Hearing Research 290 (2012) 55–63

Contents lists available at SciVerse ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker?

Boaz M. Ben-David [a,b,c,*], Vania Y.Y. Tse [a], Bruce A. Schneider [a]

[a] Centre for Research on Biological Communication Systems, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road N., Mississauga, ON L5L 1C6, Canada
[b] Oral Dynamics Lab, Department of Speech-Language Pathology, University of Toronto, 160-500 University Ave., Toronto, ON, M5G 1V7, Canada
[c] Toronto Rehabilitation Institute, 550 University Ave., Toronto, ON M5G 2A2, Canada

ARTICLE INFO

Article history:
Received 24 December 2011
Received in revised form
18 April 2012
Accepted 24 April 2012
Available online 16 May 2012

ABSTRACT

Older adults often find it more difficult than younger adults to attend to a target talker when there are other people talking. One possible reason for this difficulty is that it may take them longer to perceptually segregate the target speech from competing speech. This study investigated age-related differences in the time it takes to segregate target speech from either a speech spectrum noise masker or a babble masker (many people talking simultaneously). Specifically, we employed five different delays (0.1 s–1.1 s) between masker onset and target speech onset. Four signal-to-masker ratios were employed at each delay to determine the 50% thresholds for word recognition accuracy when target words were masked by either speech spectrum noise or multi-talker babble. Thresholds for word recognition decreased exponentially as a function of the masker-word-onset delay, at the same rate for younger and older adults, when the masker was speech spectrum noise. When the masker was babble, thresholds for younger adults decreased exponentially with delay at the same rate as they did when the masker was speech spectrum noise. The word recognition thresholds for older adults, however, did not appear to change over the range of delays explored in this study. In addition, the average difference between word recognition thresholds for younger and older adults (younger adult thresholds < older adult thresholds) was significantly larger when the masker was babble than when it was noise. These results indicate that older adults are as fast as younger adults at separating speech from a steady-state noise masker, but are not as capable as younger adults of taking advantage of the delayed onset of the speech target when the masker is babble. The potential contributions of age-related sensory and cognitive declines to these stream segregation effects are discussed. Finally, we conclude that age-related differences in the timeline for stream segregation contribute to the difficulties older adults experience in listening to speech in a background of babble.

1. Introduction

Older adults typically find it more difficult to comprehend speech than younger adults when there are competing sound sources in the environment (see Schneider et al., 2010; for a review). One possible reason for these difficulties could be that they are not as efficient as younger adults at auditory scene analysis (Bregman, 1990). To comprehend what is being said by an individual, listeners first have to perceptually segregate the target speech from other competing sound sources. If the auditory and/or cognitive systems of older adults are less efficient or more sluggish at accomplishing this task, they will be at a disadvantage vis-à-vis younger adults in auditory environments with competing sound sources. In the present study, we compare the ability of younger and older adults to benefit from a delay between the onset of an auditory masker and the onset of a speech target in a word recognition task for two different types of maskers: steady-state speech spectrum noise and a multi-talker babble.

1.1. Acoustic factors that facilitate auditory stream segregation

Generally speaking, the greater the acoustic dissimilarity between the target and competing streams, the easier it is to perceptually segregate sound sources (Brungart et al., 2001; Alain

* Corresponding author. Oral Dynamics Lab, Department of Speech-Language Pathology, University of Toronto, 160-500 University Ave., Toronto, ON, M5G 1V7, Canada. Tel.: +1 416 946 8552; fax: +1 416 978 1596.
E-mail address: boaz.ben.david@utoronto.ca (B.M. Ben-David).

0378-5955/$ – see front matter
doi:10.1016/j.heares.2012.04.022

et al., 2001, 2006; Durlach et al., 2003). Hence, it should be easier to segregate words from a steady-state noise background than from a background consisting of many different people talking simultaneously. Moreover, a number of studies have indicated that it takes time for stream segregation to develop. For example, the amount of time it takes for listeners to perceive high- and low-frequency alternating tones as two tonal streams (a low-frequency stream and a high-frequency stream) depends on the temporal and spectral proximity of the two tones (Bregman, 1990; Bregman and Campbell, 1971; Carlyon et al., 2001, 2003; Cusack et al., 2004; Miller and Heise, 1950; Snyder et al., 2006; Sussman et al., 2007). With respect to segmenting speech stimuli from babble, Heinrich et al. (2008) suggest that segregation may be complete in younger adults after 500 ms. Given that competing speech is informationally more complex than babble, it is reasonable to speculate that it will take longer to segregate target speech from competing speech than to segregate target speech from steady-state noise. Indeed, Ezzatian et al. (in press) found that word recognition for syntactically-correct but semantically-anomalous sentences (e.g., "A *rose* could *paint* a *fish*") improved as the sentence unfolded in time when the background was competing speech, but not when the background was noise. That is, word recognition scores improved from the first to the second and to the third italicized words in the semantically-anomalous sentences, only with a speech-masker, whereas with a non-speech noise masker, performance was at ceiling on the first word. Finally, there is evidence that the ability to detect a tone (McFadden and Wright, 1990; Wright, 1997; Zwicker, 1965) or to recognize a spoken word (Wagener and Brand, 2005) improves when the masking stimulus begins before the target stimulus is presented. Presumably the prior onset of the masker allows the listener to "build up" a perceptual representation of the masker before the target stimulus occurs, and the establishment of this representation, and its maintenance once the target is presented, makes it easier for listeners to detect and/or recognize the target.

1.2. The effects of age on the build up of stream segregation

Aging could affect the build up of stream segregation in several ways. First, older adults may be more susceptible to energetic or peripheral masking. Energetic masking occurs when the signal-to-noise ratio (SNR) in regions of spectral overlap between the target signal and the masker is low enough for the energy in the masker to overwhelm the energy in the signal, making it difficult for the listener to extract the target signal from the background. Age-related changes in the auditory periphery (for a review, see Schmeidt, 2010) could result in older adults being more susceptible to energetic masking. This, in turn, may reduce the amount or quality of information provided by the auditory periphery that would permit more central auditory and cognitive processes to segregate the target speech from the masker. Second, older adults could also be more susceptible to non-energetic, e.g., informational masking. Informational masking of speech occurs when competing sound sources interfere with speech recognition at more central auditory and/or cognitive processing levels. Several factors could contribute to non-energetic (informational) masking. First, the similarity between masker and masker plus target could lead to interference at more central levels (Durlach et al., 2003). Second, when the masker itself is speech, the masker may activate phonemic and semantic processes that interfere with lexical access to the target words. Such activation could occur even when the masker consists of unrecognizable speech, such as a multi-talker babble, or speech in a foreign language (for a discussion of these issues, see Schneider et al., 2007; Simpson and Cooke, 2005; Van Engen and Bradlow, 2007).

Aging could increase a listener's susceptibility to non-energetic (informational) masking in several ways. First, as suggested by the work of Hasher and Zacks (1988) it could reduce the ability to focus attention on the speech target and inhibit the processing of information from irrelevant sources (for a review, see Schneider et al., 2007). Second, as suggested by the literature on generalized slowing of cognitive processing (e.g., Salthouse, 1996) stream segregation in the presence of energetic and informational masking may take a longer time to emerge in older than in younger adults. Finally, it is possible that older adults may benefit less than younger adults from the top-down (e.g., prior knowledge of content) and/or bottom-up (e.g., spatial separation) cues that release listeners from informational masking. This latter possibility, however, is not supported in the literature. Several studies indicate that cognitively intact older adults, with normal hearing for their age, benefit as much as younger adults from both bottom-up (e.g., Li et al., 2004; Humes et al., 2006) and top-down (Ezzatian et al., 2011; Singh et al., 2008) factors that release listeners from informational masking.

1.3. The present study

The purpose of the present study was to determine the degree to which the time course and effectiveness of stream segregation is affected by age, and by the similarity between the speech target and the masker. Younger and older adults were presented with spoken words masked by either babble or steady-state speech spectrum noise and were asked to repeat them as they were presented. Specifically, each trial began with the onset of a 4 s masker. Word-onset occurred 100, 225, 350, 600, or 1100 ms after masker onset, as the literature suggests that build up of the masker stream my take up to 500 ms when the masker is babble (Heinrich et al., 2008). Word identification accuracy was measured as a function of SNR for each type of masker at each of these five delays in both age groups. Several hypotheses concerning the effects of these factors on speech recognition were constructed based on the pertinent literature.

First, we hypothesized that performance would improve as word-onset delay increased. Wagener and Brand (2005) found that speech intelligibility for short sentences presented in speech-shaped noise was better when the background was on continuously than when the background noise was gated on and off with the sentences. One interpretation of this result is that when the background was continuous, listeners had time to build up and maintain a perceptual representation of the noise, thereby facilitating and/or speeding up the segregation of the sentence from the background. Hence, we would expect that word accuracy would increase the longer the masker is on before the word occurs. Varying this delay would allow us to determine a function relating word accuracy to word-onset delay. This function would represent the timeline for stream segregation.

Second, we might expect that it takes a longer time for a babble stream (consisting of many simultaneous voices) to coalesce into an auditory object than it takes for a noise stream to emerge as an auditory object. Specifically, we might expect to see a more rapid improvement in word accuracy as word-onset delay increases when the masker is noise than when it is babble. Moreover, based on previous literature (e.g., Freyman et al., 1999; Li et al., 2004), we would expect the slope of the psychometric function relating word accuracy to SNR to be steeper for noise than for babble. Namely, the rate of the increase in word identification as SNR increases should be more rapid when words are presented in noise. We also wanted to see if the slope of the psychometric functions changed with word-onset delay. Finally, we wanted to examine the extent to which age-related declines in either auditory or cognitive processing might exacerbate these effects.

2. Method

2.1. Participants

Thirty younger adults (mean age = 20 years, SD = 1.6) and thirty older adults (mean age = 72.3 years, *SD* = 3.8) participated in the study. The younger adults were undergraduates at the University of Toronto Mississauga and received either course credit or were paid $10/hour for their participation. The older adults were volunteers from the local community and were paid $10/hour. All participants were native English speakers as assessed by a self-report and achieved a minimum score of 9/20 on the Mill Hill Vocabulary Test (Raven, 1965), corresponding to normal vocabulary levels for native English speakers (e.g., see Ben-David et al., 2011a, 2011b; Ben-David and Schneider, 2010). The average Mill Hill score was 12.8/20 ($SD = 2.8/20$) for younger adults and 15.1/20 ($SD = 2.1/20$) for older adults ($t(58) = 3.59$, $p < .001$). 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 3-kHz in the better ear (≤20 dB HL for younger adults and ≤25 dB HL for older adults). Hearing levels (for the better ear) are shown in Fig. 1 for younger and older adults.

2.2. Stimuli and apparatus

We used 520 bi-syllabic recorded words, spoken by a female actor with a southern Ontario accent, taken from Murphy et al. (2000).Words were divided into ten lists of 52. To control for the impact of linguistic characteristics on word identification (see a discussion in Ben-David et al., 2011b), lists were equated on word frequency (as measured by the Hyperspace Analog to Language frequency norms, log-HAL, $M = 8.50$, $SD = 0.03$; $F(9, 510) = 0.2$, $p = 1.0$) and on the density of their phonological neighborhood ($M = 3.21$, $SD = 0.06$; $F(9, 510) = 0.01$, $p = 1.0$). Lexical characteristics were gathered from the English lexicon project data-base (Balota et al., 2007) on September 2009. The ten lists were also matched on average word duration ($M = 622$ ms, $SE = 4$; $F(9,510) = 1.43$, $p = .2$), with word duration ranging for an individual word from 412 to 1004 ms. Spoken words were later presented on the background of 4 s of either continuous speech spectrum noise or multi-talker babble taken from the 'Revised Speech Perception in Noise' (R-SPIN) test (Bilger et al., 1984). All of the 520 digital audio files of the spoken words were equated with respect to root-mean square amplitude. Word stimuli were delivered monaurally to the best ear (to the right ear of 16 out of 30 older and 17 out of 30 younger participants, respectively), by converting the digital signal to analog form (using a 16-bit digital-to-analog converter TDT DD1), and controlling the analog output using an Enhanced Real-time Processor (TDT RP2.1) and programmable attenuator (TDT PA5), before delivering the signal to a headphone buffer (TDT HB7) and a Sennheiser HD 265 headphone.

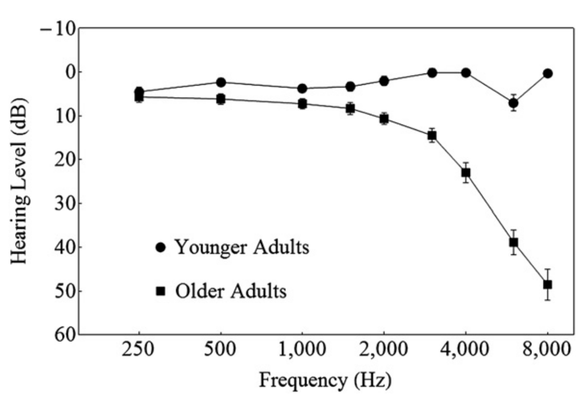

Fig. 1. Mean audiometric pure-tone air-conduction thresholds (for the best ear) as a function of frequency for the 30 younger and 30 older adults who participated in the study. The vertical bars depict the standard errors of the means.

2.3. Procedure

Participants were tested individually in a single-walled sound-attenuating booth. Thresholds for the detection of multi-talker babble were determined for each participant's better ear using an adaptive two-interval, two-alternative forced-choice paradigm taken from Heinrich et al. (2008). Throughout the experiment, words were presented to the listener's better ear at a level that was individually set to 50 dB above his/her babble threshold. However, to avoid presenting stimuli at a potentially harmful level, we adjusted the presentation level of the word stimuli to a maximum of 75 dB SPL for three older adults whose babble threshold exceeded 25 dB SPL (by 2, 3 and 6 dB). Placing this upper limit on word presentation minimized the "rollover effect" – the loss of intelligibility for some older listeners in high presentation levels (Jerger and Jerger, 1971; for a recent discussion, see Heinrich and Schneider, 2011). Average babble thresholds for younger and older adults were 17.37, and 21.18 dB SPL, respectively ($t(58) = 3.63$, $p < .001$).

In each experimental session, there were ten blocks of 52 trials. The order of the trials within blocks was randomized separately 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 in the following fashion: for half of the participants the 1st, 3rd, 5th, 7th and 9th block were babble and the 2nd, 4th, 6th, 8th, 10th were noise blocks, and for the other half this order was reversed. The word list assigned to a block was randomized. We also manipulated the word-onset delay – 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. To control for practice effects, the order of the five word-onset delays was counterbalanced across participants using a Latin square procedure (Grant, 1948). This counterbalancing procedure, when combined with the counterbalancing of the order of the masker type, generated ten experimental groups (six participants per group, three younger and three older adults), as illustrated in Table 1.

In each block, trials were presented in four different SNRs, 13 trials in each SNR. For younger adults, two different sets of four equidistant SNRs for babble and noise maskers were chosen: −23, −18, −13 and −8 dB for babble, and −10, −6, −2 and +2 dB, for noise. In a pre-test with ten younger adults (taken from the same population as our younger participants), these ranges were found to bracket the 50% intelligibility point. For older adults, we added 3 dB to each SNR point in both sets yielding values of −20, −15, −10 and −5 dB for babble, and −7, −3, +1, and +5 dB for noise. The use

Table 1
The order of word-onset delays (1100, 600, 350, 225, 100 ms). Both babble and noise maskers were presented at each delay, before proceeding to the next delay. Half of the participants in each order were presented with a babble block first, and the other half, with a noise block first.

N of participants	Order of word-onset delay (in ms) blocks				
6	1100	600	350	225	100
6	600	1100	100	350	225
6	350	225	1100	100	600
6	225	100	600	1100	350
6	100	350	225	600	1100

of different SNRs for younger and older adults (approximately by 3–4 dB) has been found to result in comparable overall word recognition scores in noise for both age groups (see Ben-David et al., 2011a; Murphy et al., 2000; Pichora-Fuller et al., 1995).

In total, participants were presented with 520 spoken word stimuli. They were asked to listen to each word and repeat it immediately. They were encouraged to guess whenever they did not hear a word properly, but if they did not hear anything, to respond “pass”. Before each block of trials commenced, three practice trials were presented, using the same type of masker and word-onset delay condition as the following block. No feedback was provided in either practice or experimental trials. Short breaks were given after each block, and a longer break was offered after the presentation of the fifth block.

2.4. Data analysis

Accuracy was coded by an experimenter listening to the participant’s responses via headphones during the experimental session. The participant’s oral responses were digitally recorded. A sample of 73% of the trials was later recoded by a different rater. Raters’ agreement reached 97.7% on average, with the lowest degree of inter-rater agreement for an individual’s recorded responses being 95%. In cases of disagreement, the words were scored from the recorded utterances.

These accuracy scores were used to construct psychometric functions relating accuracy to SNR ratio 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 (five delays X two types of maskers). However, we were not able to accurately estimate thresholds in 18 instances (17 for older adults, 1 for younger adults) out of 600 cases (30 subjects X two age groups X five delays X two types of maskers) either because accuracy did not exceed 50% correct at any of the four SNRs (10 cases for older adults in babble and 1 case for younger adults in noise) or response accuracy did not fall below 50% at any SNR (6 cases for older adults in noise and 1 case for older adults in babble). In these instances, we did not determine a threshold for that delay X masker condition because the 50% point estimated by fitting a psychometric function would be an extrapolated rather than an interpolated value. For older adults in babble there were 3, 1, 4, 2, and 1 instances corresponding to delays of 100, 225, 350, 600, and 1100 ms respectively. For older adults in noise, there were 1, 1, 2, 2, and 0 instances corresponding to delays of 100, 225, 350, 600, and 1100 ms, respectively. Finally there was one instance in which we were unable to estimate a 50% threshold for a younger adult (350 ms delay for the noise masker). These 18 instances (representing 3% of trials) were eliminated in computing the individual and average psychometric functions reported below.

To assess the extent to which these deletions might affect average performance, we first computed average accuracy scores with and without these deletions at each of the SNRs for all of the 10 delay conditions at which at least one deletion occurred. Next, at each of these 10 delay conditions we regressed the average accuracy scores with the deletions against the average accuracy scores without the deletions. The lowest of these 10 squared correlation coefficients was 0.999, which occurred for the older adults at a delay of 100 ms in the babble condition. All of the other nine conditions had squared correlation coefficients (rounded to three decimal places) of 1.000. Hence, the omission of these 18 instances is unlikely to have had a noticeable effect on the mean performance of younger and older individuals at any of the delay conditions. Thus, they were excluded in subsequent analyses.

3. Results

Fig. 2 presents the average accuracy for younger adults (leftmost panel) and older adults (rightmost panel) for word identification masked by babble (top functions) or speech spectrum noise (bottom functions) for the five word-onset delays (100, 225, 350, 600 and 1100 ms), as a function of dB SNR. For each word-onset delay, the data points (for average word identification accuracies[1] across all participants) were fit by logistic psychometric functions of the form,

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

where y represents the probability of correctly identifying the target word, x is the SNR in dB, μ represents the dB SNR level corresponding to 50% correct performance, and σ determines the slope of the fitted function. Psychometric functions were computed by minimizing Chi-Square (for a more detailed description, see Yang et al., 2007).

The 50% thresholds calculated from the data in Fig. 2 are plotted in Fig. 3 as a function of word-onset delay. To facilitate comparisons, the thresholds for the old-noise, young-babble, and young-noise conditions have been normalized to have a mean SNR of 0 dB, whereas the thresholds for old-babble have been normalized to have a mean of 2 dB. An examination of Fig. 3 suggests that thresholds decrease in an approximately identical exponential fashion as a function of delay for younger adults in both types of maskers and for older adults in noise. Fig. 3 also indicates that there is very little change in thresholds as a function of delay for older adults in babble.

To confirm this pattern of change in thresholds as a function of delay, and to show that it also characterized the individual participants in each age group, we fit *individual* psychometric functions (see Eq. (1)) to each participant’s data at each of the five word-onset delay conditions for both masker types (five delays X two masker types = 10 psychometric functions for each participant).

3.1. Word recognition thresholds

3.1.1. Thresholds as a function of word-onset delay

Average thresholds for younger (solid symbols) and older adults (open symbols) are plotted as a function of word-onset delay in Fig. 4, for noise maskers (Panel A) and for babble maskers (Panel B). The best model that fits the data is one in which: 1) thresholds do not change with word-onset delay for the older adults in babble; 2) thresholds decay exponentially with word-onset delay at the same rate for the remaining three conditions; and 3) the four conditions differ with respect to asymptotic values. The solid lines shown in Fig. 4 represent the predictions of the model specified in Eq. (2) below:

$$\begin{cases} y_{\text{old,babble}} = a_1 \\ y_{\text{young,babble}} = a_2 + be^{-c\,x} \\ y_{\text{old,noise}} = a_3 + be^{-c\,x} \\ y_{\text{young,noise}} = a_4 + be^{-c\,x} \end{cases}, \tag{2}$$

where y represents the 50% threshold in dB SNR for correct detection of a spoken word and x is the noise-onset delay. Note that the decay function be^{-cx} is unchanged across three conditions, yet the asymptote, a, is different for each function. The statistical

[1] As mentioned in the Data Analysis (Section 2.4), these averages include the 582 cases out of 600 in which we could estimate a 50% threshold.

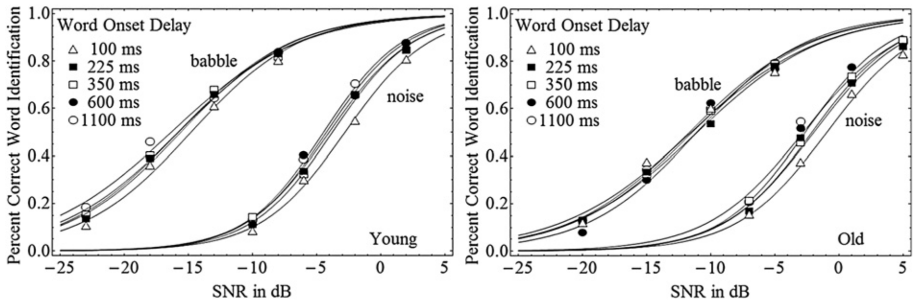

Fig. 2. Average percent-correct word identification as a function of SNR in dB for younger (left panel) and older adults (right panel) for the five different word-onset delays. The psychometric functions for the different delays are associated with the different symbols identified in the figure. The top functions represent the babble masker condition and the bottom functions the noise masker condition.

analyses that provide support for this model are presented in the Appendix A.

3.1.2. The effect of masker type and age on thresholds

Fig. 4 also indicates that 50% thresholds for word identification in babble were lower than in the noise masker. This difference in mean thresholds between noise and babble is larger for older than for younger adults at each of the five delays, i.e., there appears to be an interaction between masker type and age at each of the delays. To test for an interaction at each of the five delays, we first computed the individual differences in 50% thresholds between babble and noise masker conditions for all participants.[2] The extent of the difference between noise and babble thresholds (thresholds in noise minus thresholds in babble) for younger adults was found to be significantly smaller than the extent of the babble advantage for older adults, at each of the five delays (see Appendix B). For both younger and older adults, thresholds were also significantly lower for babble than for noise at each delay (see Appendix B). Finally, for both types of maskers, thresholds were significantly lower in younger adults than in older adults, at each of the five delays (see Appendix B).

3.2. Slopes of the psychometric functions

3.2.1. Slopes as a function of word-onset delay

The parameter σ in Eq. (1) specifies the slope of the psychometric function at the word-onset time corresponding to the 50% threshold. Average slopes at these points for younger (solid symbols) and older adults (open symbols) are plotted as a function of word-onset delay in Fig. 5, for the noise masker (top two functions) and for the babble masker (bottom two functions). An examination of these plots suggests that the slope value does not vary in any systematic way with word-onset delay. Statistical analyses (see Appendix C) confirmed that there was no evidence of change in slope with word-onset delay in any of the four conditions. The solid lines in Fig. 5 represent the predictions of a model in which slopes are independent of delay.

3.2.2. The effect of masker type and age on slopes

An examination of Fig. 5 also suggests that slope values for the noise masker are higher than for the babble masker, and that this difference in slope values is the same for younger and older adults. To test for any evidence of interaction between masker and age, we computed the individual differences in slope values between babble and noise masker conditions.[3] No evidence of any interaction between masker type and age was found at any of the five delays (see Appendix C). However, *t*-tests indicated that slopes, averaged across age groups, were significantly higher in noise than in babble at all delays (see Appendix C). Finally, there was no suggestion of any age-differences in slopes, averaged across masker type, at any of the five delays (see Appendix C). In sum, the rate of the increase in word identification as SNR increases was more rapid when words were presented on the background of a noise masker. Neither age nor word-onset delay were found to affect the slopes of the psychometric functions.

4. General discussion

The purpose of this study was to assess the role age-related differences in auditory scene analysis play in the difficulties older adults experience in understanding speech in a noisy environment. To that end, we measured age-related effects on the time course of stream segregation by varying the time-delay between the onset of an auditory masker (steady-state noise or multi-talker babble) and the onset of a spoken word.

4.1. Word recognition thresholds

When the masker was speech spectrum noise, the rate at which word recognition improved with word-onset delay appeared to be independent of age. Yet, older adults required a higher SNR for word recognition than younger adults at all of the delays tested. Hence, older adults appear to be more susceptible than younger adults to steady-state noise maskers. But, in spite of this age-related difference in the overall effects of the masker, both age groups benefitted to the same extent from the onset delay between the noise masker and the word. This suggests that even though word recognition in noise is, in general, more difficult for older than for younger adults, that there are no age-related differences with respect to the ability to use the prior onset of the noise masker to build up the percept of a noise stream. Presumably, the ability to form an auditory object of a steady-state noise masker, prior to the onset of word, facilitates stream segregation to the same extent for older and younger adults.

When the masker was multi-talker babble, word-recognition thresholds were, on average, higher for older than for younger adults. This age-related difference for babble was larger than for noise. Moreover, while thresholds improved with word-onset delay for younger adults, there was no evidence of improvement over

[2] We excluded comparisons where the data point for either the babble or noise masker condition was removed from the data set because the 50% threshold was not bracketed (see Section 2.4). This reduced the number of threshold differences (noise minus babble) obtained from older participants by 4, 2, 6, 4, and 1 comparisons for word-onset delays 100, 225, 350, 600, and 1100 ms, respectively; and reduced the number of threshold differences obtained from younger participants by 1 comparison at a delay of 350 ms.

[3] This was done only when valid threshold estimates were available (see footnote [2]).

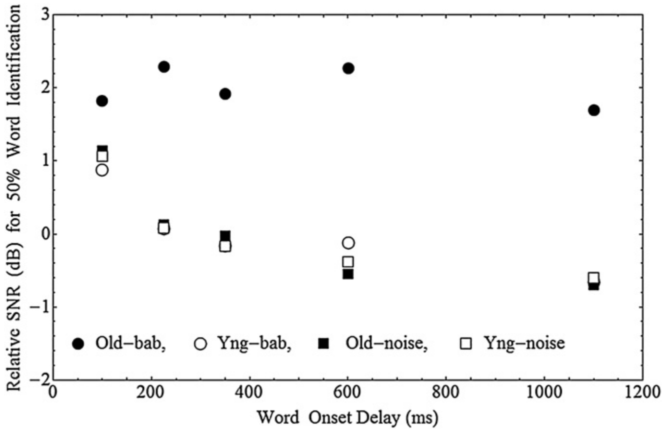

Fig. 3. Thresholds (dB SNR values for 50% correct identification) as a function of word-onset delay taken from the psychometric functions shown in Fig. 2, for older adults (filed symbols) and younger adults (empty symbols) in babble (circles) and in noise (squares). To facilitate comparisons, the thresholds for old-noise, young-babble, and young-noise have been normalized to have a mean SNR of 0 dB, whereas the thresholds for old-babble have been normalized to have a mean of 2 dB.

a 1 s period for older adults. The larger age-related difference for babble than for noise suggests that aging has a larger deleterious effect on word recognition when the masker is informational than when it is primarily energetic. Moreover, the lack of a word-onset effect in older adults with a babble masker suggests that: 1) it is more difficult for them to build up an auditory object of the babble masker than it is for younger adults; or 2) it takes them longer than 1.1 s to do so; or 3) the formation of an auditory masker is not as helpful to them as it is to younger adults in separating the babble stream from the target words.

Why is a babble masker more challenging for older than for younger adults? One reason might be the greater degree of similarity between the babble masker and the target words than between the noise masker and the target words. The babble masker and target words are acoustically similar – both are voiced and have amplitude envelopes that vary over time. By way of contrast, the noise masker is steady-state and without vocality. Age-related changes in the ability to segregate voices based on differences in the fundamental frequency (Humes et al., 2006; Vongpaisal and Pichora-Fuller, 2007; Helfer and Freyman, 2008), for instance, would make it relatively more difficult for older adults to identify words in a babble background. Yet, this age-related difference would not have any effect on their ability to segregate words from noise. In addition, age-related differences in the ability to benefit from gaps and/or amplitude fluctuations in a masker due to hearing losses (Bernstein and Grant, 2009), or to an age-related decline in auditory temporal processing (George et al., 2006; Lorenzi et al., 2006), would also make stream segregation relatively more difficult for older than for younger adults when the masker is babble. Hence, age-related differences in sensory processing could be contributing to the greater age-related difference in performance when the masker was babble than when it was noise.

It is possible that the age-related differences in performance are larger for babble than for noise because older adults are less able than younger adults to inhibit phonemic processing of the babble background (Hasher and Zacks, 1988). Such activation could impair lexical access and lead to more errors in word recognition, thereby exacerbating age-differences in performance in a babble masker compared to a noise masker. It is also possible that age-related differences in either vocal segregation and/or the ability in inhibit activation of phonemic processing of the babble could result in delays in stream segregation. Thus, it is reasonable to assume that in a cocktail party situation, the added interference presented by babble obstructs timely stream segregation for older adults, thereby disrupting speech intelligibility. Consider, for example, a conversation that takes place in a loud crowded room, where the target speech is occluded by the speech of many other speakers surrounding the listener. Phonemes from competitor speech may compete (and get confused) with the target speech utterance. This will not occur if the same conversation were to take place in a construction zone, where the noise masker is distinct from the target speech. Our data suggests that in the former example, this challenging situation may prevent seniors from taking full advantage of stream segregation cues, such as the onset of the babble prior to the speech target.

It is worth noting at this point that both older and younger adults required higher SNRs in noise than in babble for equivalent word recognition performance. Fig. 6 shows that spectral differences among words, babble and noise would predict this result. Fig. 6 plots the spectra of these three stimuli, when all three are equated with respect to average power. Note that the SNR of words to babble over most of the low-frequency region (<1500 Hz) is higher than the SNR of words to noise. Note also that in the high-frequency region (>1500 Hz), the SNR of words to babble is higher than the SNR of words to noise, again making it harder to recognize words in a background of noise than of babble. Thus, on the basis of both low- and high-frequency information, words should be harder to recognize in noise than in babble.

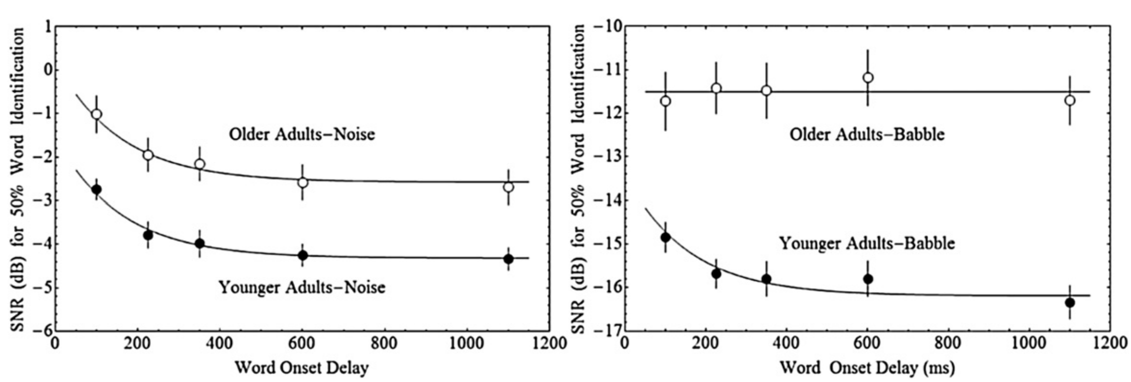

Fig. 4. Average thresholds (dB SNR values for 50% correct identification) as a function of word-onset delay, based on individual psychometric functions fit to younger (solid circles) and older (open circles) adults. The left panel presents data from the noise masker condition and the right panel presents data from the babble masker condition. The vertical bars depict the standard errors of the means. Data points for old-noise, young-noise and young-babble conditions were fit by an exponential decay functions of the type, $y = a + be^{-cx}$, where b and c are the same for all three conditions, with the asymptote, a, varying across the three conditions. Data points for old-babble condition were fit by a constant (mean threshold).

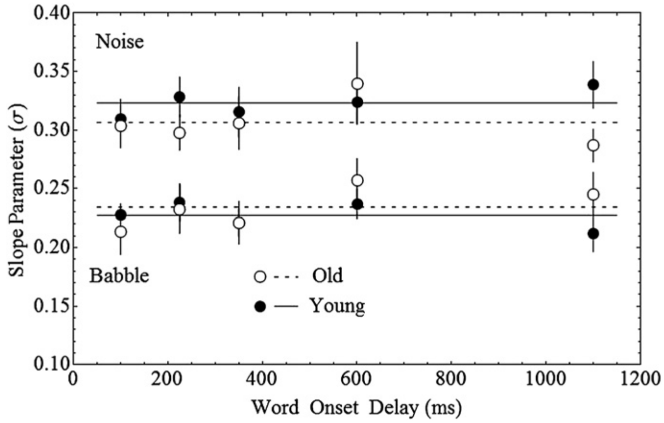

Fig. 5. Average slopes as a function of word-onset delay, based on individual psychometric functions fit to younger (solid circles) and older (open circles) adults. The top two functions present data from the noise masker condition, and the bottom two functions present data from the babble masker condition. The vertical bars depict the standard errors of the means. Data points in each condition were fit by a constant (the mean slope for each condition).

4.2. Slopes of the psychometric functions

The slopes of the functions relating percent-correct word identification to SNR portray a different picture than accuracy thresholds. Even though younger adults were found to have lower accuracy thresholds than older adults, the slopes in all noise-onset delays and masker conditions were not affected by age. The slopes were also not affected by varying the word-onset delay for either age group. In fact, a single slope was found to fit all data for babble masker conditions and another steeper slope fit all data for noise masker conditions. That is, the rate of the increase in accuracy as SNR increases was constant across conditions, given a specific auditory masker.

Our results on the invariability of slopes validate a procedure that is commonly used to control for age-related sensory (peripheral) differences in the masking of words, when there is no contextual support to aid their recognition (for a review, see Schneider et al., 2010). Specifically, 50% thresholds for the final word of low-context SPIN sentences are first determined. Then, the SNR of target speech to background masker is individually adjusted by this threshold. This procedure is based on the tacit assumption that the slope of the psychometric function does not change with age. Thus, this adjustment should provide equal word recognition for all individuals regardless of age, when there is no contextual support for the words. Indeed, the slope results in this study support this assumption, both when the masker is noise and when it is babble.

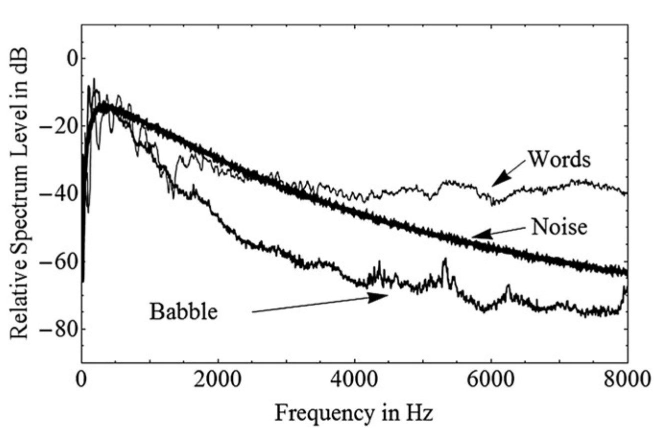

Fig. 6. The average spectra of 520 target word stimuli, babble masker and noise masker, in adjacent 10-Hz rectangular frequency bands. The three different stimuli were equated with respect to root mean square amplitude prior to spectral analysis, so that the average power is the same for the target words, babble masker, and noise masker.

5. Conclusions

The data identify age-related differences in the development of stream segregation as a contributing source for the difficulties older adults experience in listening to speech on the background of babble (but not in noise). For older adults, even after 1 s, the build up of a babble auditory stream has not been completed. As a result, the information in the beginning of a spoken sentence might not be fully processed, impairing their ability to understand the conversation. These results also suggest that multi-talker babble, made of unrecognizable speech, presents a unique challenge for older adults, above and beyond that due to energetic masking.

Acknowledgments

This study was partially supported by Canadian Institutes of Health Research Grants (STP-53875, MGC-42665, & MOP-15359), and a research opportunity program grant from the Faculty of Arts & Science at the University of Toronto Mississauga. The first author was partially supported by a grant from the Ontario Neurotrauma Foundation (2008-ABI-PDF-659). We wish to thank the following students, Wu Yan (Lulu) Li, Julio Pereira, Marissa Polidori and especially Caterina Leung for their assistance in collecting the data.

Appendix A. Models of thresholds as a function of SNR

To test whether an exponential decay function provided a better fit than a model which assumed that there were no changes in threshold as a function of delay, we first approximated the exponential decay function with a quadratic equation:

$$y = a + bx + cx^2 \tag{A1}$$

where y is the threshold and x is the word-onset delay. Next, we tested whether the quadratic model was a better predictor than a model which assumed that threshold did not change with word-onset delay,

$$y = a. \tag{A2}$$

For younger adults, in both noise and babble, and for older adults, in noise, model A1 provided a significantly better fit to the data than model A2 ($F(2,146) = 10.52$, $p < .001$; $F(2,141) = 5.97$, $p = .003$; $F(2,147) = 3.29$, $p = .04$, respectively), but not for the old-babble condition ($F(2,136) = 0.29$, $p = 0.75$). A comparison of the exponential model (Eq. (2)) with the quadratic model in these three instances showed that the exponential model provided the better fit to the data. Hence, we can conclude that an exponential decay model provides a better description of the data in these instances than a model in which thresholds do not change as a function of delay. To test whether these three instances (young-babble, old-noise and young-noise) could be fit by exponential functions that differed only in asymptote values, we again approximated the exponential function with a quadratic function to define a model in which

$$\begin{cases} y_{\text{old,babble}} = a_1 \\ y_{\text{young,babble}} = a_2 + b_2x + c_2x^2 \\ y_{\text{old,noise}} = a_3 + b_3x + c_3x^2 \\ y_{\text{young,noise}} = a_4 + b_4x + c_4x^2 \end{cases} \tag{A3}$$

Note that in this 10-parameter model, the rate of decay with word-onset delay can differ for younger adults in babble and in noise, and for older adults in noise. We then compared this model with a 6-parameter model, in which the rate of decay, but not the asymptote, was constrained to be the same in these three instances ($b_2 = b_3 = b_4$ and $c_2 = c_3 = c_4$), whereas, again, in the old-babble condition thresholds do not change with word-onset delay. Increasing the number of parameters from 6 to 10 did not result in a significant reduction in error variance ($F(4,572) = 0.19$, $p = .94$).Hence, we cannot reject a model in which: 1) for older adults in babble, the thresholds do not change with word-onset delay; 2) the rate of decay in threshold values is the same for younger adults in noise and in babble, and for older adults in noise; 3) these three conditions differ only in their asymptote value. The four functions that fit to the data of Fig. 3 are:

$$\begin{aligned} y_{\text{old,babble}} &= -11.5 \\ y_{\text{young,babble}} &= -16.2 + 1.46e^{-.0064\,x} \\ y_{\text{old,noise}} &= -2.8 + 1.46e^{-.0064\,x} \\ y_{\text{young,noise}} &= -4.3 + 1.46e^{-.0064\,x} \end{aligned} \tag{A4}$$

where y is the threshold in dB SNR and x is the word-onset delay in milliseconds.

Appendix B. Age-related and masker effects on threshold

To test for interactions between masker type and age, we compared noise minus babble thresholds for younger participants (a within-subject comparison) to noise minus babble thresholds for older participants using t-tests at each delay. The results were $M = 1.91$ dB, $t(54) = 3.34$, $p = .002$; $M = 2.65$ dB, $t(56) = 5.1$, $p < .001$; $M = 3.12$ dB, $t(51) = 5.24$, $p < .001$, $M = 3.11$ dB, $t(54) = 6.01$, $p < .001$; $M = 2.94$ dB, $t(57) = 4.88$, $p < .001$ for word-onset delays of 100, 225, 350, 600, and 1100 ms, respectively. After applying a Bonferroni correction for five tests, all were significant at $p < .01$.

To show that thresholds were lower for babble than for noise at each of the delays we conducted t-tests of the difference between noise and babble thresholds for each group at each of the delays. The values of these tests were: younger adults: $M = 12.1$ dB, $t(29) = 31.8$, $p < .001$; $M = 11.9$ dB, $t(29) = 41.89$, $p < .001$; $M = 12.0$ dB, $t(28) = 28.67$, $p < .001$; $M = 11.55$ dB, $t(29) = 31.38$, $p < .001$; $M = 12.0$ dB, $t(29) = 27.49$, $p < .001$; older adults: $M = 10.19$ dB, $t(25) = 23.85$, $p < .001$; $M = 9.26$ dB, $t(27) = 20.93$, $p < .001$; $M = 8.87$ dB, $t(23) = 21.29$, $p < .001$; $M = 8.43$ dB, $t(25) = 23.45$, $p < .001$; $M = 9.06$, $t(28) = 21.82$, $p < .001$, for delays 100, 225, 350, 600, and 1100 ms, respectively. After applying a Bonferroni correction for 10 tests, all were significant at $p < .001$.

To show that thresholds for older adults were higher than those of younger adults for both kinds of maskers at each delay, we compared the mean thresholds for younger versus those for older adults for each masker at each delay condition. The values of these tests were, babble: $M = 3.25$ dB, $t(54) = 4.79$, $p < .001$; $M = 4.57$ dB, $t(56) = 7.14$, $p < .001$; $M = 4.64$, $t(51) = 7.53$, $p < .001$; $M = 4.74$ dB, $t(54) = 7.14$, $p < .001$; $M = 4.63$ dB, $t(57) = 6.85$, $p < .001$; noise: $M = 1.34$ dB, $t(54) = 3.06$, $p = .003$; $M = 1.92$ dB, $t(56) = 4.00$, $p < .001$; $M = 1.52$ dB, $t(51) = 4.01$, $p < .001$; $M = 1.62$ dB, $t(54) = 3.66$, $p < .001$; $M = 1.69$ dB, $t(57) = 3.42$, $p = .001$, for delays 100, 225, 350, 600, and 1100 ms respectively. After applying a Bonferroni correction for 10 tests, all were significant at $p < .05$.

Appendix C. Slopes

Quadratic functions relating slope values to word-onset delay were fit to the data in all four conditions. There was no evidence that these quadratic functions provided a better fit to the data than a model in which slopes remained unchanged as a function of word-onset delay ($F(2,136) = 1.75$, $p = .18$; $F(2,147) = 0.8$, $p = .45$; $F(2,141) = 1.46$, $p = .24$; and $F(2, 146) = 0.53$, $p = .59$, for old in babble, young in babble, old in noise and young in noise, respectively).

To test for interactions between masker type and age, we compared noise minus babble slopes for younger participants (a within-subject comparison) to noise minus babble slopes for older participants using t-tests at each delay. The values of these tests were: $M = -0.0069$, $t(54) = -0.23$, $p = .82$; $M = 0.0269$ dB, $t(56) = 0.77$, $p = .45$; $M = 0.0150$, $t(51) = 0.49$, $p = .63$; $M = -0.0049$, $t(54) = -0.13$, $p = .90$; $M = 0.0822$, $t(57) = 2.43$, $p = .019$, for delays of 100, 225, 350, 600, and 1100 ms, respectively. The Bonferroni correction at $\alpha = .05$ for 5 tests requires a significance level of 0.01. Hence, there is no evidence of any interaction.

To show that slopes were higher for noise than for babble, we pooled slopes for young and old (because there was no evidence of any interaction effect) at each of the five delays, and compared the noise and babble slopes. The value of these tests were: $M = 0.0852$, $t(55) = 5.76$, $p < .001$; $M = 0.0768$, $t(57) = 4.40$, $p < .001$; $M = 0.0848$, $t(52) = 5.61$, $p < .001$; $M = 0.0887$, $t(55) = 4.62$, $p < .001$; $M = 0.0862$, $t(58) = 4.90$, $p < .001$, for delays 0, 125, 250, 500, and 1000 ms, respectively. After applying the Bonferroni correction for five tests, all were significant at $p < .001$.

To test for any age-related effects on slopes we collapsed over noise and babble conditions for younger and older adults separately at each delay, and then compared these scores between younger and older adults. The values of these tests were: $M = 0.0103$, $t(54) = 0.69$, $p = .49$; $M = 0.0167$, $t(56) = 1.079$, $p = .28$; $M = 0.0114$, $t(51) = 0.73$, $p = .47$; $M = -0.0193$, $t(54) = -0.92$, $p = .36$; $M = 0.0080$, $t(57) = 0.45$, $p = .65$, for delays 100, 225, 350, 600, and 1100 ms respectively. Hence there is no evidence of any age-related difference.

References

Alain, C., Arnott, S.R., Picton, T.W., 2001. Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. J Exp Psychol Hum 27 (5), 1072–1089.

Alain, C., Dyson, B.J., Snyder, J., 2006. Aging and the perceptual organization of sounds: a change of scene? In: Conn, M.P. (Ed.), Handbook of Models for Human Aging. Academic Press, New York, NY, pp. 759–770.

Balota, D., Yap, M., Cortese, M., Hutchison, K., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., Treiman, R., 2007. The English lexicon project. Behav Res Meth 39 (3), 445–459.

Ben-David, B.M., Schneider, B.A., 2010. A sensory origin for aging effects in the color-word Stroop task: simulating age-related changes in color-vision mimic age-related changes in Stroop. Aging Neuropsychol C 17 (6), 730–746.

Ben-David, B.M., Chambers, C.G., Daneman, M., Pichora-Fuller, M.K., Reinglod, E., Schneider, B.A., 2011a. 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., van Lieshout, P.H.H.M., Leszcz, T., 2011b. A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. Brain Inj 25 (2), 206–220.

Bernstein, J.G.W., Grant, K.W., 2009. Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. J Acoust Soc Am 125 (5), 3358–3372.

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., Rzeczkowski, C., 1984. Standardization of a test of speech perception in noise. J Speech Hear Res 27 (1), 32–48.

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

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.

Carlyon, R.P., Cusack, R., Foxton, J.M., Robertson, I.H., 2001. Effects of attention and unilateral neglect on auditory stream segregation. J Exp Psychol Hum 27 (1), 115–127.

Carlyon, R.P., Plack, C.J., Fantini, D.A., Cusack, R., 2003. Cross-modal and non-sensory influences on auditory streaming. Perception 32 (11), 1393–1402.

Cusack, R., Deeks, J., Aikman, G., Carlyon, R.P., 2004. Effects of location, frequency region, and time course of selective attention on auditory scene analysis. J Exp Psychol Hum 30 (4), 643–656.

Durlach, N.I., Mason, C.R., Shinn-Cunningham, B.G., Arbogast, T.L., Colburn, H.S., Kidd Jr., G., 2003. Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. J Acoust Soc Am 114 (1), 368–379.

Ezzatian, P., Li, L., Pichora-Fuller, K., Schneider, B.A., 2011. The effect of priming on release from informational masking is equivalent for younger and older adults. Ear Hear 32 (1), 84–96.

Ezzatian, P., Li, L., Pichora-Fuller, K., Schneider, B.A. The effect of energetic and informational masking on the time-course of stream segregation: evidence that streaming depends on vocal fine structure cues. Lang Cognitive Proc, in press.

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 Am 106 (6), 3578–3588.

George, E.L.J., Festen, J.M., Houtgast, T., 2006. Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. J Acoust Soc Am 120, 2295–2311.

Grant, D., 1948. The Latin square principle in the design and analysis of psychological experiments. Psychol Bull 45 (5), 427–442.

Hasher, L., Zacks, R.T., 1988. Working memory, comprehension, and aging: a review and a new view. In: Bower, G.H. (Ed.), 1988. The Psychology of Learning and Motivation, vol. 22. Academic Press, San Diego, CA, pp. 193–225.

Heinrich, A., Schneider, B.A., 2011. The effect of presentation level on memory performance. Ear Hear 32 (4), 524–532.

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 61 (5), 735–751.

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

Humes, L.E., Lee, J.H., Coughlin, M.P., 2006. Auditory measures of selective and divided attention in young and older adults using single-talker competition. J Acoust Soc Am 120 (5), 2926–2937.

Jerger, J., Jerger, S., 1971. Diagnostic significance of PB word functions. Arch Otolaryngol 93 (6), 573–580.

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 Hum 30 (6), 1077–1091.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B.C.J., 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci USA 103, 18866–18869.

McFadden, D., Wright, B.A., 1990. Temporal decline of masking and comodulation detection differences. J Acoust Soc Am 88 (2), 711–724.

Miller, G.A., Heise, G.A., 1950. The trill threshold. J Acoust Soc Am 22 (5), 637–638.

Murphy, D.R., Craik, F.I.M., Li, K.Z., Schneider, B.A., 2000. Comparing the effects of aging and background noise on short-term memory performance. Psychol Aging 15 (2), 323–334.

Pichora-Fuller, M.K., Schneider, B.A., Daneman, M., 1995. How young and old adults listen to and remember speech in noise. J Acoust Soc Am 97 (1), 593–608.

Raven, J.C., 1965. The Mill Hill Vocabulary Scale. H. K. Lewis, London, UK.

Salthouse, T., 1996. The processing-speed theory of adult age differences in cognition. Psychol Rev 103 (3), 403–428.

Schmeidt, R.A., 2010. The physiology of cochlear presbycusis. In: Gordon-Salant, S., Frisina, R.D., Popper, A.N., Fay, R.R. (Eds.), 2010. The Aging Auditory System: Springer Handbook of Auditory Research, vol. 34. Springer, New York, pp. 9–38.

Schneider, B.A., Li, L., Daneman, M., 2007. How competing speech interferes with speech comprehension in everyday listening situations. J Am Acad Audiol 18, 578–591.

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.), 2010. The Aging Auditory System: Springer Handbook of Auditory Research, vol. 34. Springer, New York, pp. 167–201.

Simpson, S., Cooke, M., 2005. Consonant identification in N-talker babble is a nonmonotonic function of N L. J Acoust Soc Am 118 (5), 2775–2778.

Singh, G., Pichora-Fuller, M.K., Schneider, B.A., 2008. The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. J Acoust Soc Am 124 (2), 1294–1305.

Snyder, J.S., Alain, C., Picton, T.W., 2006. Effects of attention on neuroelectric correlates of auditory stream segregation. J Cogn Neurosci 18 (1), 1–13.

Sussman, E.S., Horvath, J., Winkler, I., Mark, O., 2007. The role of attention in the formation of auditory streams. Percept Psychophys 69 (1), 136–152.

Van Engen, K., Bradlow, A., 2007. Sentence recognition in native- and foreign-language multi-talker background noise. J Acoust Soc Am 121 (1), 519–526.

Vongpaisal, T., Pichora-Fuller, M.K., 2007. Effect of age on F-0 difference limen and concurrent vowel identification. J Speech Lang Hear Res 50, 1139–1156.

Wagener, K.C., Brand, T., 2005. Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: influence of measurement procedure and masking parameters. J Audiol 44 (3), 144–156.

Wright, B.A., 1997. Detectability of simultaneously masked signals as a function of masker bandwidth and configuration for different delays. J Acoust Soc Am 101 (1), 420–429.

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.

Zwicker, E., 1965. Temporal effects in simultaneous masking by white-noise bursts. J Acoust Soc Am 37, 653–663.