



## Research Paper

# Does the degree of linguistic experience (native versus nonnative) modulate the degree to which listeners can benefit from a delay between the onset of the maskers and the onset of the target speech?



Boaz M. Ben-David <sup>a, b, c, d, \*</sup>, Meital Avivi-Reich <sup>a, e</sup>, Bruce A. Schneider <sup>e</sup>

<sup>a</sup> Communication, Aging and Neuropsychology Laboratory (CANlab), Baruch Ivcher School of Psychology, Interdisciplinary Center (IDC) Herzliya, PO Box 167, Herzliya, 4610101, Israel

<sup>b</sup> Oral Dynamics Lab, Department of Speech-Language Pathology, University of Toronto, 160-500 University Ave., Toronto, Ontario, M5G 1V7, Canada

<sup>c</sup> Toronto Rehabilitation Institute, 550 University Ave., Toronto, Ontario, M5G 2A2, Canada

<sup>d</sup> Rehabilitation Sciences Institute (RSI), University of Toronto, 160-500 University Ave., Toronto, Ontario, M5G 1V7, Canada

<sup>e</sup> Human Communication Laboratory, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road N., Mississauga, Ontario, L5L 1C6, Canada

## ARTICLE INFO

## Article history:

Received 24 June 2015

Received in revised form

24 July 2016

Accepted 25 July 2016

Available online 2 August 2016

## Keywords:

Nonnative speakers

Aging

Spoken word recognition

Babble

Noise

Stream segregation

## ABSTRACT

Background noise has a greater adverse effect on word recognition when people are listening in their second language (L2) as opposed to their first language (L1). The present study investigates the extent to which linguistic experience affects the ability of L2 listeners to benefit from a delay between the onset of a masker and the onset of a word. In a previous study (Ben-David, Tse & Schneider, 2012), word recognition thresholds for young L1s were found to improve with the increase in the delay between the onset of a masker (either a stationary noise or a babble of voices) and the onset of a word. The investigators interpreted this result as reflecting the ability of L1 listeners to rapidly segregate the target words from a masker. Given stream segregation depends, in part, on top-down knowledge-driven processes, we might expect stream segregation to be more “sluggish” for L2 listeners than for L1 listeners, especially when the masker consists of a babble of L2 voices. In the present study, we compared the ability of native English speakers to those who had either recent or long-term immersion in English as L2, to benefit from a delay between masker onset and word onset for English words. Results show that thresholds were higher for the two L2s groups than for the L1s. However, the rate at which word recognition improved with word-onset delay was unaffected by linguistic status, both when words were presented in noise, and in babble. Hence, for young listeners, stream segregation appears to be independent of linguistic status, suggesting that bottom-up sensory mechanisms play a large role in stream segregation in this paradigm. The implications of a failure of older L1 listeners (in Ben-David et al.) to benefit from a word-onset delay when the masker is a babble of voices are discussed.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Daily communication often occurs in noisy environments (e.g., classrooms, restaurants, stores, offices), where competing sound sources could interfere with one's ability to communicate effectively. The presence of competing sound sources is especially challenging to those operating in their second language (L2,

Bradlow and Bent, 2002). As a result, everyday noisy situations present more of a barrier to communication and social interaction for this group than they do for native speakers of a language (L1).

One possible reason for the difficulties L2 listeners experience in the noisy backgrounds characteristic of everyday life, is that it may be more difficult for them than for L1 listeners to segregate the speech stream from the acoustic background. To recognize and comprehend speech in noise, listeners have to be able to parse the auditory scene into its component sound sources (stream segregation, Bregman, 1990), so that they can focus their attention on the target speech. Successful segregation of the acoustic input into separate auditory streams will improve speech perception in the

\* Corresponding author. Communication, Aging and Neuropsychology Laboratory (CANlab), Baruch Ivcher School of Psychology, Interdisciplinary Center (IDC) Herzliya, PO Box 167, Herzliya, 4610101, Israel.

E-mail address: [boaz.ben.david@idc.ac.il](mailto:boaz.ben.david@idc.ac.il) (B.M. Ben-David).

presence of competing sound sources, leading to a reduction in the interference caused by the maskers, commonly referred to as “release from masking” (e.g., Brungart et al., 2001; Durlach et al., 2003). However, stream segregation is not achieved instantly, and the time it takes for it to develop depends both on the listener and on the stimuli used (see the seminal work by Bregman, e.g., Bregman, 1990; Bregman and Campbell, 1971).

There are a number of acoustic cues that affect the degree of stream segregation that is achieved, such as the temporal proximity of successive sounds, as well as their similarity in  $F_0$  and spectrum. In general, the greater the acoustic dissimilarity between the target and the competing sounds, the easier it is to perceptually segregate the streams. Beyond these acoustic, or bottom-up, factors, there are a number of top-down assisting cues, such as attention, expectations, and prior exposure which could also affect streaming (e.g., Ragert et al., 2014; Shinn-Cunningham and Best, 2008). One possible reason for the difficulties L2 listeners experience in noisy backgrounds is that their ability to parse the auditory scene is either less efficient and/or slower in their L2 than in their L1, because the top-down processes supporting word recognition are not as well developed in their L2 as in their L1.

In a previous study (Ben-David et al., 2012), we investigated the degree to which the ability of L1 listeners to identify a single word masked by a competing sound source improved with increase in the delay between the onset of the masker and the onset of a word. In young L1 listeners, word identification was found to improve as the delay between the onset of the masker and word onset increased (for delays up to approximately 600 ms) for two types of maskers: a stationary broadband noise, and a babble of voices. However, the pattern of results differed for older L1 listeners. Although older L1 listeners were able to benefit from a delay between masker onset and word onset when the masker was stationary speech-spectrum noise, when the competitor was a babble masker, they were not able to benefit from this cue, even when the onset of the speech target followed babble onset by 1.1 s Ben-David et al. (2012) interpreted this pattern of results as indicating that the acoustic and phonetic similarity between the babble of voices and the words to-be-identified interfered with stream segregation more in older than in young adults.

### 1.1. Linguistic experience and spoken word identification

The present study investigated whether young L2s might also experience greater interference when listening to speech presented in babble than in noise. In these listeners, we would expect their L2 lexicon to be less well established than their L1 lexicon. Hence, access to the meaning of a word might be slowed in L2 compared to L1. In addition, the babble masker could initiate activity in both L1 and L2 lexicons, making it more difficult to segregate the target word from the background babble when listening to L2 words. This simultaneous activation might slow access to the target lexicon, and increase the competition between lexical candidates. We would also expect the babble to be a more effective masker the less fluent the listeners were in their L2. Hence, there are reasons to expect stream segregation to be more difficult when listening for L2 words in the presence of a babble of voices, than it would be when listening for L1 words in the same babble of voices (for a more complete discussion of these issues, see Avivi-Reich et al., 2014).

There are a number of reasons why we might expect word identification to be poorer and slower in L2 listeners than in L1 listeners, and for L2 listeners' performance to be more sensitive to the nature of the competing sounds than their L1 counterparts. Previous studies have shown that when asked to listen to L2 speech in a noisy environment, L2 listeners require substantially more

favorable signal-to-noise ratios than L1 listeners. For example, Nakamura and Gordon-Salant (2011) found that young native-Japanese listeners had significantly poorer English speech perception ability than native-English listeners, in both quiet and noise. The psychometric function for the native-Japanese listeners was shifted by 3–4 dB SNR from that of the native-English listeners. The difficulties experienced by L2 listeners may be the result of a reduced ability to make fine phonemic discriminations in L2 and/or to make use of language-specific cues (for a review, see Garcia et al., 2010). The degree of threshold elevation for L2 relative to L1 listeners has been found to be affected by several factors. These include, but are not limited to, the age at which they were exposed to L2, the duration of exposure to L2, the listener's individual vocabulary size and knowledge of the grammatical structure of L2, as well as extent of L2 use. It is important to note that, even in adulthood, L2's acoustic-phonetic characteristics may not be fully acquired (e.g. Florentine, 1985; Mayo et al., 1997). This might result in a reduced ability to discriminate fine phonemic information (Bialystok and Luk, 2012; Garcia Lecumberri et al., 2010), which is crucial for successful speech perception (Bradlow and Pisoni, 1999; Meador et al., 2000).

Failing to identify sounds as different phonemes in L2, could lead to an activation of additional lexical candidates as the word unfolds in time (Weber and Cutler, 2004). The additional lexical candidates could be either: (1) Intra-lingual lexical candidates, due to inefficient phonological processing or a phonemic confusion in L2; or (2) Inter-lingual lexical candidates, due to concurrent activation of words in both L1 and L2 (FitzPatrick and Indefrey, 2009; Spivey and Martin, 1999; Chambers and Cooke, 2009). In other words, those who are competent in more than a single language are likely to experience much greater competition because of simultaneous activation across the languages. For example, Hebrew does not distinguish between short and long vowels as English does. Thus, native-Hebrew listeners might activate words starting with/mee/as well as/mi/while listening to the English word/mint/unfolds. In addition, the initial sounds in the word/mint/may activate lexical candidates in Hebrew as well (e.g.,/mi.ta/, bed in Hebrew). In the case described, the listener will be forced to face a larger competition for the activation of the target word spoken in his/her L2, than a native-English listener will face. Hence, L2 listeners are likely to experience additional cross-language interference due to the activation of lexical processes in more than a single language (e.g., Weber and Cutler, 2004).

In summary, L2 listeners face difficulties with lexical access and competition, have smaller vocabulary size (Portocarrero et al., 2007) and poorer phonemic discrimination in L2 (Bialystok and Luk, 2012). The lesser L2 experience and the greater competition are possible sources for the more preferable signal-to-noise-ratios (SNRs) that L2 listeners require in order to successfully recognize speech in noise (Ezzatian et al., 2010). Indeed, it has been found that L2 listeners achieve lower scores than native listeners on a number of speech recognition measures (Bradlow and Bent, 2002; Bradlow and Pisoni, 1999; Cooke et al., 2008; Mayo et al., 1997; Meador et al., 2000; Rogers and Lopez, 2008; Ezzatian et al., 2010). L2 listeners also tend to be slower than L1 listeners, even at the level of identifying single words in L2 (e.g., Scarborough et al., 1984; FitzPatrick and Indefrey, 2009), as well as less confident (Schulpen et al., 2003). All of these factors are likely to increase the signal-to-noise ratio (SNR) required to recognize and comprehend words. We may conclude that, in the case of L2 listeners, it is reasonable to assume that the difficulties they experience in L2 environments are due to the fact their L2 lexical processes may not be as completely instantiated and differentiated from the lexical processes that are usually invoked when listening in their L1 (Kroll and Steward, 1994).

## 1.2. The current study

The main goal of the current study is to examine how linguistic experience might modulate the benefit from a delay between the onset of the maskers and the onset of the target speech. This will provide an estimate for the extent to which young L2 listeners use this cue to better distinguish the target speech stream from a noise or a babble masker stream. Two groups of EL2 (English as second language) listeners were tested: EL2 longterm residents of Canada, who arrived in Canada between the ages of 7 and 14; and EL2 recent resident, who arrived in Canada after age 14. We used the same paradigm and stimuli as employed in Ben-David et al. (2012). Namely, listeners were asked to repeat spoken words presented over headphones. There were five different delays (100, 225, 350, 600 and 1100 ms) between masker onset and target speech onset and two types of maskers: 12-talker babble and a speech-spectrum noise.

The performance of the EL2 adults, tested in the current study, was compared to that for young EL1s in our previous study. Note that although these young EL2s are highly likely to differ from young EL1 listeners on linguistic abilities, they are not likely to differ with respect to basic auditory abilities. If the ability to use the word-onset delay cue to segregate speech from a masker is based mainly on sensory processing, one might expect that young EL2s will be as adept at forming auditory objects for babble and noise maskers as the young EL1s. However, to the extent that the ability to segregate the target words from the competing background involves central linguistic and/or cognitive level processes, EL2 listeners may not reap the same benefit from word onset delays as EL1 listeners.

## 2. Materials and methods

### 2.1. Participants

Sixty young adults (age range 17–25 years old), undergraduates at the University of Toronto Mississauga, participated in the study and received either course credit or were paid \$10/h for their participation. All participants were asked to answer detailed questionnaires regarding their linguistic background before the experimental session began. These questionnaires included self-reports on linguistic background (e.g., age of acquisition of each language, languages spoken by parents at home, languages used at school), proficiency in each language and daily language use (e.g., language use when communicating with friends, at work, and when consuming media). Only EL2 participants who were born and raised in a language other than English, and did not attend an English or an American school before relocating to an English-speaking country were recruited. Following Ezzatian et al. (2010), participants were divided into two groups: A) 30 EL2-longterm participants who arrived in Canada from a non-English speaking country between the ages 7–14 years old ( $M = 9.2$ ,  $SD = 3.4$ ), at which point they were educated in English and judged themselves to be fluent in English; and B) 30 EL2-recent participants who were raised in a non-English speaking country for at least 15 years ( $M = 17.8$ ,  $SD = 2.1$ ) before immigrating to an English speaking country. Members of both groups had received some level of instruction in English as a foreign language in their native country. These participants came from various linguistic backgrounds: 28 Chinese (Mandarin or Cantonese), 8 Korean, 4 Spanish, 3 Farsi, 2 Russian and one person speaking each of the following languages: Sinhalese, Tagalog, Thai, Punjabi, Turkish, Hindi, Armenian, Vietnamese, Indonesian, Albanian, Italian, Ukrainian, Lithuanian, Dutch and Tamil. All of the EL1 listeners, taken from the previous study (Ben-David et al., 2012), were raised in an English speaking country,

learned English as their first and primary language, and reported that they use English as their only language for communicating with friends and at work. Twenty of the 30 young EL1s reported that they have some knowledge of another language, yet only two of them reported to be fluent speakers of a language other than English.

All participants in the current study completed the abbreviated version of the Mill Hill vocabulary test (for a review of the test, see Ben-David et al., 2015), and the vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS, Wechsler, 1997). The EL2-longterm group achieved significantly higher English vocabulary scores than the EL2-recent group, as detailed in Table 1. Table 1 also indicates that these groups did not vary in age or in the distribution of gender. Both current groups did not vary in age from the young adult EL1 group examined in our previous study ( $M = 20$  years,  $SD = 1.6$ ; Ben-David et al., 2012).

As expected, Table 1 indicates that the Mill Hill vocabulary scores are higher in young EL1 listeners compared with the two EL2 groups. A questionnaire was used to ensure all participants had good health and no history of auditory pathologies. All participants had pure-tone air-conduction thresholds within clinically normal limits from 0.25- to 8-kHz in the better ear ( $\leq 15$  dB HL, four out of the sixty participants had a threshold of 20 dB in a single frequency). Hearing levels (for the better ear; left ear for 19, and 14 of longterm and recent EL2s) are shown in Fig. 1 for both groups, alongside hearing levels for young native speakers taken from Ben-David et al. (2012). Note that thresholds differences between the three young-adult groups were small (did not exceed 2.5 dB on any of the tested frequencies) and no overall group effect was observed ( $F < 1$ ).

### 2.2. Stimuli, apparatus and procedure

Stimuli, apparatus and procedure replicated those used in Ben-David et al. (2012). In short, we used 10 linguistically equated lists of 52 bi-syllabic recorded spoken English words, matched on average word duration. Spoken words were later presented on the background of 4 s of either continuous stationary speech spectrum noise or 12-talker babble. All of the 520 digital audio files of the spoken words were equated with respect to root mean square amplitude (for the average spectra of the target word stimuli, babble and noise masker, see Fig. 6, p. 61, in Ben-David et al., 2012).

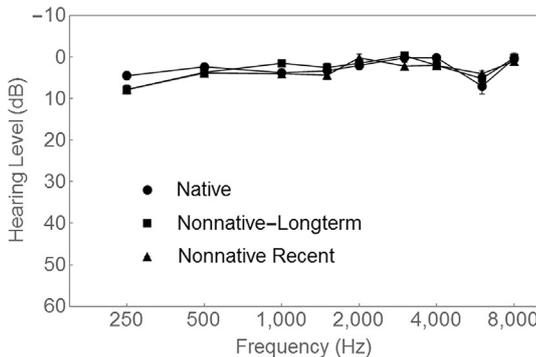
Participants were tested individually in a double-walled sound-attenuating booth. Throughout the experiment, words were presented to the listener's better ear via headphones at a level that was individually set to 50 dB above his/her 12-talker babble threshold, as measured by an adaptive two-interval, two-alternative forced-choice paradigm taken from Heinrich et al. (2008). Average babble thresholds did not vary significantly between the EL2-recent and EL2-longterm groups (averages of 16.2 and 16.5 dB SPL, respectively;  $t(58) = 0.33$ , ns).

The experimental session included ten blocks of 52 trials. The word list assigned to a block and the order of the trials within a block were randomized for each participant. In five of these blocks, words were masked by babble, and in the other five by speech spectrum noise. Noise and babble blocks were intermixed and counterbalanced across participants. We also manipulated the amount of time-delay between the onset of the masker and the onset of the word. The target word was presented 100, 225, 350, 600 or 1100 ms after the masker onset. The order of the five word-onset delays was counterbalanced across participants as well.

In each block, trials were presented in four different SNRs, 13 trials at each SNR. Two different sets of four equidistant SNRs for babble and noise maskers were chosen:  $-23$ ,  $-16$ ,  $-9$  and  $-2$  dB SNR for babble, and  $-10$ ,  $-4$ ,  $+2$  and  $+8$  dB SNR, for noise. In a pre-

**Table 1**  
Average (and sd) demographics for all four groups: Vocabulary scores on the abbreviated Mill Hill (number of words recognized correctly out of 20), Wechsler Adult Intelligence Scale (with a maximum score of 66), age in years and gender.

	Current data		Ben-David et al. (2012)		Comparing groups
	EL2-recent	EL2-longterm	EL1		
	Young	Young	Young	Older	
Mill Hill	8.4 (2.7)	11.2 (3.5)	12.8 (2.8)	15.1 (2.1)	$t(58) \geq 2.0, p \leq 0.05$ , for all 6 pairwise comparisons.
WAIS	25.2 (8.8)	36.6 (10.8)	—	—	$t(58) = 4.48, p < 0.001$
Age	20.6 (1.8)	19.8 (1.5)	20 (1.6)	72.3 (3.8)	$F(2, 87) = 1.83, p = 0.17$ , young groups.
Women	60%	66%	60%	66%	Recent - longterm, $\chi^2(1) = 2.8$ , ns. For all four, $\chi^2(3) = 0.57$ , ns



**Fig. 1.** Mean audiometric pure-tone air-conduction thresholds (for the best ear) as a function of frequency for data collected in the current study, alongside data collected in Ben-David et al. (2012), for older adults (empty symbols) and young adults (filled symbols), EL1 (circles), EL2-longterm (squares) and EL2-recent (triangles) listeners. The vertical bars depict the standard errors of the means. Note the three groups of young adults have highly similar audiograms.

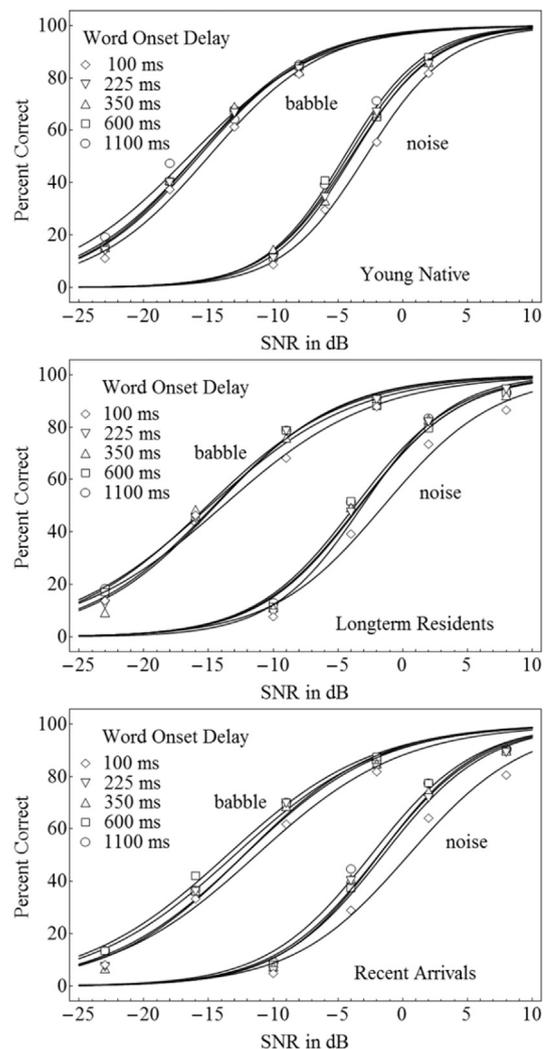
test with eight participants, representing both tested EL2 groups, these SNR ranges were found to bracket the 50% correct word-reception point (based on correct repetition). In total, participants were presented with 520 spoken word stimuli. They were asked to listen to each word and repeat it immediately, as best as they could. Three practice trials were presented at the beginning of each block of trials (using the same type of masker and word-onset delay condition as the following block, with the most favorable SNR in the respective range), but no feedback was provided in either practice or experimental trials.

2.3. Data analysis

Data analysis again fully replicated that of the previous study. Accuracy was coded by an experimenter listening to the participant's responses via headphones during the experimental session. The participant's aural responses were digitally recorded. All the trials for 53 participants (recordings for the other 7 participants were corrupted) were later recoded by two different raters. Coding-recoding agreement reached 98.3% on average, with the lowest degree of agreement for an individual's recorded responses being 93.5%. In cases of disagreement, the words were scored from the recorded utterances. The inter-rater agreement rate for the two raters of recorded responses was also checked with a sample of 5 participants (taken from both experimental groups) and reached an average of 98.9%.

These accuracy scores were used to construct psychometric functions relating accuracy to SNR for each participant at each of the five word-onset delays for the two different maskers (speech-spectrum noise and babble). In turn, these psychometric functions were used to determine the SNR corresponding to a word-identification accuracy of 50% at each of the 10 conditions (5

delays  $\times$  2 types of maskers) for each of the groups. We were able to accurately estimate 50% speech reception thresholds in all but one instance. For one of the EL1 listeners at one of the delays, performance did not exceed 50% accuracy at any of the SNRs. Hence the data from this listener was not included in this analysis (for details, see Ben-David et al., 2012). Similarly, the slopes of the psychometric functions for each condition in each group were recorded and



**Fig. 2.** Average percent-correct word identification as a function of SNR in dB for EL1 listeners (top panel) and EL2-longterm listeners (middle panel), and EL2-recent (bottom panel), for five different word-onset delays and two types of masker (noise vs. babble). The top panel are data from Ben-David et al. (2012). The psychometric functions for the different delays are associated with the different symbols identified in the figure. The top functions represent the babble masker condition and the bottom functions the noise masker condition.

analyzed. Because the slope estimate for one of the longterm EL2 listeners at one of the delays exceeded 2, the data from this listener was also removed. Hence there were 29 EL1, 29 longterm EL2, and 30 recent EL2 listeners included in the following analyses.

### 3. Results

The average accuracy for EL1 (top panel), EL2-longterm (middle panel) and EL2-recent (bottom panel) for word identification masked by babble or stationary speech spectrum noise is presented in Fig. 2 for the five word-onset delays (100, 225, 350, 600 and 1100 ms), as a function of dB SNR. As aforementioned, average word identification accuracies, for each word-onset delay, were fit by logistic psychometric functions of the form.

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

$$\text{slope} = \frac{\text{threshold at the shorter delay} - \text{threshold at the longer delay (in dB SNR)}}{\text{onset delay region(in Seconds)}} \quad (2)$$

where  $y$  is the probability of correctly identifying the target word,  $x$  represents the SNR in dB,  $\mu$  determines the dB SNR level corresponding to 50%-correct performance, and  $\sigma$  marks the slope of the fitted function at the 50% point. This figure indicates performance improves with onset delay (psychometric functions are shifted to the left as the delay increases) for all three groups of participants in each of the two masker conditions. In addition, there is a tendency for the slopes of the psychometric functions to be steeper for EL1 listeners than for EL2 listeners, and steeper for the noise masker than for the babble masker.

To investigate changes in sensitivity across groups and delays, individual psychometric functions (see Equation (1)) were fit to each participant's data at each of the five word-onset delay conditions for both masker types (5 delays  $\times$  2 masker types = 10 psychometric functions for each participant). These individual 50% word recognition thresholds, and their associated slopes, were used to determine how word-onset delay, the type of masker, and the degree of linguistic experience affected performance.

#### 3.1. Word-recognition thresholds as a function of word-onset delay

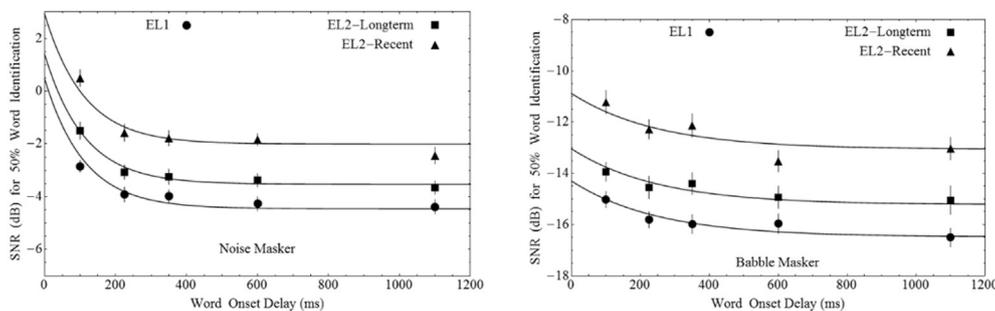
Fig. 3 plots the average 50% thresholds calculated in these

individual functions as a function of word-onset delay, for the noise background (left panel) and for the babble background (right panel). An examination of this figure suggests that performance improves with onset delay for both types of maskers, with the rate of improvement being most rapid at the shortest delays. Moreover, within a masker condition, the rate of improvement as a function of delay appears to be the same for all three participant groups. Fig. 3 also suggests that the rate of improvement is somewhat steeper when the masker is noise than when it is babble. Finally, overall, performance appears to be poorest (highest thresholds) for the recent EL2 listeners, and best for the native listeners with the longterm EL2 listeners falling between these two groups.

To confirm this description of how the rate of improvement differs with onset delay, we divided the function into four different regions: 100–225 ms delays, 225–350 ms delays, 350–600 ms delays, and 600–1100 ms delays. We then computed, for each individual, and for each noise condition, the slope over each of these onset delay regions. That is

These four slopes represent the rate of decay over the four regions (100–225, 225–350, 350–600, and 600–1100 ms) for each individual in the two types of maskers. A mixed-factorial ANOVA was then conducted on these slopes with Group (EL1s, longterm EL2s, and recent EL2s) as a between-subjects factor, and Onset Delay Region (100–225, 225–350, 350–600, and 600–1100 ms), and Masker Type (Noise versus Babble) as within-subject factors. Both the main effect of Onset Delay Region ( $F[3, 258] = 18.5, p < 0.001, \eta_p^2 = 0.177$ ) and Masker Type ( $F[1, 86] = 6.2, p = 0.015, \eta_p^2 = 0.067$ ) were significant. Neither the main effect of Group, nor any of the two or three-way interactions were statistically significant at the 0.05 level.

The average slopes associated with the four onset-delay regions were  $-9.6$  dB/s,  $-0.3$  dB/s,  $-1.6$  dB/s, and  $-0.3$  dB/s for the onset regions of 100–225 ms, 225–350 ms, 350–600 ms, and 600–1100 ms, respectively. Post-hoc pairwise t-tests for slope differences among these regions (Bonferroni corrected), indicated that the slopes over the region 100–225 ms were significantly steeper than were the slopes over the other three regions (all  $p$  values  $< 0.001$ ). None of the comparisons among the remaining three regions approached significance (all  $p$  values  $> 0.5$ ). This



**Fig. 3.** Thresholds (dB SNR values for 50% correct identification) as a function of word-onset delay taken from the psychometric functions shown in Fig. 2 for EL1 (circles), EL2-longterm (squares) and EL2-recent (triangles) listeners. The left panel presents data from the noise masker condition and the right panel presents data from the babble masker condition. The vertical bars depict the standard errors of the means. In the left panel, data points for all three groups were fit by an exponential decay functions of the type,  $y = a + be^{-cx}$ , where  $b$  and  $c$  are the same for all groups, with the asymptote,  $a$ , varying across the three groups. In Panel B, a different exponential decay function was fit to the same three groups subject to the same restrictions.

analysis supports the visual impression that there is a rapid improvement in 50% detection thresholds over the 125–250 ms region for all three groups and two masker conditions, with a smaller improvement thereafter. The main effect of Masker Type indicated that the slopes were significantly steeper for the noise masker than for the babble masker. Finally, the lack of a Group effect, or any two or three way interactions (between Group, Masker Type, and Onset-Delay Region) is consistent with the description that the rate of improvement in performance with Onset Delay is the same for all three groups, but differs over Masker Types.

Fig. 3 also shows that the asymptotic values of the functions are highest for recent EL2s, and lowest for EL1s, with the longterm EL2s falling between the other two groups. Because there was no statistically significant evidence of slope changes beyond a delay of 225 ms, we computed the average threshold of each individual in each of the two noise conditions over the three delays corresponding to 350, 600, and 1100 ms. A mixed-factorial ANOVA with Masker Type (babble vs noise) as a within-subject factor, and Group (EL1s, longterm EL2s, recent EL2s) as a between-subjects factor found a significant main effect of Masker Type ( $F[1, 86] = 5581.0$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.985$ ) and of Group ( $F[2, 86] = 32.5$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.430$ ) and a significant interaction of the two ( $F[2, 86] = 4.0$ ,  $p = 0.022$ ,  $\eta_p^2 = 0.085$ ). On average, native listeners had the lowest asymptotes, recent EL2s the highest asymptotes, with longterm EL2s in the middle. Post-hoc Sidak test revealed that recent EL2s had significantly higher asymptotic thresholds than either EL1s ( $p < 0.001$ ), or longterm EL2s ( $p < 0.001$ ), and longterm EL2s have higher thresholds than EL1s ( $p = 0.006$ ). The significant interaction effect indicated that the differences in asymptotic thresholds among the three groups were not the same for noise as they were for babble.

The exponential decay functions fit to the data in Fig. 3 incorporate the results of these two ANOVAs (see the Appendix for a description of the fitting procedure). Note, there was no statistical evidence that the rate of decline in threshold values with increasing onset delay differed among the three groups within a masker, however, the threshold values were different across maskers. Thus, the exponential functions fit to the three groups in the noise masker were restricted to having the same rate of decline for all three groups (same values of  $b$  and  $c$ , see Equation (3)), but were allowed to have different asymptotic values. Similarly, the exponential functions fit to the three groups in babble were required to have the same rate of decline for all three groups, but this rate of decline was allowed to differ from the rate of decline in a noise masker. Finally, because of the significant interaction between Group and Masker Type for asymptotic values, the asymptotes were allowed to differ among the 3 Group  $\times$  2 Masker Types. Hence, the exponential equations shown in Fig. 3 were:

$$\begin{aligned} y[t]_{EL1,noise} &= a_{EL1,noise} + b_{noise}e^{-c_{noise}t} \\ y[t]_{longtermEL2,noise} &= a_{longtermEL2,noise} + b_{noise}e^{-c_{noise}t} \\ y[t]_{recentEL2,noise} &= a_{recentEL2,noise} + b_{noise}e^{-c_{noise}t} \\ y[t]_{EL1,babble} &= a_{EL1,babble} + b_{babble}e^{-c_{babble}t} \\ y[t]_{longtermEL2,babble} &= a_{longtermEL2,babble} + b_{babble}e^{-c_{babble}t} \\ y[t]_{recentEL2,babble} &= a_{recentEL2,babble} + b_{babble}e^{-c_{babble}t} \end{aligned} \quad (3)$$

where  $t$  is onset delay, and  $y[t]$  is the 50% SNR threshold at delay  $t$ . See the Appendix for a fuller description of the data fitting procedures employed here.

Fig. 3 shows that this exponential decay model provides a good fit to the data. It illustrates that the rate of decline in thresholds with onset delay is more rapid when the masker is noise than when the masker is babble, with the decay rates in each masker condition

being the same for all three groups. Asymptotic levels are lower for babble (−13.1, −15.2, and −16.5 dB SNR for recent EL2s, longterm EL2s, and EL1s respectively) than for noise (−2.0, −3.5, and −4.5 dB SNR for recent EL2s, longterm EL2s, and EL1s respectively) with the extent of the group differences in asymptotic values being slightly larger in babble than in noise.

### 3.2. Slopes of the psychometric functions

Fig. 2 suggests that the slopes of the psychometric functions differ between the two masker types and are steeper for EL1s than for the two EL2 groups. In addition, the difference in slopes between the noise and babble conditions appears to be somewhat larger in the EL1s and longterm EL2s than in the recent EL2s. There is also some indication that there might be a small change in slope as a function of delay with the slopes being shallower at the shortest delay. To determine whether these observations were statistically significant, we conducted a mixed-factorial ANOVA with Onset Delay (100, 225, 350, 600, and 1100 ms) and Masker Type (babble vs noise) as within-subject factors, and Group (EL1, longterm EL2, recent EL2) as a between-subjects factor. The main effects of Masker Type ( $F[1, 85] = 220.4$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.722$ ) and Group ( $F[2, 85] = 23.9$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.360$ ) were highly significant. The main effect of Onset Delay was also significant ( $F[4, 340] = 3.3$ ,  $p = 0.012$ ,  $\eta_p^2 = 0.037$ ). The only significant interaction to be found was between Group and Masker Type ( $F[2, 85] = 4.44$ ,  $p = 0.015$ ,  $\eta_p^2 = 0.094$ ). Fig. 4 shows that the reason for this interaction effect is that the difference in slope between the noise and babble maskers is larger for EL1s than it is for the EL2 groups. When averaged across noise masker conditions, the slopes were significantly steeper for the EL1s than for both EL2 groups ( $p < 0.001$  for both groups, Bonferroni corrected), but the difference between the two EL2 groups was not ( $p > 0.25$ , Bonferroni corrected). Finally, across Groups and Masker Types, the slope was significantly shallower for an onset delay of 100 ms than it was for the average of the four longer onset delays ( $p < 0.005$ ).

## 4. General discussion

The current study assessed the contribution of linguistic status

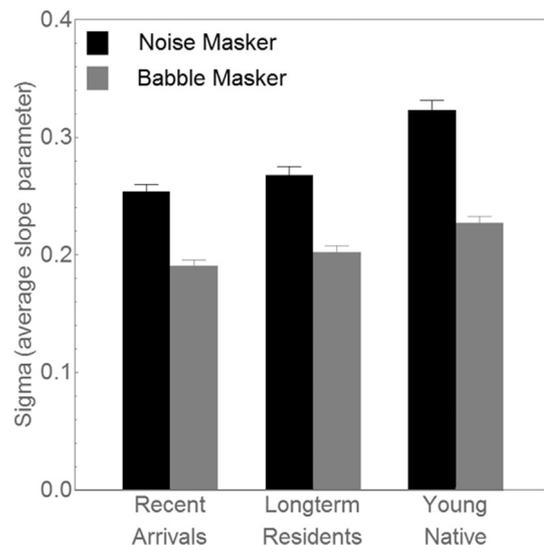


Fig. 4. Average slopes across the different word onset delays, based on individual psychometric functions for the three experimental groups, for the noise (black) and babble (gray) maskers. The vertical bars depict the standard errors of the means.

to the time course of stream segregation by varying the time-delay between the onset of auditory masker (babble or noise) and the onset of a spoken target word. We tested two groups of 30 young (aged 17–25) EL2 listeners, longterm residents and recent arrivals (arrived at the age of 7–14, and after 15, respectively). These data were compared to that of young EL1 listeners taken from our previous study (Ben-David et al., 2012). The results of this comparison show that the rate at which word recognition improved with word-onset delay was unaffected by linguistic status. This was evident both when words were presented in noise, and in babble. However, the asymptotic values (50% spoken-word recognition), and the slopes of the psychometric functions (how much benefit a listener reaps from a more favorable dB SNR) were affected by linguistic status.

#### 4.1. Word recognition thresholds

As expected, young EL2 listeners required a significantly higher SNR than their young EL1 counterparts to reach 50% correct spoken-words recognition. This was evident both in babble and in noise, irrespective of the word-onset delay. Similar findings have been noted repeatedly in the literature, using a variety of tasks and stimuli (e.g., Ezzatian et al., 2010; Avivi-Reich et al., 2014). This advantage of L1 listeners over L2 listeners, even longterm residents who have for all intents and purposes become fluent, is consistent with the critical period hypothesis – the notion that phonological encoding (Florentine, 1985; Mayo et al., 1997), and the acquisition of the statistical properties of a language (Saffran et al., 1996) must occur before a certain age.

In the current study, we found a gradient of performance, with higher dB SNR for EL2-recent arrivals than for EL2-longterm residents (by 2.13 dB for babble and 1.54 dB for noise). In turn, we found higher dB SNR for EL2-longterm residents than for EL1 listeners (by 1.28 dB for babble and 0.91 dB for noise). This improved performance for EL2-longterm residents over EL2-recent residents has also been previously documented (Ezzatian et al., 2010). There are several possible reasons which could account for this difference. Mainly, an improved ability to pre-attentively process the spectral cues of L2 due to training (Ylinen et al., 2010) and a greater familiarity and more extensive experience with the L2 lexicon.

Examining the *rate* at which spoken word recognition thresholds improved with word-onset delay presented a different picture – it appeared to increase independently of linguistic status and experience. Specifically, the same exponential decay function was found to govern the segregation of the target speech from the masker for all three groups (EL1, EL2-recent, EL2-longterm). These results are consistent with the hypothesis that the rapidity with which young listeners are able to form either a noise or babble object that can be distinguished from the speech target, is independent of their linguistic status.

To place our results in a broader theoretical framework, we examine Bregman's stream segregation models (1990), which distinguished between primitive and schema-based mechanisms. The primitive mechanism is a bottom-up, pre-attentive, sensory partitioning mechanism in the auditory system, which uses basic cues present in the stimulus. The schema-based one is a top-down, selection mechanism, which uses acquired knowledge about the sound to process and organize the incoming sensory input. The current findings may imply that this primitive mechanism plays a large role in stream segregation in the tested paradigm. As linguistic experience had no significant effect on the rate of improvement with word-onset delay, this suggests that more central cognitive mechanisms (which might be affected by linguistic experience) play, at best, a minor role with respect to how rapidly stream segregation is achieved.

#### 4.2. Slopes of the psychometric functions

The slopes of the psychometric functions that relate percent-correct spoken word recognition to SNR reveal a difference between EL1s and EL2s, which was not reflected in the thresholds' decay rate. EL1s had steeper slopes than EL2s (recent and long-term). The steeper slope values imply that EL1 listeners' spoken word recognition improved to a greater extent as the SNR values increased than those of EL2 listeners, irrespective of masker type. That is, EL1 listeners reaped a greater gain from increasingly more favorable SNRs under all masker types compared to that reaped by EL2s. This native-listener advantage in slope was also found in previous studies with noise maskers (van Wijngaarden et al., 2002; Bradlow and Bent, 2002; Ezzatian et al., 2010), as well as with babble maskers (Mayo et al., 1997).

How can this effect of linguistic status on the psychometric slopes be explained? The increase in dB SNR might boost the number of phonemes correctly heard in a word, irrespective of linguistic status. But it will likely reduce lexical competition to a greater extent in L1 than in L2 listeners, because phonemes in the L1 are likely represented more accurately than in L2 (Florentine, 1985; Mayo et al., 1997). Note that there is some evidence to show that phonemes activate word alternatives even at the end of the word (Sommers and Amano, 1998; Allopenna et al., 1998; Hadar et al., 2016), as supported by continuous activation models (e.g., TRACE model McClelland and Elman, 1986). Thus, the additional phonemes heard in L2 (due to the increase in dB SNR) may activate alternatives in both L1 and L2, whereas additional phonemes heard in L1 are more likely to activate alternatives in L1 only.

The current study also found an interaction between the linguistic status and the type of masker with respect to slopes. Specifically, the advantage in slope for noise over babble was significantly larger for EL1s than for EL2s. This may indicate that decreasing the resemblance between the target and the masker (moving from babble to noise masker) creates a larger release from masking for EL1 listeners, as the dB SNR became more favorable. Thus, EL1 listeners do not only need a smaller dB SNR in order to identify the target, they also benefit more from an increase in the dB SNR and from a reduction of the similarity of the background to the target. However, this EL1 advantage in slope may be limited to these types of maskers. Ezzatian et al. (2010) found that the slope difference between EL1s and EL2s disappeared when listeners were asked to repeat nonsense sentences masked by two other talkers. Hence, the native-listener advantage in slope, as evidenced in our study, might be lost when the informational content of the masker is very closely related to the target speech, and the task is more complex (repeating sentences rather than individual words).

#### 4.3. Stream segregation across the lifespan

Comparing the results obtained with EL1 and EL2 young adults with those obtained by older adults in our previous study (Ben-David et al., 2012) provides some intuitions on the nature of stream segregation in aging in different listening environments. Namely, with a noise masker, the benefit listeners can gain from a word-onset delay seems to be independent of linguistic status and age. Older EL1 listeners (mean age 71 years) benefitted from the early onset of a noise masker at the same rate as young EL1 listeners (Ben-David et al., 2012). For a babble masker, however, there appears to be an age-related effect. Young adults, irrespective of linguistic status, appear to benefit from a word-onset delay; however, older adults were not able to benefit at all from delays up to 1.1 s in length.

In general, the ability to separate a target stream from the acoustic background depends on a number of bottom-up acoustic

cues that differentiate the target from the background. Stream segregation is also facilitated when the listener has prior knowledge as to the location, acoustic characteristics, or content of the target speech (top-down; [Sohoglu et al., 2012](#); [Sohoglu et al., 2013](#)). Therefore, age-related difficulties in stream segregation could arise because of age-related changes in the auditory system, and/or from a failure to make effective use of top-down knowledge to help isolate the target stream from the background. When the acoustic cues differentiating the two streams are bountiful (e.g., target speech in a steady-state noise background), we would expect age-related changes in basic auditory processes to have less of an impact on the time course of stream segregation, than when there are fewer acoustic cues that differentiate the target stream from the acoustic background (e.g., target speech in a babble background). The results reported here, appear to support a sensory source – an age-related decline in basic auditory processes that have a more profound effect on stream segregation the greater the similarity between target and masker. When the target is speech and the background is noise, stream segregation appears to be equally rapid in older and young adults, independent of linguistic status. However, when the background is babble, which is more similar to the speech signal than is steady-state noise, stream segregation appears to take longer in older adults than it does in both EL1 and EL2 young adults.

The current study implies that bottom-up sensory mechanisms play a large role in stream segregation in the tested paradigm (as linguistic experience was not found to have a significant effect on the rate of stream segregation). Our previous study shows that sensory similarity between the target and the background (i.e., speech and babble) impeded stream segregation for older but not for young adults. Taken together, the results of the two studies can be taken as supporting the sensory degradation hypothesis. This hypothesis argues that age-related changes at different levels of the auditory system (see [Schneider, 1997](#); [Schneider and Pichora-Fuller, 2000](#) for reviews) degrade the emergence and representation of auditory objects. This, in turn, impedes the rapid and efficient operation of more central linguistic and cognitive functions that are necessary for speech recognition and comprehension. The notion that changes in perception in older adults may cause changes in cognition has been supported by a number of studies, which found that perceptual and cognitive processes tend to be highly correlated in older adults ([Lindenberger and Baltes, 1994](#); [Baltes and Lindenberger, 1997](#)). Moreover, several studies indicate that age-related cognitive differences could be minimized, if not eliminated, by equating for the perceptual differences ([Humes and Christopherson, 1991](#); [Schneider et al., 2000](#); [Murphy et al., 2006](#); [Avivi-Reich et al., 2014](#); [Ben-David et al., 2011](#); for an equivalent example in vision, see [Ben-David and Schneider, 2009, 2010](#); [Rabaglia and Schneider, 2016](#)). The current data could be seen as an additional example of how age-related changes in basic auditory processes could have secondary effects on more complex abilities, such as scene analysis.

In sum, both older L1 listeners and young L2 listeners will find it more difficult to recognize and comprehend speech in noisy environments than do their young L1 counterparts, but probably for different reasons. Age-related losses in basic auditory abilities are most likely to be the cause of higher thresholds in older adults, whereas insufficient or incomplete lexical development and knowledge seems to be responsible for the higher thresholds in young L2 listeners.

## 5. Conclusions

Young EL2 adults do not seem to find it more difficult to rapidly segregate a speech target from either a noise or babble background

than do young EL1s. However, EL2s were found to have higher word recognition SNR thresholds and lower slopes (the gain reaped from increasingly favorable SNRs) compared with EL1s, implying that lexical access, but not the time course of stream segregation, depends on the linguistic status of the listener. In other words, stream segregation appears to be rapid in young adults and independent of their linguistic status. Within EL2 listeners, we found that the extent of linguistic experience (recent vs. longterm residents) affects speech reception thresholds, but not the time course of stream segregation. Comparing our current study with a previous one ([Ben-David et al., 2012](#)), suggests that older EL1 listeners are able to use the delay of the target from a noise masker in a similar fashion as EL1 and EL2 young adults. However, when the masker was babble, they did not gain from a 1.1 s delay. Finally, the appropriate remedy for all groups (young and older, EL1s and EL2s) is to reduce background maskers, and enhance the cues (e.g., clear speech, visible speech cues, context, spatial separation of talkers, [Schneider et al., 2010](#)) that will better differentiate the signal from the background.

## Acknowledgments

This study was partially supported by a group grant on Sensory and Cognitive Aging, funded by Canadian Institutes of Health Research Grants (STP-53875, MGC-42665, & MOP-15359), the Natural Sciences and Engineering Research Council of Canada (RGPIN 9952-13) and by the European Commission's ERA-AGE 2 call for multidisciplinary research applications (JCRA) on "Active and Healthy Ageing across the Life Course" funded by the CIHR (HEARATTN). The first author was partially supported by a Marie Curie Career Integration Grant (FP7-PEOPLE-2012-CIG) from the European Commission.

We wish to thank the following students, Stephanie Assman, Karin Multani, and especially Michelle Bae for their assistance in collecting the data. We also wish to thank Dr. James Qi for his assistance in creating the stimuli.

## Appendix

Exponential decay functions were used in [Fig. 3](#) to describe how the 50% thresholds for word recognition declined as a function of the delay between masker and word onset. In fitting these functions to the data we assumed that the rate of decay differed between the two types of maskers (noise vs babble), but was the same for each of the three groups within a masker condition. Each point on this graph represents the mean of the thresholds of the 29, 30, and 30 individual thresholds at that combination of delay and masker type for the EL1s, EL2-Longterm, and EL2-Recent participants, respectively. The exponential decay functions shown in this graph are based on the least squares estimates of the parameter values that minimized the sum of squared deviations of the 445 threshold values in each of the two panels from the fitted function.

The exponential decay functions shown in [Fig. 3](#) were fit to all of the threshold values obtained from each of individuals in this experiment rather than fitting exponential decay functions to each individual and then averaging the individual exponential functions for the following reasons. First, the rate of decay of an exponential function is a joint function of two parameters ( $b$  and  $c$ ). If two sets of data are to have the same rate of decay it must be the case that  $b_1 = b_2$ , and  $c_1 = c_2$  in Equation (3). Hence an analysis based on the individual determined  $c$  values which shows that the  $c_{1,i}$  coefficients do not differ significantly from the  $c_{2,i}$  coefficients is not necessarily consistent with the null hypothesis that rate of decay is the same in both groups 1 and 2 for the  $n$  individuals in each group ( $1 \leq i \leq n$ ). Second, fitting a 3-parameter exponential decay

functions based on only five observations does not sufficiently constrain the 3 parameters, especially when the rate of decay is rather shallow. For instance, if threshold did not change with delay, the best-fitting exponential decay function would be

$$y_i = \bar{y} + 0 * e^{-c * t}$$

Note that as  $b$  approaches 0,  $c$  becomes unconstrained and can take on any value. Hence a small number of points cannot sufficiently constrain  $c$  when the function relating thresholds to delays is flattened. For these reasons we took a different approach to support our hypothesis that the rate of decay is identical for the three groups in the babble masker, and for the three groups in the noise masker, with the rate of decay differing between the two masker types. Specifically, we estimated the local rate of decay at four points, by determining the slope of the line between adjacent points on the graph relating thresholds to delay. The slope of the line between a delay of 100 and 225 ms for a single individual is given by the change in threshold between these two points divided by the delay difference between these two points measured in seconds ( $0.225 - 0.125 = 0.125$  s). Next we determined the slopes of the lines between 225 and 350 ms, 350 ms and 600 ms, and 600 and 1100 ms in the same fashion. Hence four slopes were determined for an individual for each of the two types of maskers. These four slopes then represent estimates of the local slope of the exponential decay at 162.5, 287.5, 475, and 850 ms, respectively. These individually determined slopes were then subjected to a 3 Group (EL1, EL2-longterm, EL2-recent)  $\times$  4 Local Slope Values  $\times$  2 Masker Types (noise vs babble) mixed-factorial ANOVA with Group as a Between-Subjects Factor, and Local Slope Values and Masker Type as Within-Subject factors, the results of which are reported in the text.

## References

- Alloppenna, P.D., Magnuson, J.S., Tanenhaus, M.K., 1998. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439.
- Avivi-Reich, M., Daneman, M., Schneider, B.A., 2014. How age and linguistic competence alter the interplay of perceptual and cognitive factors when listening to conversations in a noisy environment. *Front. Syst. Neurosci.* 8 (21) <http://dx.doi.org/10.3389/fnsys.2014.00021>.
- Baltes, P.B., Lindenberger, U., 1997. Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? *Psychol. Aging* 12 (1), 12–21.
- Ben-David, B.M., Erel, H., Goy, H., Schneider, B.A., 2015. 'Older is always better:' age-related differences in vocabulary scores across 16 years. *Psychol. Aging* 30 (4), 856–862.
- Ben-David, B.M., Schneider, B.A., 2009. A sensory origin for aging effects in the color-word Stroop task: an analysis of studies. *Aging Neuropsychol. Cognit.* 16 (5), 505–534.
- Ben-David, B.M., Schneider, B.A., 2010. A sensory origin for color-word Stroop effects in aging: simulating age-related changes in color-vision mimics age-related changes in Stroop. *Neuropsychology. Dev. Cognit. Sect. B Aging Neuropsychol. Cognit.* 17 (6), 730–746.
- Ben-David, B.M., Chambers, C.G., Daneman, M., Pichora-Fuller, M.K., Reingold, E.M., Schneider, B.A., 2011. Effects of aging and noise on real-time spoken word recognition: evidence from eye movements. *J. Speech Lang. Hear. Res.* 54 (1), 243–262.
- Ben-David, B.M., Tse, V.Y.Y., Schneider, B.A., 2012. Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? *Hear. Res.* 290, 55–63. <http://dx.doi.org/10.1016/j.heares.2012.04.022>.
- Bialystok, E., Luk, G., 2012. Receptive vocabulary differences in monolingual and bilingual adults. *Biling. Lang. Cognit.* 15 (02), 397–401. <http://dx.doi.org/10.1017/S136672891100040X>.
- Bradlow, A.R., Bent, T., 2002. The clear speech effect for non-native listeners. *J. Acoust. Soc. Am.* 112, 272–284. <http://dx.doi.org/10.1121/1.1487837>.
- 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. Am.* 106, 2074–2085. <http://dx.doi.org/10.1121/1.427952>.
- Bregman, A.S., 1990. *Auditory Scene Analysis: the Perceptual Organization of Sound*. MIT Press, Cambridge, Mass.
- Bregman, A.S., Campbell, J., 1971. Primary auditory stream segregation and perception of order in rapid sequences of tones. *J. Exp. Psychol.* 89 (2), 244–249.
- 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. Am.* 110 (5), 2527–2538.
- Chambers, C.G., Cooke, H., 2009. Lexical competition during second-language listening: sentence context, but not proficiency, constrains interference from the native lexicon. *J. Exp. Psychol. Learn. Mem. Cognit.* 35, 1029–1040.
- Cooke, M., Garcia Lecumberri, M.L., Barker, J., 2008. The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. *J. Acoust. Soc. Am.* 123, 414–427. <http://dx.doi.org/10.1121/1.2804952>.
- Durlach, N.I., Mason, C.R., Kidd, G., Arbogast, T.L., Colburn, H.S., Shinn-Cunningham, B.G., 2003. Note on informational masking. *J. Acoust. Soc. Am.* 113 (6), 2984–2987. <http://dx.doi.org/10.1121/1.1570435>.
- Ezzatian, P., Avivi, M., Schneider, B.A., 2010. Do nonnative listeners benefit as much as native listeners from spatial cues that release from speech masking? *Speech Commun.* 5, 919–929.
- FitzPatrick, I., Indefrey, P., 2009. Lexical competition in nonnative speech comprehension. *J. Cognit. Neurosci.* 22, 1165–1178.
- Florentine, M., 1985. Speech perception in noise by fluent, non-native listeners. In: *Proceedings of the Acoustic Society of Japan*. H-85–16.
- Garcia Lecumberri, M.L., Cooke, M., Cutler, A., 2010. Non-native speech perception in adverse conditions: a review. *Speech Commun.* 52, 864–886. <http://dx.doi.org/10.1016/j.specom.2010.08.014>.
- Hadar, B., Skrzypek, J.E., Wingfield, A., Ben-David, B.M., 2016. Working memory load affects processing time in spoken word recognition: evidence from eye-movements. *Front. Neurosci. Sect. Audit. Cognit. Neurosci.* 10, 220.
- Heinrich, A., Schneider, B.A., Craik, F.I.M., 2008. Investigating the influence of continuous babble on auditory short-term memory performance. *Q. J. Exp. Psychol.* 2006 61 (5), 735–751. <http://dx.doi.org/10.1080/17470210701402372>.
- Humes, L.E., Christopherson, L., 1991. Speech identification difficulties of hearing-impaired elderly persons: the contributions of auditory processing deficits. *J. Speech Hear. Res.* 34 (3), 686–693.
- Kroll, J.F., Steward, E., 1994. Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* 33, 149–174. <http://dx.doi.org/10.1006/jmla.1994.1008>.
- Lindenberger, U., Baltes, P.M., 1994. Sensory functioning and intelligence in old age: a powerful connection. *Psychol. Aging* 9, 339–355.
- Mayo, L.H., Florentine, M., Buus, S., 1997. Age of second language acquisition and perception of speech in noise. *J. Speech Lang. Hear. Res.* 40 (3), 686–693.
- McClelland, J., Elman, J., 1986. The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86.
- Meador, D., Flege, J.E., Mackay, I.R.A., 2000. Factors affecting the recognition of words in a second language. *Biling. Lang. Cognit.* 3 (1), 55–67.
- Murphy, D.R., Daneman, M., Schneider, B.A., 2006. Why do older adults have difficulty following conversations? *Psychol. Aging* 21 (1), 49–61. <http://dx.doi.org/10.1037/0882-7974.21.1.49>.
- Nakamura, K., Gordon-Salant, S., 2011. Speech perception in quiet and noise using the hearing in noise test and the Japanese hearing in noise test by Japanese listeners. *Ear Hear.* 32 (1), 121–131. <http://dx.doi.org/10.1097/AUD.0b013e3181eccdb2>.
- Portocarrero, J.S., Burright, R.G., Donovick, P.J., 2007. Vocabulary and verbal fluency of bilingual and monolingual college students. *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.* 22 (3), 415–422. <http://dx.doi.org/10.1016/j.acn.2007.01.015>.
- Rabaglia, C., Schneider, B.A., 2016. Age-related inhibitory deficit, or lack of familiarity benefit? Evidence from letter identification among visual distractors. *Atten. Percept. Psychophys.* 78, 542–565.
- Ragert, M., Fairhurst, M.T., Keller, P.E., 2014. Segregation and integration of auditory streams when listening to Multi-Part Music. *PLOS ONE* 9 (1), e84085. <http://dx.doi.org/10.1371/journal.pone.0084085>.
- Rogers, C.L., Lopez, A.S., 2008. Perception of silent-center syllables by native and non-native English speakers. *J. Acoust. Soc. Am.* 124, 1278–1293. <http://dx.doi.org/10.1121/1.2939127>.
- Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. *Science* 274 (5294), 1926–1928. <http://dx.doi.org/10.1126/science.274.5294.1926>.
- Scarborough, D., Gerard, L., Cortese, C., 1984. Independence of lexical access in bilingual word recognition. *J. Verbal Learn. Verbal Behav.* 23, 84–99.
- Schneider, B.A., 1997. Psychoacoustics and aging: implications for everyday listening. *J. Speech Lang. Pathol. Audiol.* 21, 111–124.
- Schneider, B.A., Daneman, M., Murphy, D.R., 2000. Listening to discourse in distracting settings: the effects of aging. *Psychol. Aging* 15 (1), 110–125. <http://dx.doi.org/10.1037/0882-7974.15.1.110>.
- Schneider, B.A., Pichora-Fuller, M.K., 2000. Implications of sensory deficits for cognitive aging. In: Craik, F.I.M., Salthouse, T. (Eds.), *The Handbook of Aging and Cognition*, second ed. Erlbaum, Mahwah, NJ, pp. 155–219.
- Schneider, B.A., Pichora-Fuller, M.K., Daneman, M., 2010. 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*. Springer, New York, NY, pp. 167–210.
- Schulpen, B., Dijkstra, T., Schriefers, H.J., Hasper, M., 2003. Recognition of interlingual homophones in bilingual auditory word recognition. *J. Exp. Psychol. Hum.*

- Percept. Perform. 29 (6), 1155–1178. <http://dx.doi.org/10.1037/0096-1523.29.6.1155>.
- Shinn-Cunningham, B.G., Best, V., 2008. Selective attention in normal and impaired hearing. *Trends Amplif.* 12 (4), 283–299. <http://dx.doi.org/10.1177/1084713808325306>.
- Sohoglu, E., Peelle, J.E., Carlyon, R.P., Davis, M.H., 2013. Top-down influences of written text on perceived clarity of degraded speech. *J. Exp. Psychol. Human Percept. Perform.* <http://dx.doi.org/10.1037/a0033206> (Advance online publication).
- Sohoglu, E.I., Peelle, J.E., Carlyon, R.P., Davis, M.H., 2012. Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci.* 32 (25), 8443–8453. <http://dx.doi.org/10.1523/JNEUROSCI.5069-11.2012>.
- Sommers, M.S., Amano, S., 1998. Lexical competition in spoken word recognition by younger and older adults: a comparison of the rime cognate, neighborhood, and cohort. *J. Acoust. Soc. Am.* 103, 2984–2984.
- Spivey, M., Martin, V., 1999. Crosstalk between native and second languages: partial activation of an irrelevant lexicon. *Psychol. Sci.* 10, 281–284.
- van Wijngaarden, S.J., Steeneken, H.J., Houtgast, T., 2002. Quantifying the intelligibility of speech in noise for non-native talkers. *J. Acoust. Soc. Am.* 112, 3004–3013.
- Wechsler, D., 1997. Wechsler Adult Intelligence Scale (WAIS–IIIuk): Wechsler Memory Scale (WMS–IIIuk). Psychological Corporation, San Antonio, TX.
- Weber, A., Cutler, A., 2004. Lexical competition in non-native spoken word recognition. *J. Mem. Lang.* 50, 1–25.
- Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., Näätänen, R., 2010. Training the brain to weight speech cues differently: a study of finnish second-language users of english. *J. Cognit. Neurosci.* 22 (6), 1319–1332. <http://dx.doi.org/10.1162/jocn.2009.21272>.