Age-Related Differences in Listening Effort During Degraded Speech Recognition

Kristina M. Ward, Jing Shen, Pamela E. Souza, and Tina M. Grieco-Calub

Objectives: The purpose of the present study was to quantify age-related differences in executive control as it relates to dual-task performance, which is thought to represent listening effort, during degraded speech recognition.

Design: Twenty-five younger adults (YA; 18–24 years) and 21 older adults (OA; 56–82 years) completed a dual-task paradigm that consisted of a primary speech recognition task and a secondary visual monitoring task. Sentence material in the primary task was either unprocessed or spectrally degraded into 8, 6, or 4 spectral channels using noise-band vocoding. Performance on the visual monitoring task was assessed by the accuracy and reaction time of participants' responses. Performance on the primary and secondary task was quantified in isolation (i.e., single task) and during the dual-task paradigm. Participants also completed a standardized psychometric measure of executive control, including attention and inhibition. Statistical analyses were implemented to evaluate changes in listeners' performance on the primary and secondary tasks (1) per condition (unprocessed vs. vocoded conditions); (2) per task (single task vs. dual task); and (3) per group (YA vs. OA).

Results: Speech recognition declined with increasing spectral degradation for both YA and OA when they performed the task in isolation or concurrently with the visual monitoring task. OA were slower and less accurate than YA on the visual monitoring task when performed in isolation, which paralleled age-related differences in standardized scores of executive control. When compared with single-task performance, OA experienced greater declines in secondary-task accuracy, but not reaction time, than YA. Furthermore, results revealed that age-related differences in executive control significantly contributed to age-related differences on the visual monitoring task during the dual-task paradigm.

Conclusions: OA experienced significantly greater declines in secondary-task accuracy during degraded speech recognition than YA. These findings are interpreted as suggesting that OA expended greater listening effort than YA, which may be partially attributed to age-related differences in executive control.

Key words: Listening effort, Dual-task paradigm, Cognitive capacity, Executive function

(Ear & Hearing 2017;38;74–84)

INTRODUCTION

To understand speech, a listener must match incoming acoustic information with their internal lexical representation. Under ideal listening conditions in listeners with normal hearing, this process is largely automatic because the high-fidelity, bottom-up representation of speech is easily matched to the long-term representations of the listener's native language. As the acoustic signal or its internal representation is degraded by signal processing, background noise, hearing loss, or a combination of these factors, there is a concomitant increase in the demand for top-down cognitive processes necessary for speech recognition (e.g., Broadbent 1958; Rabbitt 1966; Downs & Crum 1978; Rakerd et al. 1996; Sarampalis et al. 2009; Rönnberg et al., 2013). The extent to which listeners allocate cognitive resources for speech recognition has previously been referred to as *listening effort* (e.g., Hicks & Tharpe 2002; Fraser et al. 2010; Gosselin & Gagné 2010; Picou et al. 2011). It is thought that listeners experience minimal listening effort in ideal listening conditions and greater listening effort in degraded listening conditions (e.g., Gordon-Salant & Fitzgibbons 1997; Eisenberg et al. 2000; Desjardins & Doherty 2013; Pals et al. 2013). Broader views of this process have also conceptualized the allocation of resources as related to cognitive spare capacity (e.g., Rönnberg et al. 2011) and fatigue (e.g., McGarrigle et al. 2014). These concepts are an emerging source of discussion among researchers. For convenience, we will use the term listening effort here.

There has been recent interest in understanding how listening effort is related to the speech-recognition difficulties demonstrated by many older listeners. Advancing age results in declines in sensory acuity, suprathreshold sensory processing, and cognitive function, all of which interfere with speech recognition. Declines in audibility have been strongly linked to poorer speech recognition in quiet (Humes & Roberts 1990). Older adults (OA) have even more difficulty recognizing speech in degraded listening situations, including those in which the signal is degraded by competing noise (Pichora-Fuller et al. 1995; Dubno & Ahlstrom 1997; Gordon-Salant & Fitzgibbons 1997; Füllgrabe et al. 2015). It is probable that one source of such differences is a decline in suprathreshold auditory processing. For example, the ability to perceive temporal cues declines with age, even in OA with audiometrically normal hearing (e.g., Walton 2010; Füllgrabe et al. 2015). It has also been suggested that some OA have difficulty tracking dynamic spectral cues, including formant transitions (e.g., Souza et al. 2011; Schvartz-Leyzac & Chatterjee 2015). Regardless of the specific causes, existing data suggest that age-related declines in degraded speech recognition are likely to place greater demand on cognitive processing. That is, OA seem to dedicate a greater portion of their finite cognitive resources to the speech recognition task compared to younger adults (YA; Tun et al. 2009; Gosselin & Gagné 2011; Desjardins & Doherty 2013; Degeest et al. 2015). Accordingly, tasks designed to vary listening effort may offer insight into age-related performance differences.

Dual-Task Paradigms and Listening Effort

There is a growing body of work aimed at quantifying listening effort using behavioral, physiological, and subjective measures (see McGarrigle et al. 2014 and Francis & Füllgrabe 2015 for a review). A commonly used behavioral methodology, the dual-task paradigm, is contingent on the assumption of delimited cognitive capacity (Kahneman 1973). As conceptualized by Kahneman (1973), *cognitive capacity* refers to the finite pool of shared cognitive resources that a listener can use to perform

---

The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois.

0196/0202/2017/381-0074/0 • Ear & Hearing • Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved • Printed in the U.S.A.

various tasks across modalities. When degraded speech recognition demands greater cognitive effort, the Kahneman model predicts that fewer resources are available in the *cognitive spare capacity* to maintain simultaneous cognitive processes. In dual-task paradigms designed to quantify listening effort, participants perform two tasks—a primary speech recognition task and a secondary task—in isolation and then simultaneously. When performing both tasks simultaneously, listeners are instructed to prioritize the primary task over the secondary task. In this scenario, the primary task occupies a portion of cognitive resources, and therefore, performance on the secondary task is dependent on cognitive spare capacity. While the primary and secondary tasks that comprise the dual-task paradigm dictate the domain-specific cognitive processes involved, the sharing of attention between the two tasks necessitates the involvement of executive control. Specifically, as the primary task becomes more difficult, cognitive resources, including attention, are reallocated to the primary task to maintain performance. This leaves a reduced cognitive spare capacity available to maintain performance on the secondary task. If a listener's cognitive spare capacity does not contain sufficient resources to meet the cognitive demand imposed by the secondary task, performance on that task will decline. Thus, tracking listeners' cognitive spare capacity with changing demands of speech recognition may serve as an indicator of listening effort.

Previous studies have used dual-task paradigms as a surrogate for listening effort in various listener populations and under a variety of conditions of degraded auditory input. For instance, Sarampalis et al. (2009) found that young adults with normal hearing experienced declines in performance on a visual task, indicating greater listening effort, in the presence of background noise. The authors argued that this result is consistent with the idea that listeners allocated a relative portion of their limited cognitive resources to the speech recognition task in noise and, consistent with Kahneman's theory of delimited cognitive capacity, had insufficient cognitive spare capacity to maintain performance on the visual task. Similarly, Pals et al. (2013) showed that performance on two separate visual tasks—each taxing different aspects of working memory—declined in young adults with normal hearing as the speech signal became more spectrally degraded. Finally, Downs (1982) demonstrated that adult listeners with hearing loss exerted less effort, as indicated by faster responses to a visual probe, when performing a simultaneous speech discrimination task with hearing aids as opposed to unaided. Considered together, these studies provide support for the hypothesis that listeners' cognitive spare capacity—and thus secondary-task performance—decreases as the demands for speech recognition increase.

Listening Effort Across the Life Span

Previous studies have used dual-task paradigms to evaluate age-related differences in listening effort. This work has suggested that OA, regardless of hearing sensitivity, have significantly greater declines in secondary-task performance (i.e., experience greater listening effort) during sentence recognition in background noise than YA (Gosselin & Gagné 2011; Desjardins & Doherty 2013). Gosselin and Gagné (2011) found that OA were less accurate and slower on a secondary tactile task than YA and proposed that these secondary-task declines resulted from of a combination of age-related differences in sensory and cognitive function. Similarly, Desjardins and Doherty (2013) showed that OA had greater declines on a secondary visual–motor task and proposed that this decline was related to age-related differences in working memory and processing speed. Given that OA consistently perform more poorly than YA in the presence of background noise, these findings are not surprising. Rather, they affirm the idea that degraded speech recognition has a greater effect on concurrent cognitive processing in OA.

There is an open question, however, as to whether previous results are specific to speech recognition in noise or whether they can be generalized to other forms of degraded speech. The present study aimed to address this issue by investigating differences in dual-task performance, and thus listening effort, between YA and OA during recognition of speech that is noise-band vocoded. Noise-band vocoding spectrally degrades the speech signal without adding another *object* to attend, which is the case when there is acoustic competition (e.g., background noise). Background noise, a form of acoustic competition to the target speech, draws an individual's cognitive resources (e.g., attention), particularly if the competing sounds contain relevant information like speech (e.g., Durlach et al. 2003). Interference with cognitive processes may be more problematic for OA (e.g., Helfer & Freyman 2008). The use of noise-band vocoded speech eliminates this potential confound while still providing a way to systematically degrade the fidelity of the speech input.

The Present Study

The purpose of the present study was to explore differences in dual-task performance in YA and OA during degraded speech recognition. Participants completed a dual-task paradigm consisting of a speech recognition (primary) task and a visual monitoring (secondary) task. By incorporating a primary and secondary task that use different perceptual modalities (i.e., an auditory/verbal primary task and a visual/motor secondary task), the dual-task paradigm used in the present study assumes a global (i.e., not modality specific) cognitive capacity including executive control resources (i.e., attention) that can be reallocated to perform tasks across modalities as cognitive demand increases. The central hypotheses were that secondary-task performance during degraded speech recognition is dependent on (1) the fidelity of the speech input and (2) participants' hearing acuity and executive control, which are factors that are often correlated to listening effort and are expected to change across the lifespan. Based on these hypotheses, we predicted that OA would exhibit greater declines in secondary-task performance and therefore greater listening effort, than YA as spectral degradation increased.

METHODS

Participants

Twenty-five YA (18–24 years, mean 21.5 years, SD = 2.5 years) and 21 OA (56–82 years, mean 66.5 years, SD = 6.3 years) participated in the present study. Two additional participants, 1 YA and 1 OA, were excluded from data analyses due to prior experience with the speech material and difficulty completing study tasks, respectively. While the majority of the YA group consisted of undergraduate and graduate students attending Northwestern University, the OA group comprised people from within the community of Northwestern University as well as from the greater Chicagoland area. The average number of years of education was 19.2 years (SD = 1.9 years) for the YA group and 21.1 years (SD = 2.4 years) for the OA group. All participants were native English speakers, had normal or corrected-to-normal visual acuity as tested using a Snellen eye chart, and had typical hearing for their age. For the YA group, criteria for normal hearing was pure-tone thresholds of ≤20 dB HL at octave frequencies from 0.25 kHz through 8 kHz and the interoctave frequency of 6 kHz. Audiometric criteria for the OA group were motivated by Cruickshanks et al. (1998) and consisted of pure-tone thresholds of ≤25 dB HL at octave frequencies from 0.25 kHz to 2 kHz and ≤45 dB HL at 4 kHz to 8 kHz in each ear (Fig. 1). All OA in the present study had thresholds that were either better than or within +1 SD of the mean age-specific hearing thresholds reported in Cruickshanks et al. (1998). Testing occurred within a single session that lasted a maximum of 2 hours, including breaks. Participants were compensated at an hourly rate for their time. Before data collection, approval for all study procedures was obtained from the Institutional Review Board at Northwestern University, and all participants completed an informed consent process.

Stimuli

Speech Stimuli • The speech stimuli consisted of a corpus of 180 Bamford–Kowal–Bench phonetically balanced short sentences (e.g., *They are looking at the clock.*) produced by a male speaker (Bench et al. 1979; Etymotic Research, Elk Grove Village, IL, USA). Sentences were noise-band vocoded using custom MATLAB software to generate 4 conditions that varied in spectral degradation (Souza & Rosen 2009). During vocoding, each sentence was root-mean-square equalized at 65 dB SPL and digitally filtered into the respective number of frequency channels (i.e., 4-, 6-, or 8-ch). The output of each filtered waveform was half-wave rectified and low-pass filtered at 30 Hz using a fourth-order Butterworth filter with a slope of −24 dB per octave to extract the amplitude envelope of the signal. The envelope was then multiplied by a broadband noise carrier and passed through its respective band-pass filter. This process stripped the speech stimuli of its original temporal fine structure while maintaining the slow-changing temporal features of the speech waveform. In addition to the 3 noise-band-vocoded conditions, testing incorporated an unprocessed condition to establish a baseline measure of speech recognition.

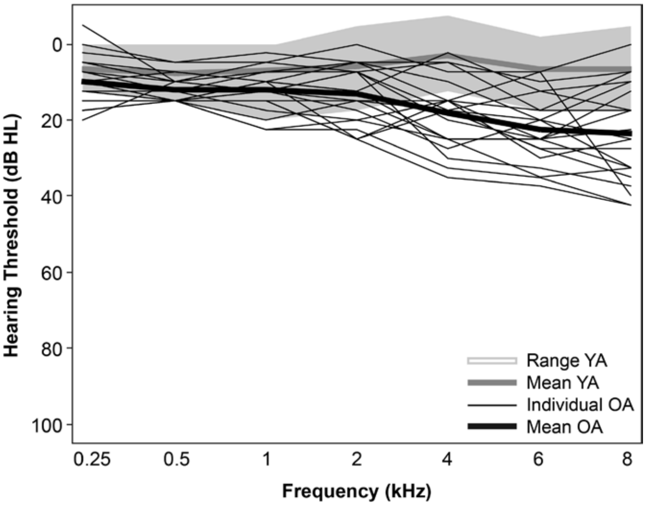

Figure 1. Pure-tone thresholds (in dB HL) averaged across the right and left ears for younger adults (YA; range shown as gray shaded region) and older adult (OA; individual data indicated by thin black lines). Average thresholds for YA (thick gray line) and OA (thick black line) are also shown.

Visual Stimuli • The visual stimuli were 260 individual grayscale illustrations (281 by 197 pixels) of familiar animate and inanimate objects (e.g., tiger, balloon) adapted from Snodgrass & Vanderwart (1980). The images were scaled to 3″ by 4″ and individually displayed on a computer monitor using custom software adapted from Wright et al. (2014) and modified with MATLAB PsychToolbox (Brainard 1997; Pelli 1997; Kleiner et al. 2007).

Testing Apparatus

Testing was conducted in a sound-attenuating booth with only the participant present. Participants sat at a desk facing a Dell Ultrasharp U2413 24″ monitor attached to an Apple wired USB keyboard. Stimuli were presented from a Macbook Pro laptop positioned externally to the sound booth. Speech stimuli were directly routed from the laptop to supra-aural Sennheiser HD 25-SP headphones worn by the participant, and visual stimuli were displayed on the Dell computer monitor within the sound booth. For all visual tasks, participants indicated their response by pressing a designated key on the keyboard.

Executive Control Assessment

The NIH Toolbox Flanker Inhibitory Control and Attention Test (McDonald 2014) was used in the present study to assess participants' selective attention and inhibitory control. This standardized psychometric measure of executive control has been normed on individuals from 3 to 85 years old. In this task, a row of 5 arrows was displayed on the screen and the participant had to press either the leftward or rightward arrow key as quickly as possible to indicate the directionality of the middle arrow. Test trials consisted of intermixed congruent (i.e., all arrows facing the same direction) and incongruent (i.e., middle arrow facing the opposite direction) trials for a total of 20 trials. Unadjusted scale scores were derived for each participant by combining the reaction time scores (in milliseconds) from the trials on which they correctly responded with the accuracy of each response. The age-adjusted scores were generated by comparing participants' scores to those of a nationally representative normative sample within the same age range (Beaumont et al. 2013). The normal distribution associated with these scores has a mean of 100 and a standard deviation of 15. For both unadjusted and adjusted Flanker scores, higher values indicate better performance and a greater overall level of functioning.

Procedures

Speech Recognition (Primary Task) • Sentences were presented binaurally at 65 dB SPL with a fixed intersentence silent interval of 5 sec. Participants were instructed to listen to each sentence and repeat it aloud as accurately as possible, even if it required guessing. Responses were scored live with a point awarded for each of the 3 or 4 predetermined keywords correctly repeated for each sentence. Before testing, each subject was familiarized to a preset list of 20 sentences that were distinct from the test lists. The familiarization list contained 5 sentences from each of the 4 listening conditions, with spectral degradation increasing as the list progressed. If participants' performance fell below the criterion score of 50% keywords correctly identified in the most degraded (i.e., 4-ch) condition, they received an additional training block. This block consisted of the 20 sentences used during familiarization presented in the 4-ch condition. Feedback was only provided for the first 10 sentences; participants' performance on the final 10 sentences was used to ensure that they met the 50% correct criterion before progressing to the test trials.

After familiarization, participants were presented with one list of 20 sentences in each of the 4 listening conditions (i.e., unprocessed, 4-, 6-, 8-ch noise-band vocoded) for a total of 80 sentences. Sentence lists were counterbalanced for the condition in which they were presented, and the order of test conditions was randomized across subjects. Within each test condition, speech recognition accuracy was calculated as the proportion of keywords correctly identified out of 62 possible keywords. Participants completed the speech recognition task in each condition first in isolation and then as part of the dual task.

Visual Monitoring (Secondary Task) • During the visual monitoring task, grayscale images were presented sequentially on a computer monitor at a rate of 300 msec per image with an interstimulus interval of equivalent duration. Participants were instructed to press a key as quickly as possible when the same picture occurred twice in a row. Performance was based on reaction time, defined as the duration of time between the onset of the display of the repeated image and the participant's key press, and accuracy, defined as the proportion of trials in which a repeated image elicited a key press. Reaction times were only extracted from correct responses, or key presses that occurred in response to a repeated image. The majority of participants (36 out of 46) performed the visual monitoring task in isolation at two points during the study—immediately before and after the dual-task conditions—to investigate the presence of a learning or fatigue effect over the course of the experiment. Data from 8 YA and 2 OA were collected before the implementation of the second administration of the visual monitoring task in isolation after the dual-task conditions. In addition to when performed in isolation, the visual task was performed concurrently with the speech recognition task throughout the dual-task conditions.

Dual-Task Paradigm • In addition to the primary and secondary task baseline measurements, participants performed these tasks simultaneously as part of the dual-task paradigm. Participants were instructed to listen to each sentence and repeat it aloud as accurately as possible—the designated "main goal"—while simultaneously monitoring the visual stream for repeating images to the best of their ability. Similar to the speech recognition task in isolation, the dual-task paradigm consisted of a total of 80 sentences presented across 20-sentence lists, 1 list per listening condition. Similar to the single-task condition, sentence lists and conditions were counterbalanced for order of presentation. Although sentence lists stemmed from the same corpus, speech material was not redundant between the single and dual-task conditions. Throughout the duration of each 20-sentence list, participants performed the visual monitoring task in which they were presented with approximately 206 images with repeated images occurring about 10% of the time. The visual monitoring task was administered continuously throughout the dual-task conditions, such that repeated pictures could occur within any of the following three contexts relative to the speech recognition task: (1) during the auditory presentation of the target sentence; (2) during the participant's verbal recall of the target sentence and (3) during the remaining period of silence within the 5-sec intersentence interval. Although there were brief periods of time (i.e., approximately 1–2 sec) following the verbal recall of each sentence when participants were only actively performing the visual monitoring task, it is expected that the demand for shared attentional resources was consistently maintained due to anticipation of the subsequent sentence. Performance scores for the speech recognition task (i.e., accuracy) and visual monitoring task (i.e., accuracy and reaction time) were recorded for each listening condition.

RESULTS

Hearing Sensitivity

To assess whether there are group differences in hearing sensitivity, pure-tone audiometric thresholds of YA and OA (Fig. 1) were compared using a 2 (age group: YA, OA) × 7 (frequency: 0.25, 0.5, 1, 2, 4, 6, 8 kHz) mixed analysis of variance (ANOVA) with age group as the between-subjects factor. This analysis revealed a main effect of frequency (*F*[3.39, 149.25] = 11.55, $p < 0.0001$, $\eta^2_p = 0.21$, following Greenhouse–Geisser correction for violation in sphericity), suggesting that the hearing sensitivity of YA and OA changed across frequency. There was also a significant main effect of age group (*F*[1, 44] = 49.91, $p < 0.0001$, $\eta^2_p = 0.53$), indicating that there was a difference in hearing sensitivity between YA and OA. In addition, there was a significant age group-by-frequency interaction (*F*[3.39, 149.25] = 15.65, $p < 0.0001$, $\eta^2_p = 0.26$, following Greenhouse–Geisser correction for violation in sphericity), suggesting that the differences in pure-tone audiometric thresholds between YA and OA vary by frequency. The average difference in thresholds between YA and OA at each frequency tested was (in dB HL) 3.3 (0.25 kHz), 4 (0.5 kHz), 4.8 (1 kHz), 7.4 (2 kHz), 15.2 (4 kHz), 15.7 (6 kHz), and 17.1 (8 kHz). Post hoc paired comparisons using a Bonferroni-adjusted critical alpha level of 0.007 showed that these differences were statistically significant for 0.5, 2, 4, 6, and 8 kHz. These results indicate that although all participants demonstrated age-typical hearing, there were differences in hearing sensitivity between groups.

Executive Control

A one-way ANOVA revealed that the average unadjusted Flanker score for YA (122.7 ± 10.7, mean ± SD) was significantly greater than that for OA (102.8 ± 11.3; *F*[1, 45] = 37.53, $p < 0.001$, $\eta^2 = 0.46$), confirming age-related differences between the two groups. This age-related difference persisted when comparing age-adjusted Flanker scores (YA = 115.6 ± 13.5; OA = 104.5 ± 13.8; *F*[1,45] = 7.56, $p < 0.01$, $\eta^2 = 0.147$). Based on the NIH Toolbox performance measure standard score mean of 100 and standard deviation of 15, these data suggest that executive control abilities are age-typical for OA but are, on average, greater than one standard deviation above the mean for YA. Overall, this indicates better executive control for the YA group as compared to the OA group.

Speech Recognition (Primary Task)

The average percentage of keywords correctly identified in each of the 4 listening conditions when performed alone (i.e., single task) and simultaneously with the visual monitoring task (i.e., dual task) for YA and OA are displayed in Figure 2. Speech

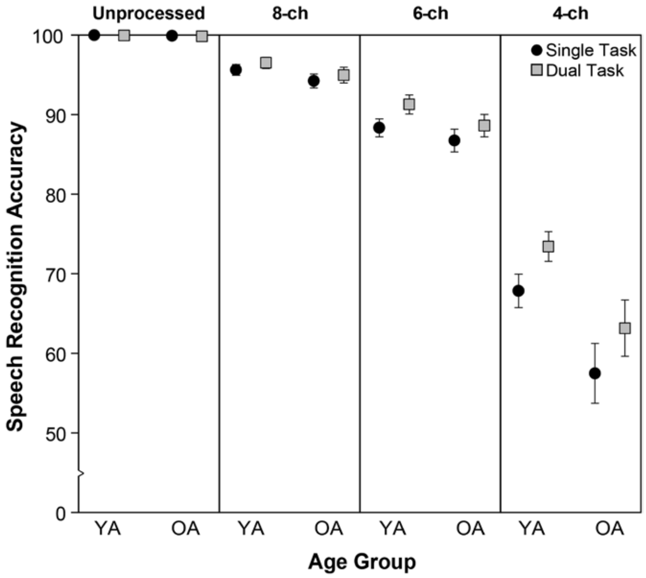

Figure 2. Mean single- and dual-task speech recognition accuracy (% correct) for younger adults (YA) and older adults (OA) across primary-task conditions (unprocessed, 8-ch, 6-ch, and 4-ch). Error bars represent ±1 SE from the mean.

recognition accuracy data were folded-square-root transformed before statistical analysis to allow for closer examination of ceiling effects without requiring adjustment of extreme data values (i.e., 0%, 100%; Tukey 1960).

The transformed speech recognition scores were submitted to a 2 (age group: YA, OA) × 2 (task: single, dual) × 4 (condition: unprocessed, 8-ch, 6-ch, 4-ch) mixed ANOVA, with task and condition as the within-subjects factors and age group as the between-subjects factor. As predicted, there was a main effect of condition ($F[1.85, 81.49] = 522.76$, $p < 0.0001$, $\eta^2_p = 0.92$, following Greenhouse–Geisser correction for violation in sphericity), suggesting that speech recognition decreased with increasing spectral degradation. Post hoc paired comparisons using a Bonferroni-corrected critical alpha level of 0.0042 revealed significantly poorer speech recognition accuracy with each increase in spectral degradation for both YA and OA that was observed in both the single- and dual-task conditions (i.e., a total of 12 comparison, ps < 0.0001). Additionally, there was a main effect of age group ($F[1, 44] = 5.46$, $p < 0.05$, $\eta^2_p = 0.11$) reflecting better speech recognition by YA as compared to OA across conditions. Post hoc paired comparisons revealed that YA and OA had significantly different speech recognition accuracy only in the 4-ch condition during the dual-task paradigm ($p = 0.05$). This difference failed to remain significant after correcting for multiple comparisons (Bonferroni-adjusted critical alpha = 0.00625). All other paired comparisons between the YA and OA within each condition were not statistically significant for single-task conditions ($p = 0.33$ [unprocessed], 0.36 [8-ch], 0.85 [6-ch], 0.13 [4-ch]) or dual-task conditions ($p = 0.33$ [unprocessed], 0.23 [8-ch], 0.42 [6-ch]). There was also a significant age group-by-condition interaction ($F[1.85, 81.49] = 4.73$, $p < 0.05$, $\eta^2_p = 0.09$, following Greenhouse–Geisser correction for violation in sphericity), which was likely driven by group differences in the 4-ch condition. Taken together, these results indicate statistically similar speech recognition accuracy for YA and OA across conditions. This result in the presence of a significant main effect of group and a group-by-condition interaction is likely due to the greater statistical power observed when performing the ANOVA, thus allowing for increased sensitivity to detect smaller between-group variability in listening effort across conditions. Finally, there was a significant main effect of task ($F[1, 44] = 25.56$, $p < 0.001$, $\eta^2_p = 0.37$), as well as a significant task-by-condition interaction ($F[2.30, 101.05] = 5.37$, $p < 0.01$, $\eta^2_p = 0.11$). These results suggest that both YA and OA correctly identified more keywords during the dual-task condition as opposed to the single-task condition and that the magnitude of the difference between single- and dual-task conditions was not consistent with changing spectral fidelity of the speech input. Post hoc paired comparisons using a Bonferroni-adjusted critical alpha level of 0.00625 revealed better speech recognition accuracy during the dual task than the single task for YA in both the 6-ch ($p < 0.001$) and 4-ch conditions ($p = 0.002$). YA's speech recognition did not differ between single and dual tasks for the unprocessed or 8-ch conditions ($p = 0.33$ and 0.25, respectively). Post hoc comparisons also failed to reveal differences in OA's single- and dual-task performance in the unprocessed, 8-ch, 6-ch, or 4-ch conditions ($p = 0.6$, 0.14, 0.11, and 0.04, respectively). This finding suggests that YA may have been able to improve their recognition in the 6-ch and 4-ch conditions with increased exposure, akin to perceptual learning (e.g., Shannon et al. 1995; Davis et al. 2005), given that dual-task conditions always followed single-task conditions. Despite this potential training effect, the fact that performance on the speech recognition task was better in the dual-task conditions, in both groups, confirms that participants were prioritizing it as the primary task, as instructed.

Visual Monitoring (Secondary Task)

Paired *t* tests using a Bonferroni-adjusted critical alpha of 0.025 were conducted to test whether there was a learning or fatigue effect on secondary-task performance. Results failed to reveal differences in accuracy (% correct, folded-square-root transformed before analysis) on the visual monitoring task in isolation when performed before and after the dual-task conditions for YA ($p = 0.15$) or OA ($p = 0.29$). There were, however, significant differences in reaction time (in milliseconds) on the visual monitoring task in isolation when performed before and after the dual-task conditions for YA (before = 500 ms ± 57 ms [mean ± SD] vs. after = 538 ms ± 60 ms, $t[1,16] = -4.05$, $p = 0.001$) but not OA (before = 588 ms ± 114 ms vs. after = 577 ± 65, $t[1, 18] = 0.7$, $p = 0.49$). This suggests a small, but consistent slowing in performance in the YA over the course of the study. To provide a better estimate of participants' baseline accuracy and reaction time on the visual monitoring task, these scores were calculated as the average performance across the two time points for the 17 YA and 19 OA who completed both administrations of the visual task in isolation. Table 1 shows the average accuracy and reaction time of responses on the visual monitoring task when performed in isolation and concurrently with the speech recognition task across each of the 4 listening conditions for YA and OA. Accuracy data were folded-square-root transformed before statistical analysis. Separate one-way ANOVAs with age group as a between-subjects factor revealed a significant main effect of group for both accuracy ($F[1, 45] = 7.13$, $p < 0.05$, $\eta^2 = 0.14$) and reaction time ($F[1, 45] = 11.78$, $p < 0.01$,

TABLE 1. Performance on the visual monitoring task for YA and OA across conditions.

	Visual Monitoring Task							
	Accuracy (% correct)				Reaction Time (ms)			
	Single Task		Dual Task		Single Task		Dual Task	
Condition	M	SD	M	SD	M	SD	M	SD
Younger	93.8	5.6			516.6	55.0		
Unpr.	—	—	93.5	9.8	—	—	522.8	49.4
8-ch	—	—	92.7	8.2	—	—	533.2	56.8
6-ch	—	—	88.4	10.0	—	—	537.1	47.2
4-ch	—	—	81.9	12.9	—	—	557.4	65.4
Older	85.0	15.2			586.6	82.5		
Unpr.	—	—	83.4	20.1	—	—	591.2	55.6
8-ch	—	—	77.2	18.1	—	—	584.1	66.0
6-ch	—	—	74.9	21.8	—	—	627.7	118.0
4-ch	—	—	59.4	21.3	—	—	603.8	70.8

OA indicates older adults; YA, younger adults.

$\eta^2 = 0.21$) on the visual monitoring task when performed in isolation. These results suggest that YA were significantly faster and more accurate than OA on the visual monitoring task when performed in isolation.

As the secondary task is used to measure changes in cognitive spare capacity during dual-task speech recognition, performance on the visual monitoring task when performed in isolation should coincide with baseline cognitive differences between YA and OA. Pearson correlations were used to determine the strength of the relation between participants' performance on the Flanker test of executive control and their single-task performance on the visual monitoring task. Results revealed that Flanker task performance was significantly positively correlated ($r = 0.36$, $p = 0.015$) with participants' accuracy scores on the visual monitoring task and significantly negatively correlated ($r = -0.36$, $p = 0.014$) with participants' reaction time scores on the visual monitoring task (Fig. 3). Taken together, these results suggest that YA have greater baseline executive control than OA and that performance on the visual monitoring task reflects, though weakly, these age-related differences.

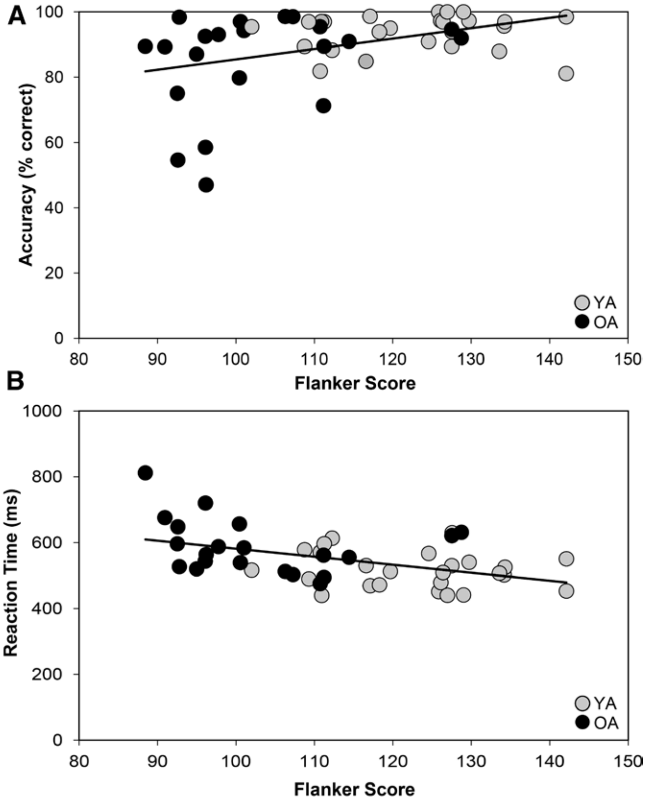

Figure 3. Participants' accuracy (A) and reaction time (B) on the visual monitoring task when performed in isolation as a function of unadjusted Flanker score. Higher Flanker scores reflect better performance. Flanker score was found to significantly correlate with both accuracy ($r = 0.36$, $p = 0.015$) and reaction time ($r = -0.36$, $p = 0.014$). OA indicates older adults; YA, younger adults.

Visual Monitoring (Single vs. Dual Task)

A prediction for the present study is that OA will show greater declines in secondary-task performance during degraded speech recognition and that these results will be interpreted as increased listening effort. To quantify these declines, and thus estimate listening effort, participants' single-task (i.e., baseline) performance on the visual monitoring task was compared to performance during the dual-task conditions and normalized to their baseline performance (i.e., [{Single Task − Dual Task}/Single Task] × 100; Somberg & Salthouse, 1982). The advantage of using this formula to quantify changes in secondary task performance is that the formula accounts for individual differences in secondary-task performance at baseline (Kemper et al. 2009; Gosselin & Gagné 2011; Desjardins & Doherty 2013).

In the present study, listening effort was calculated separately from participants' accuracy and reaction time on the visual monitoring task. Capturing both components of participants' performance on the visual monitoring task provides an opportunity to observe subtle differences in response behavior, which has been shown to change with age (e.g., Gordon-Salant 1986). As spectral degradation increases, it is expected that participants will respond less accurately and more slowly on the visual monitoring task because a relative portion of their limited cognitive resources are allocated to the speech recognition task. Thus, it is predicted that listening effort will increase as speech recognition accuracy declines, with greater spectral degradation

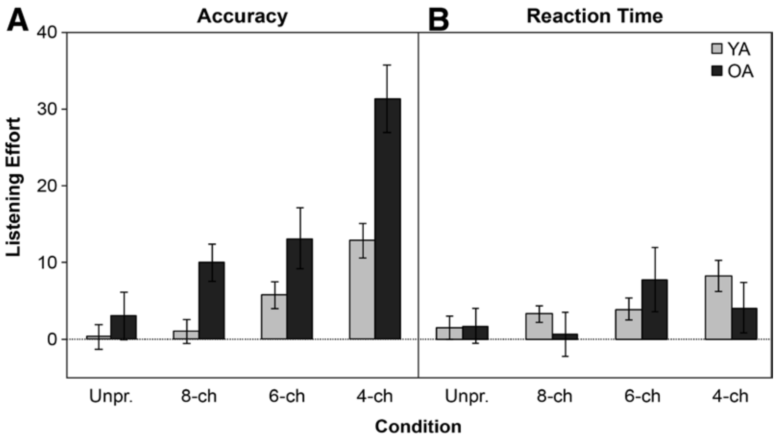

Figure 4. Mean listening effort scores as measured by accuracy (A) and reaction time (B) for younger adults (YA) and older adults (OA) across conditions. Error bars represent ±1 SE from the mean. Reaction time data have been transformed to allow for display along the same *y*-axis as accuracy.

for both YA and OA. Alternatively, participants may attempt to preserve one aspect of performance (accuracy or reaction time) while sacrificing the other as the speech input becomes more degraded (Brouwer et al. 1991). The following analysis will probe these possibilities.

Accuracy and Reaction Time • To facilitate the comparison of listening effort scores for accuracy and reaction time along the same scale (i.e., higher values indicating greater listening effort), the scores derived from changes in reaction time were each subtracted from zero, effectively inverting the sign (Fig. 4). Accuracy and reaction time data were subjected to independent 2 (age group: YA, OA) × 4 (condition: unprocessed, 8-ch, 6-ch, 4-ch) mixed ANOVAs with age group as a between-subjects factor to investigate the effect of age group on listening effort across conditions. There was a main effect of condition for both listening effort measured by accuracy ($F[2.68, 117.79] = 27.71$, $p < 0.001$, $\eta^2_p = 0.39$) and reaction time ($F[2.28, 100.37] = 3.40$, $p < 0.05$, $\eta^2_p = 0.07$), suggesting that participants showed greater declines in secondary-task performance as spectral degradation increased. There was also a significant main effect of age for listening effort measured from changes in secondary-task accuracy ($F[1, 44] = 14.47$, $p < 0.001$, $\eta^2 = 0.25$), but not secondary-task reaction time ($F[1, 44] = 0.07$, $p = 0.79$, $\eta^2 = 0.002$). For listening effort measured from accuracy on the visual monitoring task, there was a significant age group-by-condition interaction ($F[2.68, 117.79] = 3.88$, $p < 0.05$, $\eta^2 = 0.08$), suggesting that performance on the secondary task declined differentially for YA and OA as spectral degradation increased. Post hoc paired comparisons using a Bonferroni-adjusted critical alpha level of 0.0125 failed to reveal significant differences in listening effort between YA and OA in the unprocessed ($p = 0.59$) or 6-ch ($p = 0.13$) conditions. However, there were significant between-group differences in the 8-ch ($p = 0.004$) and 4-ch conditions ($p = 0.003$), indicating that, in these conditions, OA had greater declines in secondary-task accuracy than YA during degraded speech recognition.

Hierarchical Multiple Regression • A secondary objective of the present study was to explain individual differences in listening effort, as quantified by changes in secondary-task performance. The observed differences in listening effort and Flanker score between YA and OA motivate the question of whether age-related differences in executive control account for these findings. Hierarchical multiple regression analyses were conducted to probe this relation. Listening effort in each condition, as measured by changes in accuracy on the visual monitoring task, served as the dependent variable for each regression analysis. Based on the finding that there are differences in executive control between YA and OA, Flanker score was the first predictor variable entered into each regression analysis. To examine the contribution of differences in hearing sensitivity between YA and OA, the pure-tone average (PTA; calculated from thresholds at 0.5, 1, 2, 4 kHz and averaged across ears) was added to each analysis as a secondary predictor variable. Finally, to account for other possible age-related factors not captured by the other predictor variables, such as working memory and suprathreshold auditory processing, age was entered as a tertiary predictor variable into each analysis. This resulted in three successive linear regression models tested for each condition separately: (1) Flanker score; (2) Flanker score, PTA; (3) Flanker score, PTA, age.

The results of the regression analyses are displayed in Table 2. Each row displays information regarding the overall fit of the current model as well as the individual contribution of the predictor variable added at each step. The first data column displays the value of R^2, which represents the cumulative proportion of variance explained by the variable(s) in each model while the change in R^2, displayed in the second data column, demonstrates the individual contribution of the predictor variable added at each step of the analysis. The third data column shows whether the additional variance accounted for (i.e., the change in R^2) by the added predictor variable was significant. Observed listening effort in the unprocessed condition was minimal for both YA and OA, and no group differences were observed; therefore, it is reasonable that none of the factors in the model made a significant contribution. For each of the degraded conditions, performance on the Flanker task accounted for a significant amount of the variance in listening effort, ranging from 17% to 19%. As demonstrated by the change in R^2, the model containing PTA accounted for an additional 2% of variance in the 8-ch condition and approximately 1% of variance in the 6-ch and 4-ch conditions, although none of these contributions were significant. Similarly, including age as a predictor variable accounted for a nonsignificant 2% of additional variance in listening effort in the 8-ch condition and made no additional contribution in the 6-ch condition. The only condition in which a predictor variable contributed significant variance over and above Flanker score was the 4-ch condition in which age accounted for an additional

TABLE 2. Hierarchical multiple regression

	Model Statistics			Individual Predictor Statistics			
Predictor	R^2	R^2 Change	Sig. R^2 Change*	*b*	SE	*β*†	*p*‡
Unpr.							
Flanker	0.05	0.05	0.12	−0.18	0.11	−0.23	0.12
PTA	0.05	0.00	0.85	0.07	0.37	0.04	0.85
Age	0.07	0.01	0.44	−0.10	0.12	−0.19	0.61
8-ch							
Flanker	0.17	0.17	<0.01	−0.29	0.10	−0.41	0.00
PTA	0.19	0.02	0.28	0.34	0.31	0.20	0.28
Age	0.21	0.02	0.28	0.11	0.10	0.25	0.28
6-ch							
Flanker	0.19	0.19	<0.01	−0.42	0.13	−0.44	0.00
PTA	0.20	0.01	0.52	−0.28	0.43	−0.12	0.52
Age	0.20	0.00	0.94	−0.01	0.14	−0.02	0.94
4-ch							
Flanker	0.19	0.19	<0.01	−0.53	0.17	−0.43	0.00
PTA	0.20	0.01	0.40	0.47	0.55	0.16	0.40
Age	0.27	0.07	<0.05	0.35	0.17	0.45	0.05

Each row displays information regarding the overall fit of the current model as well as the individual contribution of the predictor variable added to the previous model at each step.
**The significance of the change in R^2 for each successive model.*
†The standardized multiple regression coefficient for each predictor.
‡The significance of each added predictor's contribution to the model.

7% of the variance in listening effort. The standardized regression coefficients (β) and associated significance values for each model further demonstrate that the model containing only Flanker score accounted for the most variance in listening effort expended by YA and OA in all conditions except the most degraded (i.e., 4-ch) condition, in which age also significantly contributed to the model.

DISCUSSION

The purpose of the present study was to examine the differences in dual-task performance between YA and OA during degraded speech recognition and to determine whether age-related differences in hearing acuity as well as executive control contribute to these differences. To address these issues, YA and OA completed a dual-task paradigm consisting of a primary degraded speech recognition task and a secondary visual monitoring task. Changes in participants' accuracy and reaction time on the visual monitoring task in the dual-task conditions, relative to the single-task condition, served as an estimate of cognitive spare capacity and was interpreted as listening effort.

Speech Recognition

Consistent with previous work, both YA and OA exhibited decrements in speech recognition as spectral degradation increased (Shannon et al. 1995; Eisenberg et al. 2000; Friesen et al. 2001; Başkent 2006). There was an overall difference in speech recognition between YA and OA across conditions, suggesting that OA had greater difficulty with the primary task than YA. Planned paired comparisons between YA and OA *within* each condition, however, failed to find statistically significant age-related differences in speech recognition in the unprocessed, 8-ch, and 6-ch conditions. The statistically different performance between the two groups in the 4-ch condition (before correcting for multiple comparisons) likely dominated the main effect of age as well as the condition-by-age interaction. These results are consistent with studies that have evaluated age-related differences in speech recognition of noise-vocoded stimuli (Souza and Boike 2006; Sheldon et al. 2008). The studies converge on the idea that OA require greater spectral fidelity to achieve similar speech recognition as YA, although, in the present data, the age-related difference is only observed for the most degraded condition. It is important to note that the observed differences between the studies may reflect differences in hearing acuity among the older participants: participants in Souza and Boike (2006) study had greater hearing impairment than the participants of the present study or those in the study by Sheldon et al. (2008). An additional consideration is that the use of sentences containing contextual cues allowed a partial compensation for age-related declines in suprathreshold auditory processing in the present study.

Executive Control

As expected, there were differences in executive control with age. Results revealed better scores on the Flanker task in YA than OA. These differences paralleled differences in performance on the visual monitoring task when performed in isolation: YA responded significantly faster and more accurately than OA. The age-related differences in performance on the visual monitoring task and significant, albeit weak, correlations between visual monitoring performance and Flanker scores suggest that there are overlapping cognitive processes captured by these measures.

Visual Monitoring (Single vs. Dual Task)

The critical analysis of the present study was determining how accuracy and reaction time on the visual monitoring task changed between the single- and dual-task conditions. Though these changes were used as a proxy for listening effort (Figure 4), it is unclear as to whether the results reflect listening effort per se or rather the cognitive cost of effortful listening when performing concurrent tasks. This idea is expanded in the discussion below.

Consistent with our hypotheses, both YA and OA showed greater declines in secondary-task accuracy and reaction time, with increased spectral degradation of the primary speech recognition task. These results are consistent with previous studies that demonstrate greater listening effort under conditions of reduced spectral information (Hervais-Adelman et al. 2012; Pals et al. 2013; Winn et al. 2015). These results are also consistent with the idea that the cognitive cost of effortful listening is not necessarily modality-specific given that changes in a secondary visual monitoring task were observed with decreasing spectral degradation of the primary speech recognition task.

More relevant to the objective of the present study are the observed age-related differences in secondary-task performance during the dual-task conditions. There were significant group differences in secondary-task accuracy (but not reaction time) between YA and OA with increasing spectral degradation during the dual-task conditions. The lack of an age effect in listening effort measured from reaction time in the present study may be due to one or a combination of factors. First, it may be related to the instructions that participants were given before the task, which was to "press a key as quickly as possible when the same picture occurred twice in a row." Participants may have only responded when they were confident of the response, thus allowing them to maintain a stable response speed. Alternatively, the nature of the visual monitoring task may not have been sensitive enough to capture differences in processing speed between YA and OA. For example, a different experimental paradigm with discrete trials that required participants to respond before subsequent trials (i.e., similar to the design in Gosselin & Gagné 2011) may have yielded different results.

The observation that secondary-task accuracy was significantly poorer in OA supports the idea that OA had fewer cognitive resources available during degraded speech recognition. The findings of the present study suggest that, under certain conditions of spectral degradation, OA are able to perceive degraded speech as well as YA, but endure a greater cognitive cost to achieve this performance (Tun et al. 2009; Gosselin & Gagné 2011; Desjardins & Doherty 2013). This cognitive cost inherent to effortful listening has been shown to have associated real-life consequences, such as difficulty taking notes during a lecture or driving while holding a conversation, as well as chronic feelings of fatigue (for a commentary on effort-related fatigue, see Bess & Hornsby 2014). Whether the cognitive cost experienced by OA in the present study reflects a greater reallocation of available cognitive resources to speech recognition (compared to YA) or simply fewer available cognitive resources than YA to perform the visual monitoring task cannot be elucidated by the current data. However, given the presence of age-related differences in Flanker scores and baseline performance on the visual monitoring task, the latter explanation seems more plausible.

Notably, YA and OA exhibited statistically similar changes in secondary-task accuracy during the unprocessed and 6-ch dual-task conditions. The fact that there was no difference between YA and OA in the unprocessed condition is consistent with the idea that recognizing clear speech is minimally effortful and, therefore, results in a greater cognitive spare capacity to perform the secondary task. Though unexpected, the absence of a difference between YA and OA in the 6-ch condition is likely due to the smaller effect size observed in the 6-ch condition versus the 8-ch or 4-ch conditions. A plausible explanation for this result is that YA required greater spectral degradation to experience similar cognitive costs as OA. The significant age group-by-condition interaction is consistent with this interpretation. As displayed in Figure 4A, OA began to expend additional listening effort in the 8-ch condition, whereas YA did not experience an increase in listening effort until the more spectrally degraded 6-ch condition.

Individual Differences in Dual-Task Performance

To probe individual differences in secondary task accuracy during the dual-task paradigm within our sample, hierarchical multiple regression analyses were conducted for each condition. In the unprocessed condition, secondary-task performance was largely unimpaired during the dual task in both YA and OA. As such, it is reasonable that none of the age-related factors entered into the model had any predictive power. Age-related changes in executive control, as measured by Flanker score, accounted for a significant amount of variance in secondary-task accuracy in the 8-ch, 6-ch, and 4-ch dual-task conditions. This finding suggests that the observed age-related differences in the degree to which secondary-task performance declined during the dual-task conditions are likely due to poorer executive control in OA. Interestingly, age contributed significantly over and above Flanker score in the 4-ch condition only, indicating that the mechanisms driving secondary-task performance in this dual-task condition are not solely cognitive. A potential explanation for this finding is that the ability to perceive severely degraded speech is highly dependent on cognitive processing (to compensate for missing linguistic content) as well as suprathreshold auditory processing (to encode information from the temporal envelope of the signal). Because previous studies have shown significant age-related decrements in auditory temporal envelope sensitivity in adults with normal hearing acuity through 2 kHz (Pichora-Fuller & Schneider 1992) or 6 kHz (Strouse et al. 1998; Füllgrabe et al. 2015), it is likely that the age-related differences in secondary-task accuracy observed in the 4-ch condition are reflective of the additional resources OA needed to process the degraded speech signal.

Limitations

While the present study contributes to the knowledge of how age influences dual-task performance in a paradigm designed to quantify listening effort, there are limitations to consider. First, while both groups demonstrated age-typical hearing, the audiometric thresholds of YA and OA in the present study were not matched at all frequencies: OA had poorer high-frequency thresholds than YA on average. Even small differences in audiometric thresholds can affect degraded speech recognition (e.g., Dubno & Ahlstrom 1997), and this may be more apparent in conditions with greater acoustic degradation.

Another factor that may have influenced results is that the cognitive measure used in the present study only assessed cognitive processes related to executive control. The rationale for choosing the Flanker task as the standardized cognitive measure in this study was that the visual monitoring task was dependent on participants' ability to selectively attend (i.e., monitoring each image in the ongoing stream) and inhibit irrelevant input (i.e., only providing a response on repeating images).

Although results demonstrated age-related differences in executive control between YA and OA, there were likely other cognitive processes at play during the primary and secondary tasks, including auditory working memory and visuospatial working memory, respectively. Thus, there may be other aspects of cognitive capacity not captured by the Flanker task that contributed to participants' dual-task performance.

Finally, it is reasonable to assume that other age-related factors beyond hearing acuity and executive control contributed to the age-related differences in dual-task performance observed in the present study. For example, age-related differences in temporal envelope sensitivity, auditory filter width, and processing speed may have significantly contributed to listening effort during degraded speech recognition if measured independently.

CONCLUSIONS

The purpose of the present study was to investigate age-related differences in secondary-task performance during degraded speech recognition. The primary findings from this study are as follows: (1) both YA and OA experienced declines in secondary-task accuracy with increased spectral degradation of the primary task; (2) OA experienced greater declines in secondary-task performance than YA, which was most evident in the 8-ch and 4-ch conditions; (3) age-related differences in executive control significantly accounted for a portion of the observed differences in listening effort between YA and OA; and (4) age significantly contributed to listening effort to a greater extent than executive control in the 4-ch condition.

ACKNOWLEDGMENTS

The authors thank Julia Curato for her assistance with data processing and organization. In addition, the authors extend their gratitude to Andy Sabin for developing the original version of the visual monitoring task and Brock Ferguson for his feedback on the customized software program used in the present study.

This research was supported by the School of Communication at Northwestern University and National Institutes of Health—National Institute on Deafness and Other Communication Disorders (NIH-NIDCD) R01 DC012289 (to Pamela Souza).

The authors declare no conflict of interest.

Address for correspondence: Kristina M. Ward, The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA. E-mail: kmward@u.northwestern.edu

Received September 21, 2015; accepted June 5, 2016.

REFERENCES

Başkent, D. (2006). Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels. *J Acoust Soc Am*, *120*, 2908–2925.

Beaumont, J. L., Havlik, R., Cook, K. F., et al. (2013). Norming plans for the NIH Toolbox. *Neurology*, *80*(11 Suppl 3), S87–S92.

Bench, J., Kowal, A., Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol*, *13*, 108–112.

Bess, F. H., Hornsby, B. W. Y. (2014). Commentary: Listening can be exhausting – fatigue in children and adults with hearing loss. *Ear Hear*, *35*, 592–599.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, *10*, 433–436.

Broadbent, D. E. (1958). *Perception and Communication*. Elmsford, NY: Pergamon Press.

Brouwer, W. H., Waterink, W., Van Wolffelaar, P. C., et al. (1991). Divided attention in experienced young and older drivers: lane tracking and visual analysis in a dynamic driving simulator. *Hum Factors*, *33*, 573–582.

Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., et al. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. The Epidemiology of Hearing Loss Study. *Am J Epidemiol*, *148*, 879–886.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., et al. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *J Exp Psychol Gen*, *134*, 222–241.

Degeest, S., Keppler, H., Corthals, P. (2015). The effect of age on listening effort. *J Speech Lang Hear Res*, *58*, 1592–1600.

Desjardins, J. L., & Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear Hear*, *34*, 261–272.

Downs, D. W. (1982). Effects of hearing and use on speech discrimination and listening effort. *J Speech Hear Disord*, *47*, 189–193.

Downs, D. W., & Crum, M. A. (1978). Processing demands during auditory learning under degraded listening conditions. *J Speech Hear Res*, *21*, 702–714.

Dubno, J. R., & Ahlstrom, J. B. (1997). Additivity of multiple maskers of speech. *Modeling sensorineural hearing loss*, 253–272.

Durlach, N. I., Mason, C. R., Kidd Jr, G., et al. (2003). Note on informational masking (L). *The Journal of the Acoustical Society of America*, *113*(6), 2984–2987.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., et al. (2000). Speech recognition with reduced spectral cues as a function of age. *J Acoust Soc Am*, *107*(5 Pt 1), 2704–2710.

Francis, A. L., & Füllgrabe, C. (2015). Research on listening effort: History and methods, theory, and practice. *J Acoust Soc Am*, *137*, 2209.

Fraser, S., Gagné, J. P., Alepins, M., et al. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. *J Speech Lang Hear Res*, *53*, 18–33.

Friesen, L. M., Shannon, R. V., Baskent, D., et al. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am*, *110*, 1150–1163.

Füllgrabe, C., Moore, B. C., Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front Aging Neurosci*, *6*, 347.

Gordon-Salant, S. (1986). Effects of aging on response criteria in speech-recognition tasks. *J Speech Hear Res*, *29*, 155–162.

Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *J Speech Lang Hear Res*, *40*, 423–431.

Gosselin, P. A., & Gagné, J. P. (2010). Use of a dual-task paradigm to measure listening effort. *Can J Speech Lang Pathol Audiol*, *34*, 43–51.

Gosselin, P.A., & Gagné, J. P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *J Speech Lang Hear Res*, *54*, 944–958.

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

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., et al. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Lang Cogn Process*, *27*, 1145–1166.

Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *J Speech Lang Hear Res*, *45*, 573–584.

Humes, L. E., & Roberts, L. (1990). Speech-recognition difficulties of the hearing-impaired elderly: the contributions of audibility. *J Speech Hear Res*, *33*, 726–735.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Kemper, S., Schmalzried, R., Herman, R., et al. (2009). The effects of aging and dual task demands on language production. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*, *16*, 241–259.

Kleiner, M., Pelli, D. G., Brainard, D. H. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 1–16.

McDonald, S. (Ed.). (2014). Special series on the Cognition Battery of the NIH Toolbox. *J Int Neuropsych Soc*, *20*, 487–651.

McGarrigle, R., Munro, K. J., Dawes, P., et al. (2014). Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *Int J Audiol*, *53*, 433–440.

Pals, C., Sarampalis, A., Baskent, D. (2013). Listening effort with cochlear implant simulations. *J Speech Lang Hear Res*, *56*, 1075–1084.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis*, *10*, 437–442.

Pichora-Fuller, M. K., & Schneider, B. A. (1992). The effect of interaural delay of the masker on masking-level differences in young and old adults. *J Acoust Soc Am*, *91*(4 Pt 1), 2129–2135.

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*, 593–608.

Picou, E. M., Ricketts, T. A., Hornsby, B. W. (2011). Visual cues and listening effort: individual variability. *J Speech Lang Hear Res*, *54*, 1416–1430.

Rabbitt, P. M. (1966). Recognition: Memory for words correctly heard in noise. *Psychonomic Science*, *6*, 383–384.

Rakerd, B., Seitz, P. F., Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing losses. *Ear Hear*, *17*, 97–106.

Rönnberg, J., Lunner, T., Zekveld, A., et al. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front Syst Neurosci*, *7*, 31.

Rönnberg, N., Stenfelt, S., Rudner, M. (2011). Testing listening effort for speech comprehension using the individuals' cognitive spare capacity. *Audiol Res*, *1*, e22.

Sarampalis, A., Kalluri, S., Edwards, B., et al. (2009). Objective measures of listening effort: effects of background noise and noise reduction. *J Speech Lang Hear Res*, *52*, 1230–1240.

Schvartz-Leyzac, K. C., & Chatterjee, M. (2015). Fundamental-frequency discrimination using noise-band-vocoded harmonic complexes in older listeners with normal hearing. *J Acoust Soc Am*, *138*, 1687–1695.

Shannon, R. V., Zeng, F. G., Kamath, V., et al. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Sheldon, S., Pichora-Fuller, M. K., Schneider, B. A. (2008). Effect of age, presentation method, and learning on identification of noise-vocoded words. *J Acoust Soc Am*, *123*, 476–488.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp Psychol Hum Learn*, *6*, 174–215.

Somberg, B. L., & Salthouse, T. A. (1982). Divided attention abilities in young and old adults. *J Exp Psychol Hum Percept Perform*, *8*, 651–663.

Souza, P., Arehart, K., Miller, C. W., et al. (2011). Effects of age on F0 discrimination and intonation perception in simulated electric and electroacoustic hearing. *Ear Hear*, *32*, 75–83.

Souza, P., & Rosen, S. (2009). Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech. *J Acoust Soc Am*, *126*, 792–805.

Souza, P. E., & Boike, K. T. (2006). Combining temporal-envelope cues across channels: effects of age and hearing loss. *J Speech Lang Hear Res*, *49*, 138–149.

Strouse, A., Ashmead, D. H., Ohde, R. N., et al. (1998). Temporal processing in the aging auditory system. *J Acoust Soc Am*, *104*, 2385–2399.

Tukey, J. W. (1960). The practical relationship between the common transformations of percentages or fractions and of amounts. Reprinted in Mallows, C.L. (ed.) 1990. *The Collected Works of John W. Tukey. Volume VI: More Mathematical*. Pacific Grove, CA: Wadsworth & Brooks-Cole, 211–219.

Tun, P. A., McCoy, S., Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychol Aging*, *24*, 761–766.

Walton, J. P. (2010). Timing is everything: temporal processing deficits in the aged auditory brainstem. *Hear Res*, *264*, 63–69.

Winn, M. B., Edwards, J. R., Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear*, *36*, e153–e165.

Wright, B. A., Conderman, J. S., Waggenspack, M. K., et al. (2014). Prevention of learning of a non-native phonetic contrast by prior exposure to the contrasting stimuli while performing an irrelevant visual task [Abstract]. *J Acoust Soc Am*, *135*, 2357.