Delayed Stream Segregation in Older Adults: More Than Just Informational Masking

Payam Ezzatian,[1] Liang Li,[2] Kathy Pichora-Fuller,[1] and Bruce A. Schneider[1]

Objective: To determine whether the time course for the buildup of auditory stream segregation differs between younger and older adults.

Design: Word recognition thresholds were determined for the first and last keywords in semantically anomalous but syntactically correct sentences (e.g., "A *rose* could *paint* a *fish*") when the target sentences were masked by speech-spectrum noise, 3-band vocoded speech, 16-band vocoded speech, intact and colocated speech, and intact and spatially separated speech. A significant reduction in thresholds from the first to the last keyword was interpreted as indicating that stream segregation improved with time.

Results: The buildup of stream segregation is slowed for both age groups when the masker is intact, colocated speech.

Conclusions: Older adults are more disadvantaged; for them, stream segregation is also slowed even when a speech masker is spatially separated, conveys little meaning (3-band vocoding), and vocal fine structure cues are impoverished but envelope cues remain available (16-band vocoding).

Key words: Auditory stream segregation, Auditory temporal processes, Hearing, Hearing and aging, Informational masking, Word recognition.

(Ear & Hearing 2015;36;482–484)

INTRODUCTION

The aim of the present analyses was to investigate the specific properties of maskers that may contribute to age-related difficulties when adults listen to speech in background noise. Specifically, we compared the time-course of stream segregation in older adults with previously reported results in younger adults (Ezzatian et al. 2012). Listeners were asked to repeat semantically anomalous, syntactically correct sentences (e.g., "A *rose* could *paint* a *fish*," keywords in italics) spoken by a female. Sentences were masked either by steady state speech-spectrum noise or by two competing female voices speaking different sentences of the same type (target followed masker onset by 1 second). The two-talker speech masker was tested in four conditions so that the relative contributions of amplitude envelope cues, fine structure cues (including voice F_0 and harmonics), spatial separation, and semantic meaning could be assessed (see Table 1).

The younger listeners' signal to noise (SNR) thresholds for 50% correct word recognition did not improve from the first to the last (third) keyword when the masker was steady state noise. When the masker was two-talker speech, improvement was observed if the masker was intact and colocated with the target, but not if there was perceived spatial separation between the target and masker or if the masker was vocoded using three or six frequency bands. Improvement in SNR thresholds from the first to last keyword in the sentence suggests that stream segregation (the ability to perceptually separate the target voice from the masker, see Bregman

1990) was slowed for younger adults only in the condition where the masker was intact, colocated two-talker speech.

The present analyses were conducted to determine whether the speech understanding difficulties of older adults involve slower stream segregation even in the conditions where younger adults demonstrated rapid stream segregation. Specifically, by examining how the different maskers affect the degree of improvement from first to last keyword in the two age groups, we could identify which masking situations slowed stream segregation only in older adults.

METHODS

The same analyses as had been conducted previously with data from younger adults were conducted using data from healthy older adults whose audiograms were "normal" for their age (International Organization for Standardization [ISO] 7029-2000; ISO 2000). Data were taken from four experiments in which word recognition accuracy was measured in steady state speech-spectrum noise and in a two-talker speech masking condition. A different two-talker speech masking condition was tested in each of the four experiments (see Table 1; full descriptions of the four experiments are provided in the Methods section of Ezzatian et al. 2012). For each masking condition in each experiment, word recognition accuracy was determined at four SNRs. Psychometric functions relating to the percentage of words correctly recognized to SNR and the SNR thresholds corresponding to 50% correct word recognition were determined separately for the first and last keywords.

RESULTS

Figure 1 plots the average SNR thresholds for the younger and older listeners for the first and last keywords in the steady state and two-talker masker conditions of the four experiments. In Figure 1, conditions in which there is improvement from the first to the last keyword are shown with dotted lines and conditions in which there is no improvement are shown with solid lines. When the masker was steady state noise, thresholds did not improve from the first to the last keyword for either age group. As expected, thresholds for the three conditions in which the steady state masker was colocated with the target were similar to each other, but higher than for the condition in which there was perceived spatial separation between the target and the masker. This pattern of results for the steady state masker conditions in the four experiments was similar for both younger and older adults, except that thresholds of the older adults were overall higher (about 3 dB) than those of younger adults, consistent with previous findings (Schneider et al. 2010).

When the masker was two-talker speech, thresholds were also lower for younger than for older adults in all four experiments. For both age groups, thresholds also varied depending on the two-talker masker condition. Thresholds in the condition with

[1]Department of Psychology, University of Toronto, Toronto, Ontario, Canada; and [2]Department of Psychology, Peking University, Beijing, China.

TABLE 1. Description of the two-talker masker conditions, participant characteristics, and the experiments from which data were taken

Two-Talker Speech Masker Condition	Effect of Two-Talker Speech Masker Condition on Available Cues	Sample Size, Age, and Source of Data for Younger Adults	Sample Size, Age, and Source of Data for Older Adults
1. Intact, colocated	Amplitude envelope, fine structure, and semantic content cues available	N = 16, mean age = 20.7 years (SD = 2.1), Ezzatian et al. (2012)	N = 16, mean age = 70.0 years (SD = 5.7), "no prime" conditions of experiments 2 in Ezzatian et al. (2011)
2. Three-band vocoded, colocated	Minimizes fine structure (including voice F_0 and harmonics) and semantic content but preserves some amplitude envelope fluctuations	N = 16, mean age = 21.7 years (SD = 2.64), Ezzatian et al. (2012)	N = 16, mean age = 71.5 years (SD = 4.5), "no prime" conditions of experiment 4 in Ezzatian et al. (2011).
3. Sixteen-band vocoded, colocated	Minimizes fine structure (including voice F_0 and harmonics) but preserves most semantic content and amplitude envelope fluctuations	N = 16, mean age = 20.4 years (SD = 1.67), Ezzatian et al. (2012)	N = 16, mean age = 68.2 years (SD = 4.7), new data
4. Perceived spatial separation created by the precedence effect	Preserves fine structure (including voice F_0 and harmonics), semantic and amplitude envelope fluctuations and adds a spatial separation cue	N = 16, mean age = 21.7 years (SD = 1.7), Ezzatian et al. (2012)	N = 12, mean age = 71.4 years (SD = 4.4), Li et al. (2004)

Additional details regarding audiometric thresholds and other participant characteristics can be found in the articles cited in this table. (Note that the data from one younger and one older participant in condition 1 were removed for the analyses because those participants' slopes were more than 4 SD higher than the mean slope. The data from one older participant in condition 4 was also removed from the analyses because the obtained values for this participant were more than 3 SD above the mean.)

the intact, colocated two-talker speech masker (condition 1) were poorer than those in the three other conditions (2–4). Strikingly, for younger adults, thresholds seemed to improve from the first to the last keyword only in the intact, colocated speech masker condition, but for older adults improvement was apparent in all four two-talker speech masker conditions. Whether there was a significant improvement in each two-talker speech masking condition was determined based on one-tailed t tests comparing results for the first and last keywords. For the younger adults, significant improvement was found only when the two-talker masker was intact and colocated, $t(14) = 3.58$, $p = 0.002$. For the older adults, there was a significant improvement in all the two-talker speech masker conditions: intact, $t(14) = 2.39$, $p = 0.016$; precedence, $t(10) = 7.98$, $p < 0.001$; 3-band noise vocoding, $t(15) = 1.98$, $p = 0.033$; 16-band noise vocoding, $t(15) = 2.93$, $p = 0.005$).

Clearly, both age groups improved across keywords in the sentences when speech was masked by intact colocated competing two-talker speech (Fig. 1, filled circles), but age-related differences in whether improvement was realized in the other three two-talker masking conditions warranted further examination. Because there were no statistically significant differences among the three other two-talker masking conditions for younger listeners (Ezzatian et al. 2012) or any statistically significant differences among these three conditions for older listeners in the present study,* we collapsed the results over these three conditions within each age group and conducted a two keyword position (first versus last keyword) by two masker type (colocated intact speech masker versus other speech masker types) by two age groups (younger versus older) repeated measures analysis of variance with keyword position as a within-subject factor and masker type and age group as between-subject factors. This analysis of variance revealed significant main effects of keyword position ($F[1,117] = 45.677$, $p < 0.001$), masker type ($F[1,117] = 136.659$, $p < 0.001$), and age

($F[1,117] = 32.352$, $p < 0.001$). In addition, there was a significant two-way interaction between keyword position and masker type ($F[1,117] = 4.154$, $p = 0.044$, and a significant three-way interaction for keyword position, masker type, and age ($F[1,117] = 4.669$, $p = 0.033$). Figure 2 illustrates that the three-way interaction is due to a significant improvement in thresholds from the first to the last keyword for older adults for all masker types and for younger adults only for the colocated intact speech masker type. In the colocated speech masker condition, improvement did not differ significantly between age groups.

DISCUSSION

The present analyses are consistent with prior findings, suggesting that stream segregation in both younger and older adults is equally rapid when the masker is steady state speech-spectrum noise but that stream segregation is not as rapid in older as it is in younger adults when the masker is multi-talker babble, even though the words in the background are hardly discernible (Ben-David et al. 2012). By examining how different masker types affect streaming, the present analyses extend our understanding of the specific properties of maskers that cause older adults to have more difficulty than younger adults when listening to speech in background noise. Masker-specific age-related differences in streaming may be attributed to the reduced ability of the older group to use particular types of auditory cues. Even though the audiometric thresholds of the older adults may be poorer than those of the younger adults, especially at higher frequencies, reduced audibility would not readily explain the observed age-related differences in improvement from the first to the last keyword because both the target and the masker would be similarly affected across the sentence. Rather, auditory processing of suprathreshold cues that could be used in different ways to distinguish between the target speech and the various maskers must be considered. In particular, voice similarities between the target and the competing talkers are implicated in the effect of keyword position for both younger and older adults when the target is colocated

* A two keyword position by three condition analysis of variance revealed a significant effect of keyword position ($F[1,40] = 26.55$, $p < 0.001$) but not of condition or keyword position by condition.
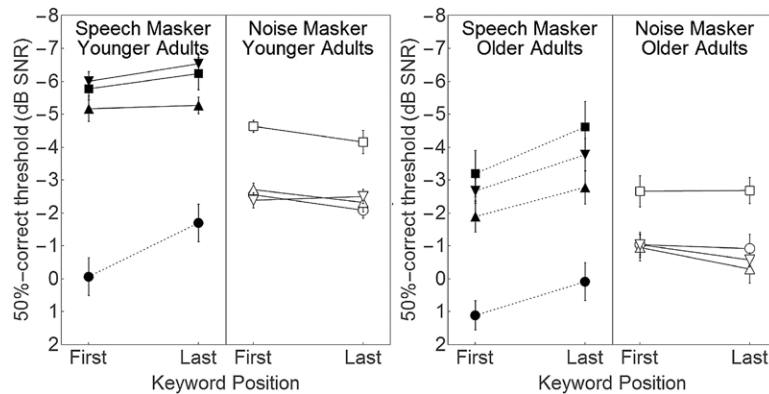
Fig. 1. Left two panels: average 50% correct word recognition thresholds in dB signal to noise (SNR) for the first and last keywords for the younger adults in the two-talker speech masking conditions of Ezzatian et al. (2012). Note that the threshold values are plotted such that a rising pattern with keyword position indicates improved performance across the sentence. Filled black circles represent data from the intact two-talker speech masker condition. Filled black upward pointing triangles represent the data from the 3-band vocoded two-talker speech masker condition. Filled black downward pointing triangles represent the data from the 16-band vocoded two-talker speech masker condition. Filled black squares represent the data from an intact two-talker speech masker condition with a spatial separation between target and masker. Unfilled symbols represent thresholds from the corresponding four experiments when the masker was steady state, speech-spectrum noise. The right two panels contain the data from the four corresponding experiments conducted on older adults. Dotted lines indicate the data from experiments in which performance significantly improved from the first to the last keyword. Solid lines indicate conditions in which there is no improvement in performance from the first to the last keyword. Error bars represent standard errors from mean performance. A portion of the data for younger adults has been re-plotted with permission from Figure 5 of Ezzatian et al. (2012).

with the intact two-talker masker. In addition, age-related reductions in auditory processing that render the target and competing speech less distinctive likely slow stream segregation by older adults in the other two-talker masking conditions. In particular, the observation of a significant effect of keyword position in both the 3-band and the 16-band vocoding conditions suggests that, unlike younger adults, similarities in the amplitude envelopes of the target and masking streams are sufficient to delay stream segregation in older adults, whether the semantic content is available (16-band vocoding) or not (3-band vocoding).

In summary, older adults have more difficulty distinguishing streams, with consequent delays in auditory stream segregation, whenever there are fluctuations in the amplitude envelopes of masker, whether or not there are voice cues in the masker, whether or not there is perceived spatial separation, and independent of whether or not the semantic contents of the masker was intelligible.

Address for correspondence: Bruce Schneider, Department of Psychology, University of Toronto, 3359 Mississauga Rd. N, Mississauga, Ontario, Canada L5L 1C6. E-mail: bruce.schneider@utoronto.ca

REFERENCES

Ben-David, B. M., Tse, V. 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.

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

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

Ezzatian, P., Li, L., Pichora-Fuller, K., Schneider, B. (2012). The effect of energetic and informational masking on the time-course of stream segregation: Evidence that streaming depends on vocal fine structure cues. *Lang Cogn Process, 27*, 1056–1088.

International Organization for Standardization. (2000). *ISO 7029-2000: Acoustics—Statistical Distribution of Hearing Thresholds as a Function of Age* (2nd ed.). Geneva, Switzerland: International Organization for Standardization.

Li, L., Daneman, M., Qi, J. G., et al. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *J Exp Psychol Hum Percept Perform, 30*, 1077–1091.

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.), *Springer Handbook of Auditory Research: The Aging Auditory System* (pp. 167–210). New York, NY: Springer.
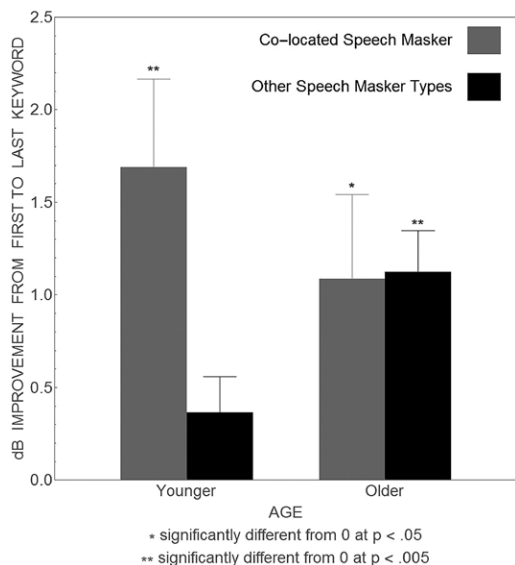
Fig. 2. Improvement (dB) in threshold from the first to the last keyword. Standard error bars are shown.