, R.K., 1999. The role of perceived spatial separation in the unmasking of speech. J. Acoust. Soc. Amer. 106, 3578–3588.

Hasher, L., Zacks, R.T., 1988. Working memory, comprehension, and aging: a review and a new view. In: Bower, G.H (Ed.), The psychology of learning and motivation: advances in research and theory, Vol. 22. Academic Press, San Diego, CA, pp. 193–225.

Hawley, M.L., Litovsky, R.Y., Culling, J.F., 2004. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. J. Acoust. Soc. Amer. 115, 833–843.

Heinrich, A., Schneider, B.A., Craik, F.I.M., 2008. Investigating the influence of continuous babble on auditory short-term memory performance. Quarterly Journal of Experimental Psychology 61, 735–751.

Helfer, K.S., 1997. Auditory and auditory-visual perception of clear and conversational speech. Journal of Speech Language and Hearing Research 40, 432–443.

Helfer, K.S., Freyman, R.L., 2005. The role of visual speech cues in reducing energetic and informational masking. J. Acoust. Soc. Amer. 117, 842–849.

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

Kalikow, D.N., Stevens, K.N., Elliott, L.L., 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J. Acoust. Soc. Amer. 61, 1337–1351.

Kidd, G., Arbogast, T.L., Mason, C.R., Gallun, F.J., 2005. The advantage of knowing where to listen. J. Acoust. Soc. Amer. 118, 3804–3815.

Kidd, G., Mason, C.R., Rohtla, T.L., Deliwala, P.S., 1998. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. J. Acoust. Soc. Amer. 104, 422–431.

Li, A., Daneman, M., Qi, J.G., Schneider, B.A., 2004. Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? J. Exp. Psychol. Human Perception Perform. 30 (6), 1077–1091.

Litovsky, R.Y., 2005. Speech intelligibility and spatial release from masking in young children. J. Acoust. Soc. Amer. 117, 3091–3099.

Marrone, N., Mason, C.R., Kidd, G., 2008. The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. J. Acoust. Soc. Amer. 124, 3064–3075.

Marslen-Wilson, W., 1989. Access and integration: projecting sound onto meaning. In: Marslen-Wilson, W. (Ed.), Lexical Representation and Process. MIT Press, Cambridge, pp. 3–24.

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.

Meador, D., Flege, J.E., Mackay, I.R.A., 2000. Factors affecting the recognition of words in a second language. Bilingualism: Lang. Cognition 3, 55–67.

Noble, W., Perrett, S., 2002. Hearing speech against spatially separate competing speech versus competing noise. Perception & Psychophysics 64, 1325–1336.

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, 465–485.

Rogers, C.L., Lopez, A.S., 2008. Perception of silent-center syllables by native and non-native english speakers. J. Acoust. Soc. Amer. 124, 1278–1293.

Saffran, J.R., Newport, E.L., Aslin, R.N., 1996. Word segmentation: the role of distributional cues. J. Mem. Lang. 35 (4), 606–621.

Schneider, B.A., Daneman, M., Pichora-Fuller, M.K., 2010. The Effects of Senescent Changes in Audition and Cognition on Spoken Language Comprehension. In: Gordon-Salant, S., Frisina, R.D., Popper, A.N., Fay, R.R. (Eds.), Springer Handbook of Auditory Research: The Aging Auditory System: Perceptual Characterization and Neural Bases of Presbycusis. Springer, New York, pp. 167–210.

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.

van Wijngaarden, S.J., Steeneken, H.J.M., Houtgast, T., 2002. Quantifying the intelligibility of speech in noise for non-native listeners. J. Acoust. Soc. Amer. 111, 1906–1916.

Verhaeghen, P., Salthouse, T.A., 1997. Meta-analyses of age-cognition relations in adulthood: estimates of linear and nonlinear age effects and structural models. Psychological Bull. 122 (3), 231–249.

Yang, Z., Chen, J., Huang, Q., Wu, X., Wu, Y., Schneider, B.A., Li, L., 2007. The effect of voice cuing on releasing Chinese speech from informational masking. Speech Comm. 49, 892–904.