AGING AND AUDITORY SCENE ANALYSIS: ARE OLDER ADULTS AS GOOD AS YOUNGER ADULTS?

Boaz M. Ben-David [1, 2, 3], Vania, Y. Y. Tse [1] and Bruce A. Schneider [1]
[1] Centre for Research on Biological Communication Systems, Department of Psychology, University of Toronto Mississauga, Canada
[2] Oral Dynamics Lab, Department of Speech-Language Pathology, University of Toronto, Canada
[3] Toronto Rehabilitation Institute, Canada.
E-mail: boaz.ben.david@utoronto.ca; Homepage: www.boazbendavid.org

Abstract

Older adults experience a greater degree of difficulty than younger adults listening to a target talker, when other people are talking. Part of this difficulty may be that it takes older adults a longer period of time to separate target speech from competing sound sources. There is reason to expect that the amount of time it takes older adults to achieve segregation will depend on the complexity of the auditory scene. This study investigates age differences in the time-line for segregating target speech from either a speech spectrum noise or a babble background (many people talking simultaneously) by varying the delay between masker onset and speech onset. Results indicate that older adults are as fast as younger adults at separating speech from a noise masker, but do not benefit to the same extent that young adults do, when the masker is babble.

Older adults typically find it more difficult to comprehend speech than younger adults when there are competing sound sources in the environment (see Schneider, Pichora-Fuller, & Daneman, 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.

The Effects of Age on the Buildup of Stream Segregation

Aging could affect the buildup of stream segregation is 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 (see Schmiedt, 2010 for a review) 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 informational masking. Informational masking of speech occurs when competing sound sources interfere with speech recognition at more central auditory and/or cognitive processing levels. For example, if age reduces a listener's ability to focus attention on the speech target and inhibit the processing of information from irrelevant sources (see Schneider, Li, & Daneman, 2007 for a review), we would expect older adults to be less proficient at stream segregation than younger adults. Moreover, stream segregation may take a longer time to emerge in older adults.

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 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 second masker. Word onset occurred 100, 225, 350, 600, or 1100 ms after masker onset. 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.

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. Finally, we wanted to examine the extent to which age-related declines in either auditory or cognitive processing might exacerbate these effects.

Method

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, corresponding to normal vocabulary levels for native English speakers (e.g., see Ben-David, & al, 2011). 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).

Stimuli and Apparatus

We used 520 bi-syllabic recorded words, spoken by a female actor with a southern Ontario accent, taken from Murphy, Craik, Li and Schneider (2000). Words were divided into ten lists of 52. To control for the impact of linguistic characteristics on word identification, lists were equated on word frequency, on the density of their phonological neighborhood and on average word duration. Spoken words were later presented on the background of 4 seconds of either continuous speech spectrum noise or multi-talker babble. 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 via headphones.

Procedure

Participants were tested individually in a single-walled sound-attenuating booth. 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. 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 the other five by speech spectrum noise. Noise and babble blocks were intermixed and counterbalanced across participants. 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 also counterbalanced across participants.

In each block, trials were presented in four different signal-to-noise ratios (SNR), 13 trials at each SNR. For younger adults, two different sets of four equidistant SNRs for babble and noise maskers were chosen: -23, -18, -13 and -8dB for babble, and -10, -6, -2 and +2 dB, for noise. For older adults, we added 3 dB to each SNR point in both sets. The use of different SNRs for younger and older adults (approximately by 3-4 dB) has been found to result in comparable word recognition scores in noise for both age groups (see Ben-David et al., 2010).

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."

Results

The scope of this talk does not permit us to discuss in full the results or the statistical analyses that support our models. We have opted to focus on results related to word-recognition thresholds as computed by fitting individual psychometric functions (see Equation 1) 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).

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

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 χ^2.

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 Figure 1, for noise maskers (top panel) and for babble maskers (bottom panel). 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 significantly with respect to the asymptotic value of the exponential decay functions (the values of a in Equation 2). The solid lines shown in Figure 1 represented the predictions of the model specified in Equation 2 below:

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

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

The Effect of Masker Type and Age on Thresholds

Figure 1 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. The extent of the difference between noise and babble thresholds (thresholds in noise minus thresholds in babble) for young adults was then compared to the extent of the babble advantage for older adults, and was found to be significant at each of the five delays. For both younger and older adults, thresholds were also significantly lower for babble than for noise at each delay. Finally, for both types of maskers, thresholds were significantly lower in younger adults than in older adults, at each of the five delays.

Discussion

The purpose of this study was to assess the role age-related differences in auditory scene analysis plays in the difficulties older adults experience in understanding speech in a noisy environment. For 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.

Data on 50% thresholds for word-identification accuracy reveal an interaction of age and type of masker. When words were presented on the background of steady-state noise, an additional 1 second of delay in word presentation resulted in ~1.5 dB of release from masking for both age-groups. Moreover, for both age-groups, the time-line for a reduction in threshold as word-onset delay increased followed the same exponential decay function (differing only in asymptotic values). In other words, the longer the exposure for the noise masker, the better older and younger listeners were able to construct a noise stream separate from the target-speech stream. Conversely, when speech was presented on the background of a multi-talker babble, older adults did not benefit from an additional 1 second delay to the same extent as younger adults did. Specifically, whereas older adults received no significant gain in their ability to segregate speech from babble, younger adults benefited from the delay to the same extent as with a noise masker. In fact, for younger adults the same exponential decay function found to describe performance with a noise masker can fit performance with babble (although asymptotic values differ).

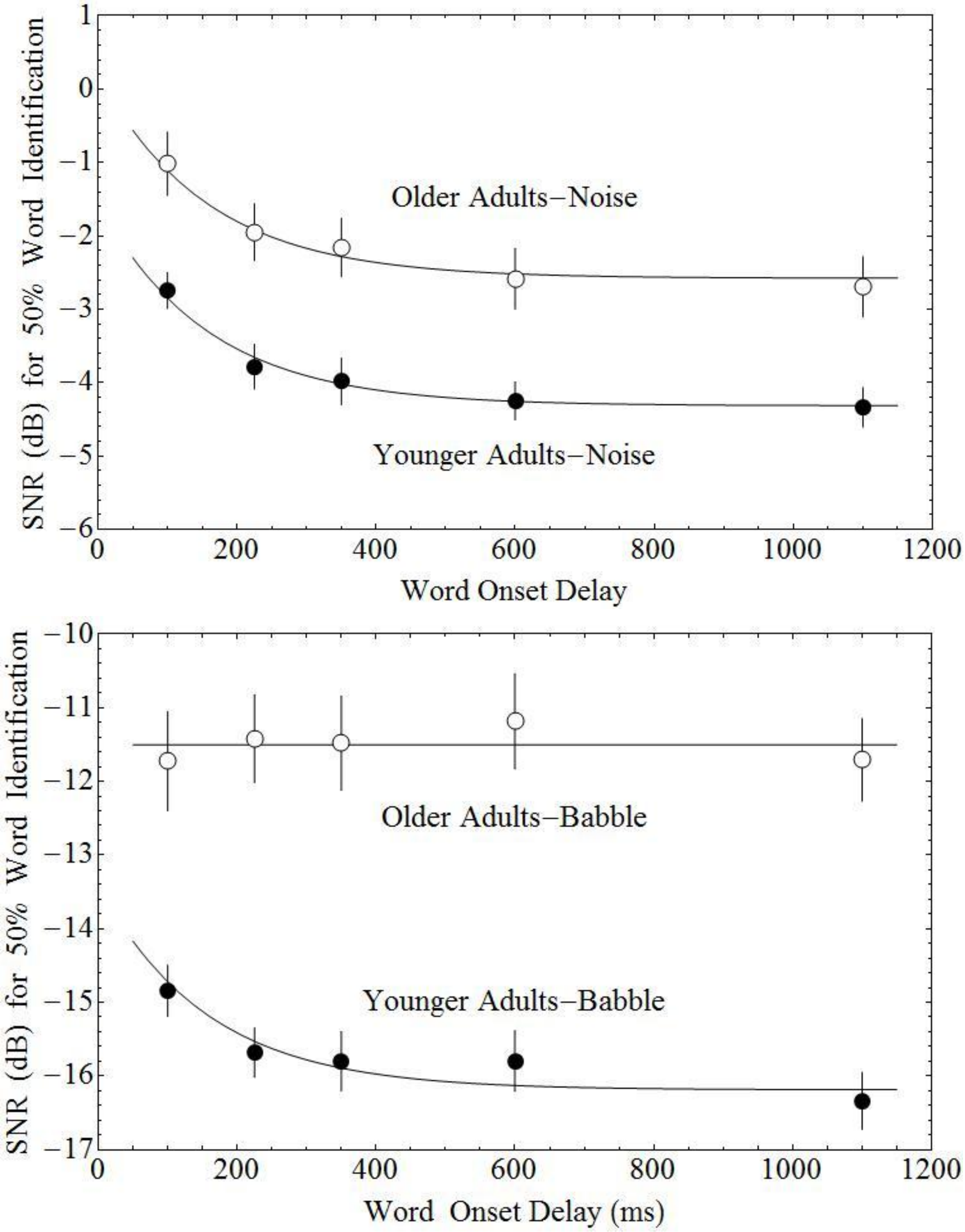

Figure 1. Average thresholds (dB SNR values for 50% correct identification) as a function of word-onset delay, based on individual psychometric functions fit to younger (solid circles) and older (open circles) adults. The top panel presents data from the noise masker condition and the bottom panel presents 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 condition, with a, varying across the three condition. Data points for old-babble condition were fit by a constant (mean threshold).

Figure 1 also shows that older adults require a higher SNR than younger adults to reach 50% identification when noise is presented 1.1 second before speech. However, the relative benefit both groups receive from this delay is the same. In other words, after tailoring presentation levels to overcome sensory degradation, older- and younger-adults behave similarly in noise. Yet on a multi-talker background, not only do older adults need a higher SNR than younger adults, they also fail to gain any benefit from an additional 1 sec of word-onset delay.

Finally, 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 1.1 second, 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. These results also suggest that informational masking presents a unique challenge for older adults, above and beyond energetic masking.

Acknowledgments

This study was partially supported by grants from the Natural Science and Engineering Research Council of Canada (RGPIN 9952-08) and from the Faculty of Arts & Science at the University of Toronto Mississauga. B. M. Ben-David was partially supported by a grant from the Ontario Neurotrauma Foundation (2008-ABI-PDF-659). We wish to thank Wu Yan (Lulu) Li, Julio Pereira, Marissa Polidori and especially Caterina Leung for their assistance.

References

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reinglod, E., & Schneider, B. A. (2011). Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research, 54*(1), 243-262. doi:10.1044/1092-4388(2010/09-0233)

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

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. *Psychology and Aging*, *15*(2), 323-334.

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

Schneider, B. A., Pichora-Fuller, M. K., & Daneman, M. (2010). Effects of senescent changes in audition and cognition on spoken language comprehension. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, & R. R. Fay (Eds.). *The aging auditory system: Springer Handbook of Auditory Research 34* (pp. 167-201). New York: Springer.

Schmiedt, R. A. (2010). The physiology of cochlear presbycusis. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, & R. R. Fay (Eds.). *The aging auditory system: Springer Handbook of Auditory Research 34* (pp. 9-38). New York: Springer.

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. *International Journal of Audiology*, *44*(3), 144–156.