

Effects of masker type, sentence context, and listener age on speech recognition performance in 1-back listening tasks

Jaclyn Schurman

Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

Douglas Brungart

National Military Audiology and Speech Pathology Center, Walter Reed National Military Medical Center, Bethesda, Maryland 20889

Sandra Gordon-Salant^{a)}

Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

(Received 21 April 2014; revised 15 October 2014; accepted 30 October 2014)

Studies have shown that older listeners with normal hearing have greater difficulty understanding speech in noisy environments than younger listeners even during simple assessments where listeners respond to auditory stimuli immediately after presentation. Older listeners may have increased difficulty understanding speech in challenging listening situations that require the recall of prior sentences during the presentation of new auditory stimuli. This study compared the performance of older and younger normal-hearing listeners in 0-back trials, which required listeners to respond to the most recent sentence, and 1-back trials, which required the recall of the sentence preceding the most recent. Speech stimuli were high-context and anomalous sentences with four types of maskers. The results show that older listeners have greater difficulty in the 1-back task than younger listeners with all masker types, even when SNR was adjusted to produce 80% correct performance in the 0-back task for both groups. The differences between the groups in the 1-back task may be explained by differences in working memory for the noise and spatially separated speech maskers but not in the conditions with co-located speech maskers, suggesting that older listeners have increased difficulty in memory-intensive speech perception tasks involving high levels of informational masking. © 2014 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4901708>]

PACS number(s): 43.71.Lz, 43.72.Dv, 43.71.Gv [DB]

Pages: 3337–3349

I. INTRODUCTION

One of the greatest challenges faced by the aging population is the systematic increase in difficulty that older listeners have understanding speech in noisy environments. In part, this increased difficulty can be explained by the poorer hearing thresholds that older adults acquire as a result of presbycusis and a lifetime of exposure to loud noise. However, these problems in speech perception do not appear to be limited only to those individuals who experience impaired hearing thresholds as they age: Older individuals who have normal audiometric thresholds also tend to report problems understanding speech in complex environments. Indeed, numerous studies have reported that older individuals have greater difficulty understanding speech in noise than younger individuals, even when both groups have similar audiometric thresholds (e.g., [Dubno et al., 1984](#)).

As alarming as the results of these studies are, there is reason to believe that they may substantially underestimate the problems that aging individuals may have while engaging in real-world conversations. In the clinic or laboratory, speech perception performance is usually measured in relatively simple tasks that require listeners to attend to a speech stimulus and immediately report what was heard. However,

in real-world conversations, listeners must receive and process the speech signal, monitor ongoing discourse, and respond appropriately during his or her turn in the conversation. The listener's turn may immediately follow a spoken message, or it may be delayed because of an intervening comment from the same or a different communication partner. If the conversation is taking place in a noisy room, the task becomes even more challenging because competing noise can directly mask the spoken message ([Brungart et al., 2001](#)) or competing speech can additionally distract the listener's attention ([Humes et al., 2006](#)). Either form of background competition can reduce the listener's ability to recognize accurately individual phonemes, thereby reducing speech intelligibility. While the listener attempts to understand the somewhat distorted signal, additional spoken information is presented that may be lost.

These demanding listening scenarios are likely to be difficult even for young adults with normal hearing. However, they are expected to be even more demanding for older adults who often experience age-related declines in working memory ([Daneman and Carpenter, 1980](#); [Baddeley, 1992](#)), selective attention ([Humes et al., 2006](#)), and central auditory processing of temporal information ([Anderson et al., 2012](#)), all of which act to reduce speech understanding in noise. As a result of these factors, there is little question that older normal-hearing listeners will require higher signal-to-noise ratios (SNRs) to achieve the same level of speech perception

^{a)}Author to whom correspondence should be addressed. Electronic mail: sgsalant@umd.edu

performance obtained by younger listeners. However, what is not well known is the extent to which these age-related hearing difficulties might be exacerbated in more complex listening tasks that require listeners to hold information in working memory while new information is presented. The general objective of the current study was to compare the performance of older and younger normal-hearing listeners in a delayed recall task to determine whether the two groups performed similarly in situations where the SNR was adjusted to produce equivalent performance in a traditional listening task requiring immediate recall of the stimulus materials. This task was repeated with high-context and “anomalous”-context sentences and four different masker types to determine if there was a significant interaction between these factors and the ability to recall information in a delayed response task.

A. Aging, memory, and speech perception

Normal aging is accompanied by progressive decline in numerous global cognitive domains, particularly working memory capacity (Park *et al.*, 2002), which is thought to be important for language processing. Rönnberg and colleagues (2013) offered the new Ease of Language Understanding (ELU) Model as a framework for examining the role of working memory (WM) capacity during a number of speech communication situations. WM is conceptualized as a finite ability involving short-term processing and storage of incoming information, which permits language comprehension, learning, and other cognitive functions (Rönnberg *et al.*, 2013). WM capacity is often measured with a type of “span” test, which indicates the number of items a person can retain in memory and recall accurately, while also making a judgment about these items. Moreover, there is a trading relation between WM processing (the processing demands of the test) and storage (the number of items recalled correctly). Numerous studies have indicated a relationship between WM capacity and speech recognition performance in noise (Zekveld *et al.* 2014), especially among older listeners or listeners with hearing impairment (e.g., Pichora-Fuller *et al.*, 1995; Lunner, 2003; Koelewijn *et al.*, 2014). The ELU model assumes that the presence of fluctuating noise backgrounds places large demands on phonological processing and hence requires listeners to use their WM system to increase attentional focus. A recent extension of this model indicates that WM capacity is especially important when semantic information is unexpected or incongruent, such that individuals with high WM capacity are better able to inhibit conflicting semantic cues than those with low WM capacity (Zekveld *et al.*, 2011). The ELU model thus predicts that fluctuating noise backgrounds (including background babble) place heavier demands on listeners’ WM than steady state backgrounds and that nonsense sentences similarly place heavier demands on listeners’ WM than predictable sentences. If WM capacity is a limited resource, then those with lower WM capacity will perform more poorly than those with higher WM capacity in fluctuating noise and in tasks requiring identification of sentences with minimal contextual cues.

One of the primary focuses of the present investigation is to measure the direct impact of increasing WM demands

on recognition accuracy during a speech perception task. Most experimental speech paradigms assess speech recognition using immediate recall of the spoken message in which the listener repeats or otherwise identifies the stimulus immediately after it is presented. In a delayed recall task, the word or sentence is presented, and the listener holds this information in memory for recall at a later time, typically after presentation of an intervening speech stimulus, which itself may need to be recalled. Such delayed recall tasks are rarely used in auditory speech recognition experiments. Rather, *N*-back tasks (1-, 2-, or 3-back) are typically used in vision experiments to increase the WM load compared to an immediate recall (0-back) condition (e.g., Braver *et al.*, 1997; Gonçalves and Mansur, 2009). In the current investigation, WM load was increased by using an auditory 1-back task in which the listener was required to hold a sentence in memory and repeat it following presentation of an intervening sentence, similar to the processing and storage demands of the conversational scenario described in the preceding text. Listener performance in this auditory 1-back task was compared to performance in the control condition (immediate recall or 0-back task) to determine the impact of everyday WM demands during conversation when the listener must monitor and maintain relevant information while awaiting their turn in the conversation. According to the ELU model, it may be predicted that increasing WM demands with the auditory 1-back task should have a much heavier toll on older than younger listeners, given that older listeners on average have a lower WM capacity than younger listeners (Park *et al.*, 2002). Combining the predictions of the ELU model, then, it was expected that older listeners (with reduced WM capacity compared to younger listeners) would exhibit poorer performance than younger listeners with each single listening/task demand on WM capacity employed in this study (fluctuating noise maskers, low-context stimuli, and 1-back tasks) and that age-related differences would be even more prominent in conditions that combine two or three of these processing requirements.

One prior study by Pichora-Fuller *et al.* (1995) employed a delayed recall task for speech presented in noise with high and low context speech materials. The objective of this previous study was to measure WM capacity for sentence-final words of the Speech Perception in Noise (SPIN) sentences in a procedure similar to that used to measure WM span for read sentences (Daneman and Carpenter, 1980). In this “listening span” test, listeners were asked to recall the sentence-final word of the preceding *n* SPIN sentences with *n* representing an increasing set size (number of sentences in a set), while also indicating if the final word was predictable from the sentence context. Variables were set size, SNR, and sentence predictability (high vs low). Older listeners consistently demonstrated lower set sizes than younger listeners, and performances of both groups were reduced at the lowest SNR (0 dB), especially for low probability sentences. Thus Pichora-Fuller *et al.* (1995) demonstrated that WM capacity (i.e., WM span) can be measured in a listening span task and that word predictability can affect set size in more adverse noise conditions. However, listeners’ accuracy in recalling an

entire sentence following presentation of an intervening sentence, as would be required in ongoing conversation was not assessed in this prior study.

B. Effect of masker type on immediate and delayed speech perception

In addition to varying WM demands in ongoing speech, the current investigation varied the type of noise in the background. Different types of background noise produce differential effects on speech perception performance (Dubno *et al.*, 1984; Festen and Plomp, 1990; Pichora-Fuller *et al.*, 1995; Başkent *et al.*, 2014). A steady-state or modulated noise produces energetic masking, in which performance declines (relative to performance in quiet) because certain parts of the acoustic structure of the target signal are rendered undetectable by a more powerful masker that overlaps the target in time and frequency. A competing speech signal also produces some energetic masking, but the effects of energetic masking are often overshadowed by the additional effects of informational masking, which occurs when the similarities between the acoustic and semantic content of the target talker and the competing speech make the task of separating the target from the masker (speech stream segregation) more difficult. Processing of speech stimuli in the presence of informational masking should be more demanding than in the presence of energetic masking because the listener is required to segregate the target speech from the competing speech maskers and recognize the target speech signal. Speech maskers are also known to have a detrimental effect on WM even in cases where the WM task occurs in a non-auditory sensory modality. Many researchers have demonstrated that participants who are asked to recall a list of visually presented numbers perform more poorly when the task is performed in the presence of an irrelevant but intelligible speech masker than they do when they perform the task in quiet or in the presence of a steady-state speech-shaped noise (Weeks and Hasher, 2014). There is also some evidence that this irrelevant speech effect may be more detrimental to the WM recall abilities of older listeners than younger listeners (Bell *et al.*, 2008).

Informational masking effects are particularly strong in speech-on-speech masking tasks where the target and masking talkers originate from the same location as they would if they were mixed together on a single telephone line or originating from a single loudspeaker. However, in most real-world situations, the spatial separation between a target speech signal and other speech maskers in the background is a powerful cue that younger listeners can use to improve speech understanding in immediate speech recall tasks (Brungart *et al.*, 2001; Brungart and Simpson, 2002). Unfortunately, the benefits of spatial separation may also be reduced among older people with age-related hearing loss (Helfer and Freyman, 2008), which might cause older listeners with hearing loss to have more difficulty in WM-intensive tasks than younger listeners in these environments.

Previous attempts to model the impacts of WM on speech perception can be used to make some predictions about the interaction one would expect to find between WM

and performance in speech perception and recall tasks with different types of maskers. For example, the ELU model has been used to explain better performance in steady-state noise conditions compared to fluctuating masker conditions because listeners with high WM capacity can attend better in the fluctuating masker conditions (Besser *et al.*, 2013). Additionally, listeners with larger WM span and better inhibitory mechanisms, as measured with the Size Comparison span (SIC span), appear to outperform those with lower WM capacity on long term speech comprehension tasks presented in a speech background (Sörqvist and Rönnerberg, 2012) and on measures of speech recognition threshold (SRT) in conditions with spatial separation (Zekveld *et al.*, 2014). Taken together, these prior findings would suggest that listeners should perform better in delayed speech recall tasks in steady-state noise than background speech and that performance differences between those with higher vs lower WM capacities should be evident. However, the benefit that listeners with higher and lower WM derive from spatial separation in a delayed recall task is currently unknown. It is possible that the reduced difficulty obtained from spatial separation of a target and masker might provide an even greater benefit for individuals with relatively low WM capacities because they have so much difficulty in the co-located condition. On the other hand, it is also possible that individuals with low WM capacity might perform relatively poorly in spatially separated tasks because they have difficulty focusing selective attention on the location of the target talker while performing the delayed recall task. One of the goals of this investigation is to help distinguish between these two possibilities.

C. Study objectives

The overall objective of the current study was to compare the relative impact of age, sentence context, and masker type on speech recognition performance under conditions in which WM demands were increased by requiring the listeners to perform an auditory 1-back task. Comparison to performance in a baseline, auditory 0-back task indicates the effect of increasing WM demands during online speech understanding. In the sentence recognition task used here, the act of holding one sentence in WM while repeating the prior sentence is expected to increase the WM demands and have a greater detrimental effect on performance of older listeners because of reduced WM capacity. Finally, because older listeners rely heavily on semantic contextual cues during challenging listening tasks (Wingfield *et al.*, 1985; Gordon-Salant and Fitzgibbons, 1997), the influence of semantic content during 0-back and 1-back speech recognition tasks was assessed by comparing recognition of the original high probability (HP) revised speech perception in noise test (R-SPIN) sentences (Bilger *et al.*, 1984) and “anomalous probability” (AP) SPIN sentences. The latter sentences had the same syntactic structure as the HP sentences but substituted alternate lexical items for the original words to create sentences that were meaningless. Predictions of the ELU model are that WM resources would be taxed more for AP stimuli than for HP stimuli, and therefore

age-related differences should be minimal for HP stimuli and maximal for AP stimuli.

To examine the effects of different context levels and masker types on performance in a delayed response task, a new experimental protocol was developed that made it possible to efficiently test performance in the 1-back task under conditions that produce equivalent levels of performance in the baseline 0-back task. First, the well-known effects of age-related hearing loss on speech-in-noise performance (e.g., Festen and Plomp, 1990; Humes and Dubno, 2010) were controlled by recruiting younger and older listeners with average normal hearing sensitivity. Within each of these two groups, performance in the 0-back task with each combination of context level and masker type was equalized by using adaptive tracking to set the overall percentage of correct responses to a fixed level (80% correct). Then the 1-back task was run with the same sequence of SNRs used in the 0-back task. If there is no interaction between listener age, context level, and masker type, then one would expect the performance decline in the 1-back task (relative to the 0-back task) to be equivalent in all of the conditions for both groups. However, if certain combinations of age, context, and masker type have different impacts on performance in working-memory-intensive listening tasks than they do in immediate recall tasks, one would expect to see performance differences in the 1-back condition. The next section describes the method in more detail.

II. METHOD

A. Participants

Two groups of participants with normal hearing sensitivity bilaterally [≤ 25 dB Hearing Level (HL), re: ANSI, 2010, at octave intervals between 250 and 4000 Hz] took part in this study: 20 young normal-hearing individuals (YNH) ages 19–24 yr ($M = 20.9$) and 15 older normal-hearing (ONH) individuals ages 66–76 yr ($M = 69.5$). The average pure-tone average (PTA) was 5.66 dB HL for the right ear and 5.41 dB HL for the left ear for the YNH listeners, and 13.33 dB HL for the right ear and 12.55 dB HL for the left ear for the ONH listeners. Note that although the mean thresholds were slightly higher for the ONH listeners, a paired samples t -test revealed that the three-frequency PTA values (500, 1000, and 2000 Hz) between the groups for both ears were not significantly different (right PTA: $p = 0.305$; left PTA: $p = 0.328$).

All participants were native speakers of English and were required to have at least a high school degree. Individuals completed a case history prior to participation to confirm the absence of otalgia, aural fullness, recent ear infections, a history of otologic surgery, and a history of noise exposure. None of the participants reported hearing problems.

B. Stimuli

Stimuli were HP R-SPIN sentences (Bilger *et al.*, 1984) and AP sentences, which were derived from the HP sentences utilizing similar procedures to those used to create the IEEE anomalous sentences corpus (Herman and Pisoni,

2003). The AP sentences were created specifically for use in this study and were never previously recorded, therefore the authors were required to record these original AP sentences. In addition, the original HP R-SPIN sentences also needed to be re-recorded for the target speaker to be the same for the HP and AP sentences. There were 200 HP sentences and 200 AP sentences with three to seven keywords within each sentence. All nouns, verbs and adjectives were considered keywords (HP example: His PLANS MEANT TAKING a BIG RISK. AP example: His DOCTOR DRANK a LOST RISK). Adjustments were made to ensure that the anomalous sentences were grammatically, syntactically, and politically correct. All sentences were recorded by one female speaker who is a native speaker of American English. The recordings were made in an IAC sound booth using a Shure SM63 microphone and a Marantz professional handheld solid-state recorder (Model No. PMD661). The stimuli were then uploaded from the recorder to a Dell PC. Each sentence was spliced and stored as an individual waveform file using Adobe Audition, and the RMS level was held constant for all sentences.

Continuous speech-shaped noise and three types of speech maskers were created to mask the R-SPIN sentences. In the continuous noise condition, the speech-shaped noise was generated with the same long-term average spectrum as the speech maskers. Two speech-shaped noise maskers were looped to play continuously throughout each trial. The same female speaker who recorded the R-SPIN sentences also produced the speech maskers. Four selected passages from Grimm's *Fairy Tale Classics* were recorded. Three types of speech maskers were created from these recordings: A single talker masker (1T), a 2-talker co-located speech masker (2T), and a 2-talker spatially separated speech masker (2T spatial). Depending on the masker type, one or two 45-s long passages were randomly selected from the four possible passages and looped to play continuously throughout the duration of each condition.

In all conditions, the target speech signal was spatially processed by convolving it with the left and right ear head related transfer functions (HRTFs) measured on a Knowles electronics manikin for acoustic research (KEMAR). The HRTFs were measured in the Auditory Localization Facility at Wright-Patterson Air Force Base, OH, using the procedures outlined in Brungart *et al.* (2011). These HRTFs were processed to correct them for the inverse transfer function of the headphones used in the experiment (Sennheiser HDA-200) and used to generate 448-point finite impulse response (FIR) filters for the left and right ears that were convolved with the target speech signal. This processing produced a binaural stimulus that appeared to originate from a location directly in front of a listener wearing headphones. In some conditions of the experiment, the masking signals were convolved with the same HRTFs used to process the target speech, resulting in a stimulus where the target and maskers appeared to be co-located at a position directly in front of the listener (1T, 2T and noise conditions). In other conditions (2T spatial), the masking signals were convolved with HRTFs that were measured at azimuth locations of $\pm 60^\circ$ relative to the KEMAR manikin. This resulted in a stimulus

where the two masking talkers appeared to originate from locations 60° to the left and 60° to the right of the target talker for the 2T spatial condition.

C. Procedures

A measure of WM, the listening span (LSPAN; Daneman and Carpenter, 1980), was administered prior to the experimental testing to assess each participant's short-term memory storage and processing abilities. Participants were asked to listen to sentences and determine if each sentence was true or false. Participants responded "yes" if the sentence was true or "no" if the sentence was false. Participants were then directed to remember the last word in each sentence to be recalled at the end of the set of sentences. As the test progressed, the number of sentences in each set increased from two to eight, requiring the listener to remember an increasing number of final words. If a participant could not successfully complete the two-sentence set trial of the LSPAN, they were not eligible to participate in the experiment. During the experimental conditions, participants were seated in a quiet room and heard sentences presented from a laptop computer (Dell). The sentences were routed from the laptop to an Andrea PureAudio USB-SA external digital sound card and delivered binaurally to the listener through Sennheiser HDA 200 headphones. Sennheiser HDA 200 headphones were chosen due to the excellent passive attention capabilities of this type of headphones.

A total of eight conditions were tested in the R-SPIN portion of the experiment: (1) HP sentences with the 2T (collocated) masker, (2) HP sentences with the noise masker, (3) HP sentences with the 1T masker, (4) HP sentences with the 2T spatial masker, (5) AP sentences with the 2T masker, (6) AP sentences with the noise masker, (7) AP sentences with the 1T masker, and (8) AP sentences with the 2T spatial masker. In each case, the signal level was fixed at 70 dB sound pressure level (SPL) while the noise level varied adaptively.

The data collection in the experiment was conducted with a custom software program (MATLAB 2007a) that generated two types of trials: 0-back trials, where listeners were asked to listen to a sentence and immediately verbally repeat the contents of the sentence back to the experimenter, and 1-back trials, where the listener was asked to listen to a sentence, hold that sentence in memory until another sentence was presented, and then repeat back the contents of the first sentence to the experimenter. Prior to the start of data collection, each listener participated in two 21-trial practice blocks to get them used to this experimental procedure. First, they were presented with 10 consecutive 0-back trials (trials 1-10), where they heard a sentence and repeated the sentence back to the experimenter. In cases where they were unsure, participants were strongly encouraged to guess all words in the sentence. At the completion of the tenth trial, they were told that they would be switching to a one-back condition, and they were played *two* sentences (trials 11 and 12) and asked to repeat the contents of the first sentence (trial 11). After responding to trial 11, they were presented another sentence (trial 13) and asked to repeat back the contents of

the sentence in trial 12. This process repeated until trial 21, where the listener was asked to first repeat back the contents of trial 20 and then to repeat back the contents of trial 21. In the training blocks, the listeners heard a different set of sentences than the ones used in the remainder of the experiment (IEEE sentences) at a relatively high SNR where they had little difficulty understanding the speech.

Once the listeners were comfortable with the experimental procedure, data collection commenced. Over the course of data collection, each listener participated in a total of three blocks of trials in each of the eight listening conditions, presented in random order. Within each listening condition, the first block of trials was always a 20-trial "tracking" block, consisting of 20 0-back trials. In this block of trials, the noise level was changed adaptively. After each sentence, the SNR was reduced by 8 dB times the percentage of correct keywords in the sentence and then increased by 2 dB times the percentage of incorrect keywords in the sentence. This resulted in an adaptive SNR track that converged on an equivalent SNR value that produced 80% correct responses.

The second two blocks in each masking condition were 21-trial "test" blocks that consisted of 10 0-back trials and 11 1-back trials, similar to the training block. The SNR at the start of the first tracking block was set to the final SNR for the tracking block of that condition to ensure that the initial SNR was set to a level that would generate approximately 80% correct keyword identifications. Within the first ten trials, the SNR continued to adapt according to a tracking rule that reduced the SNR by 8 dB times the percentage of correct keywords and decreased the SNR by 2 dB times the percentage of incorrect keywords. Then after the tenth trial of the block, a message displayed on the screen instructing the listener that the task was switching from a 0-back immediate recall task to a 1-back delayed recall task. The participants were told to nod yes when they were prepared to begin the 1-back task. Then the listener heard 11 more trials in the 1-back task. Within each of these 1-back trials, the SNR was set to be equivalent to the SNR that occurred ten trials earlier in the 0-back portion of the combined block of trials. Thus the 11th trial used the same SNR as the 1st trial, the 12th trial used the same SNR as the 2nd trial, and so on. This ensured that the ten 1-back trials were always collected with the same combination of SNR values used in the 0-back trials of the experiment. The 21st sentence in each block, which was presented at the same SNR as the 11th trial, was necessary to serve as a trailing sentence for the 20th sentence but was not included in the scoring of either type of listening trial (0-back or 1-back).

For all 0-back trials (0-back only block, and the first half of the combined blocks), two scores were derived: The SNR corresponding to 80% correct and the percent correct score (to verify that scores were 80% correct). The 1-back trials (second half of the combined block) had the same SNR values as the 0-back trials, so only the percent correct score was measured. The percent correct score was derived from all keywords correct across the ten sentences presented in that block. For the case of 1-back trials, the keywords were summed across sentences 11–20.

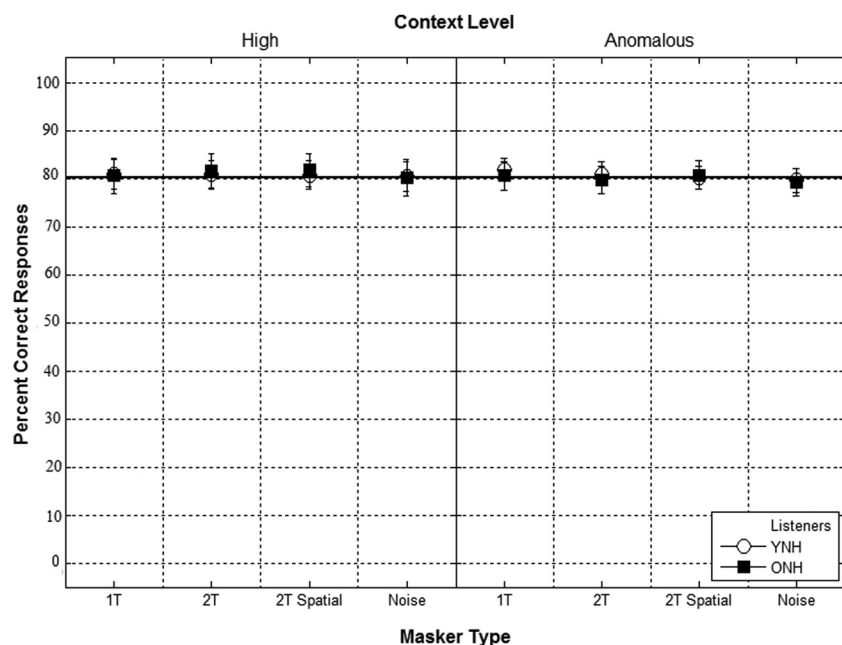


FIG. 1. Percent correct performance in the 0-back task as a function of masker type (1T, single talker; 2T, two talkers; 2T Spatial, two talkers spatially separated; Noise, speech shaped noise) and context type for both groups (YNH, young normal hearing; ONH, older normal hearing).

Testing occurred during one session lasting approximately 1.5–2 h at the University of Maryland, College Park. All participants were paid for their participation. This project had approval from the University of Maryland, College Park Institutional Review Board.

III. RESULTS

A. Evaluation of speech reception thresholds in the 0-back conditions

The overall design of the experiment is predicated on the assumption that the adaptive SNR tracking that occurred in the first half of each combined block was successful in adjusting performance in the 0-back trials to roughly the same overall percentage of correct responses for each stimulus condition tested in the experiment. The percent correct

values shown in Fig. 1 confirm that both the YNH and ONH listeners performed very close to 80% correct responses in all the 0-back listening conditions.

However, there were substantial differences in the SNR values required to achieve 80% correct responses (which we will refer to as the speech reception threshold for 80% responses or the SRT_{80}) in the different conditions of the experiment. Figure 2 and Table I show the mean SRT_{80} values averaged across all the 0-back trials in each testing block of the experiment. An analysis of variance (ANOVA) was performed to determine the effects of group, context and masker for the 0-back task only. The ANOVA revealed significant main effects of group [$F(1,33) = 238.13$, $p < 0.001$], context [$F(1,33) = 421.07$, $p < 0.001$], and masker [$F(3,99) = 356.2$, $p < 0.001$], and significant group \times masker [$F(3,99) = 54.10$, $p < 0.001$], masker \times context [$F(3,99) = 36.55$, $p < 0.001$]

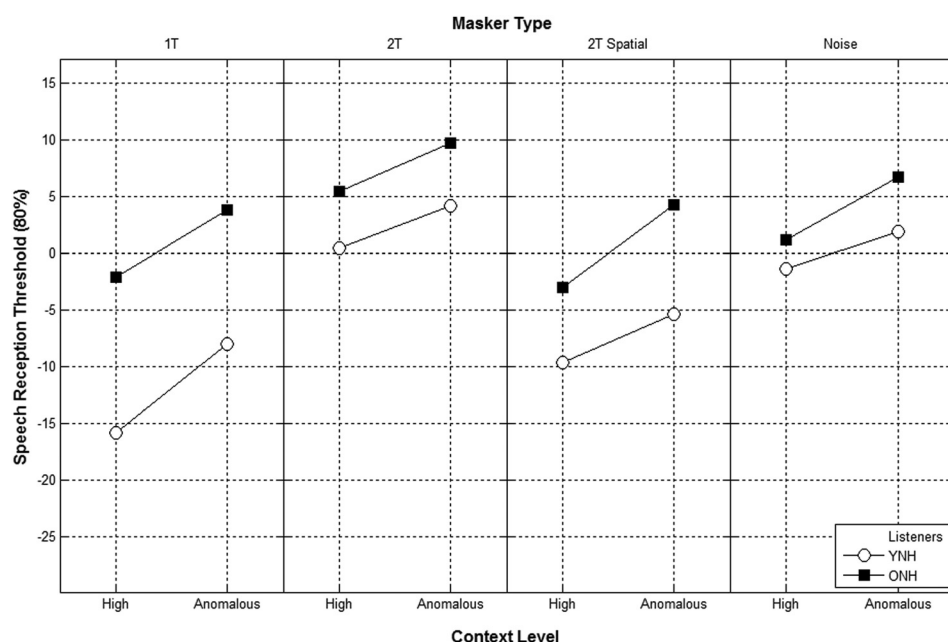


FIG. 2. SRT_{80} scores as a function of masker type (1T, single talker; 2T, two talker; 2T Spatial, two talker spatially separated; Noise, speech shaped noise) and context for both groups (YNH, young normal hearing; ONH, older normal hearing).

TABLE I. Mean SRT₈₀ values for both groups across all masker and context types.

Masker types	HP sentences		AP Sentences	
	YNH	ONH	YNH	ONH
2T co-located	0.48	5.44	4.17	9.72
2T spatial	-9.58	-2.96	-5.43	4.24
1T	-15.93	-2.07	-7.89	3.84
Noise	-1.34	1.19	1.99	6.70

and group \times masker \times context [$F(3,99) = 22.19$, $p < 0.001$] interactions.

The group effect was evaluated for each masker in each sentence context. *Post hoc* independent samples *t*-tests were conducted using the Bonferroni correction and indicated significant effects of group in each masker for the HP sentences ($p < 0.0125$) and in each masker for the AP sentences ($p < 0.0125$), reflecting an advantage for the YNH listeners. However, *post hoc* *t*-test analysis revealed that smaller differences in SNR between groups were observed with the 2T and noise maskers than either the 1T masker or the 2T spatial masker in both sentence contexts. This suggests that YNH listeners are relatively more efficient than ONH listeners at segregating speech either from a single co-located speech masker or from two spatially separated speech maskers, but that both groups have difficulty segregating the target speech from either two co-located speech maskers or a co-located noise masker.

A *post hoc* analysis of the context effect was also conducted for each group in each masker condition using paired comparison *t*-tests and the Bonferroni correction. Results showed that both the YNH and ONH groups were able to take advantage of contextual cues in each masker condition; this resulted in lower SRT₈₀ values for HP sentences than for the AP sentences in the 0-back task ($p < 0.001$).

Finally, a *post hoc* *t*-test analysis of the effect of masker type on recognition performance by both listener groups in each sentence context revealed a similar pattern of performance across both contexts but different effects of masker for the two listener groups. For the YNH listeners, SRT₈₀ performance in each masker condition was significantly different from that in each contrasting masker condition for HP sentences ($p < 0.008$) with the exception of the noise condition compared to the 2T co-located condition. The 1T masker produced the lowest SNR required to reach 80% correct performance compared to the three other maskers ($p < 0.008$). Lower SRT₈₀ values were also found for the 2T spatial masker compared to the noise masker and 2T co-located masker for YNH listeners with HP sentences. Analysis of AP sentences using paired comparison *t*-tests and the Bonferroni correction revealed that SRT₈₀ values in all masker types were significantly different from each other ($p < 0.008$) for YNH listeners, with lowest SRT₈₀ values observed for the 1T masker, followed by the 2T spatial, noise, and then the 2T co-located maskers (highest SRT₈₀ values).

An analysis of the masker effect was also conducted for ONH listeners. An ANOVA revealed significant main effects of context [$F(1,14) = 120.8$, $p < 0.001$] and masker

[$F(3,42) = 49.34$, $p < 0.001$] without a significant interaction between the two main effects. Therefore the masker effect was assessed with data collapsed across both HP and AP sentences using paired comparison *t*-tests. Significant differences in SRT₈₀ performance for ONH listeners were found among the 2T co-located, 2T spatial noise, and 1T ($p < 0.001$) maskers, where the 2T masker produced the highest SRT₈₀ value. In addition, the noise masker produced a higher SRT₈₀ value compared to the 2T spatial masker ($p < 0.001$). Overall the pattern of results indicates that both groups had the most difficulty with the 2T co-located masker with significant improvement in the noise and 2T spatial maskers. However, the YNH listeners showed the least masking with the 1T masker, whereas ONH listeners did not.

B. Evaluation of percent correct performance in the 1-back conditions

In the 1-back task, the percentage of correct responses dropped substantially for both the ONH and YNH listeners. Percent correct scores for the 1-back task are presented in Fig. 3. An ANOVA revealed significant main effects of group [$F(1,33) = 24.56$, $p < 0.001$], context [$F(1,33) = 3.28$, $p < 0.001$], and masker type [$F(3,99) = 11.32$, $p < 0.001$] and a significant interaction between group \times context [$F(1,33) = 5.86$, $p < 0.05$]. *Post hoc* analysis of the effect of masker type (paired-comparison *t*-tests with Bonferroni correction) revealed that overall performance in the 1-back task was significantly worse with the speech maskers than with the noise masker ($p < 0.008$). This result suggests that listeners in both age groups have relatively more difficulty remembering keywords in the 1-back task with a speech masker than a noise masker. The significant interaction between context and listener group was evaluated further. *Post hoc* analysis revealed a significant age effect in each context and masker type ($p < 0.008$), with the exception of performance in the 2T spatial condition with AP sentences (Fig. 3). Thus younger listeners outperformed older listeners in the majority of masker types and contexts. This was true even though the SRT for each group was adjusted to obtain the same

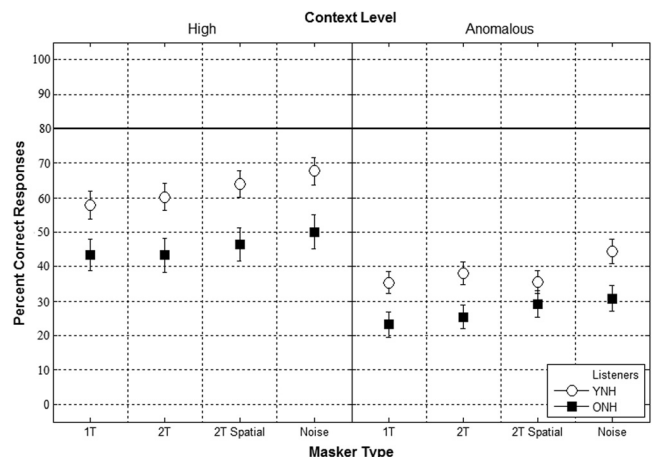


FIG. 3. Percent correct scores for two listener groups (YNH, young normal hearing; ONH, older normal hearing) on the 1-back task in two sentence contexts and four masker conditions (1T, single talker; 2T, two talker; 2T Spatial, two talker spatially separated; Noise, speech shaped noise).

overall performance in the 0-back task. The effect of context was also significant for each group averaged over all masker types ($p < 0.001$), indicating better performance for HP compared to AP sentences. The source of the interaction appears to be a greater difference between groups for percent correct performance for HP [$t(33) = 5.33, p < 0.001$] sentences compared to AP sentences [$t(33) = 3.82, p < 0.001$]. Although some studies that have examined the effects of context in more traditional 0-back tasks report that ONH listeners benefit more from contextual information compared to YNH listeners (e.g., Pichora-Fuller *et al.*, 1995), other studies suggest that the two groups benefit to the same extent when the task samples performance of both groups at below-ceiling levels (e.g., Dubno *et al.*, 2000). The current finding of more contextual benefit by younger than older listeners appears to be contradictory to both perspectives; one possible explanation might be that the additional benefits that older listeners sometimes appear to show for high-context sentences in terms of the SRT for traditional 0-back tasks [as they did particularly for the 2T spatial and noise maskers in this experiment (e.g., right two panels of Fig. 2)] may come at a cost of additional WM resources that interfere with performance in the 1-back task. However, any interaction of this type clearly does not eliminate the advantages of context in memory tasks as it is clear that both groups performed better with the HP sentences than the AP sentences during the 1-back tasks in the present experiment.

C. Interactions between listening span and performance in 1-back and 0-back tasks

Mean LSPAN scores for ONH listeners and YNH listeners were 2.73 and 4.2 ($p < 0.001$), respectively. Figure 4 shows the relationship between LSPAN scores and SRT performance in the 0-back task. Inspection of Fig. 4 indicates that the ONH group exhibited much higher (poorer) SRT₈₀ values than the YNH group even in cases where their LSPAN values overlapped (e.g., LSPAN score of 3 for each of the four masker types). A bivariate correlation was

computed to assess the relationship between LSPAN scores and SRT separately for each masker and each listener group. Results indicate that the LSPAN scores did not accurately predict SRT in the 0-back task with the exception of a correlation in the 1T masker for the YNH participants [$r(35) = -0.607, p < 0.01$]. In addition, step-wise multiple linear regression analyses showed that listener age was identified as the most significant predictor variable of the SRT₈₀ values for each masker type with the LSPAN score either not identified as significant or identified with much lower significance than listener age across all masker types.

The LSPAN did, however, appear to be a better predictor of performance in the 1-back conditions than the 0-back conditions of the experiment. Figure 5 shows the relationship between the percent correct score in the 1-back task and the LSPAN score for the YNH and ONH listeners in the experiment. The large symbols in the figure show the mean percent correct scores when the individual listeners were grouped together into four categories on the basis of their LSPAN scores. For example, the leftmost filled square in each panel shows the mean percent correct performance in the 1-back condition for the ONH listeners who had LSPAN scores in the range from 2 to 2.5. These binned data are helpful for showing the overall trend in the data for each listener group. The data for the individual subjects in each group are also shown by the small symbols in each panel. A bivariate correlation was computed to assess the relationship between LSPAN scores and these individual subject percent correct scores in the 1-back task. This correlation revealed strong and significant correlations with percent correct scores in the 1-back task in each masker condition: Noise [$r(35) = 0.707, p < 0.01$], 2T masker [$r(35) = 0.690, p < 0.01$], 2T spatial masker [$r(35) = 0.642, p < 0.01$], and 1T masker [$r(35) = 0.593, p < 0.01$]. One notable feature of these data is that the lines that best fit the relationship between the LSPAN scores and percent correct responses in the 1-back task appear to be co-linear for the ONH and YNH groups in the 2T spatial and noise conditions but appear to be roughly parallel with an approximately 10% lower y intercept for the

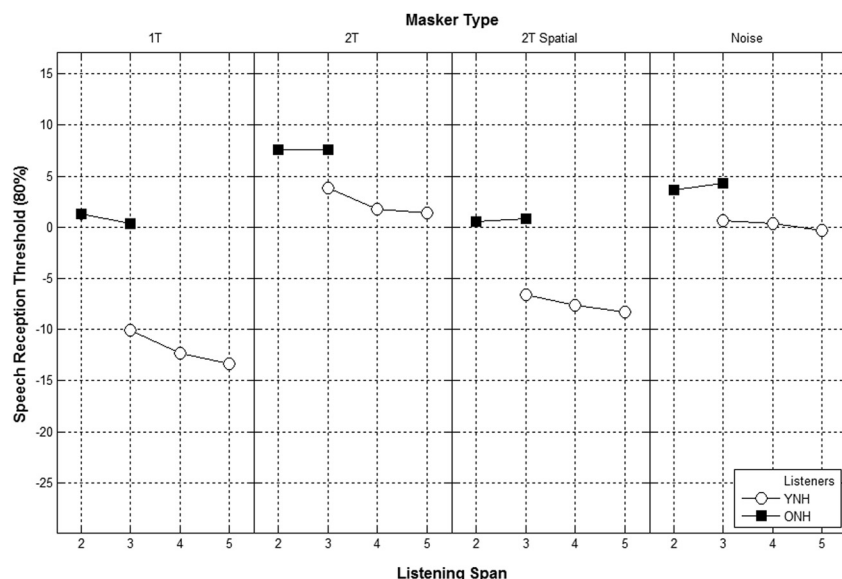


FIG. 4. Correlation data between LSPAN scores and SRT₈₀ scores for the two listener groups (YNH, young normal hearing; ONH, older normal hearing) in four masker types (1T, single talker; 2T, two talker; 2T Spatial, two talker spatially separated; Noise, speech shaped noise). LSPAN scores were collapsed into four score categories (2 = scores of 2 and 2.5; 3 = scores of 3 and 3.5; 4 = scores of 4 and 4.5; 5 = scores of 5 and 5.5).

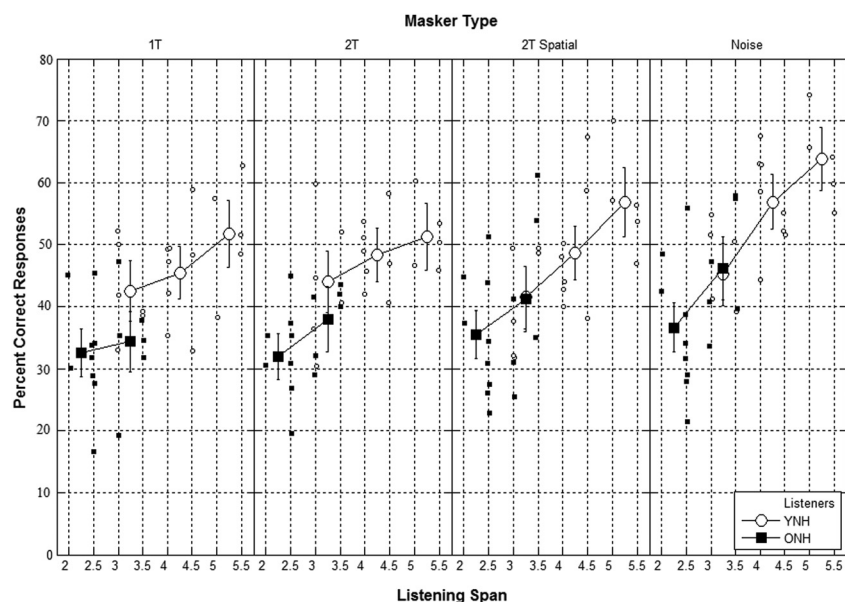


FIG. 5. Symbols represent the average percent correct scores in the 1-back task and LSPAN scores collapsed across four listening span categories (2–2.5, 3–3.5, 4–4.5, 5–5.5) for each of the two listener groups, shown separately for the four masker types. Individual listener data points are also plotted.

ONH group in the 1T and 2T conditions. This suggests that the WM component assessed by the LSPAN test is a good predictor of performance for both age groups in cases with a noise masker or with spatially separated speech maskers but that the ONH group has additional difficulties that cannot be accounted for by the LSPAN when the task required them to segregate a speech target from a co-located speech masker.

D. Performance as a function of trial order

A final analysis of performance in the two tasks is given in Fig. 6, which shows how performance varied with trial order across the 21 trials in each test block of the experiment. These results, which have been averaged across both context levels and all four masker types, show that performance was relatively stable for the ten 0-back trials for both groups of listeners. In trial 11, which was the first 1-back

trial in the block, performance dropped off only slightly in the YNH listeners but dropped off substantially more for the ONH listeners. On the 12th trial, performance dropped substantially for both groups and remained roughly constant for both listening groups for trials 12–19. Then on trial 20, there appears to be a divergence in performance between the ONH and YNH groups, where the YNH listeners improved and the ONH listeners declined. On trial 21, which was the additional trial that was added as an interferer for the 20th trial of the experiment, performance improved much more for the YNH group than for the ONH group. The fact that the YNH group performed nearly as well in the first 1-back task as they did in the last 0-back task, and the ONH listeners did not, suggests that much of the greater difficulty experienced by the ONH listeners was the result of WM issues rather than a result of having greater difficulty extracting a speech signal from noise *while* remembering an earlier sentence. The source of the substantial decline in performance of the ONH listeners on trial 11 is unclear. One possibility is that the ONH listeners had more difficulty switching to a new task than the YNH listeners despite the written instructions on the computer monitor prior to the onset of the 1-back task. Task switching is considered to be an executive function because it involves cognitive control and shifting attention. Executive function is a cognitive ability that is known to decline with aging (e.g., Zelazo *et al.*, 2004); however, it was not measured specifically in this investigation, and thus its contribution to the older listeners' decline in performance at the juncture between tasks is unclear. A second possibility is that the ONH listeners experienced a decline in attentional focus (listener fatigue) after the tenth sentence, unlike the YNH listeners. This possibility is unlikely for three reasons: (1) The listeners were provided as long as they needed to make the responses in trials 1–10 of each block, (2) the listeners were given as long as they desired to acknowledge the switch to a one-back task in trial 11, and (3) both the ONH and YNH listeners appear to have relatively stable performance over the course of trials 12 through 19, where one would expect the effects of cumulative listener fatigue to

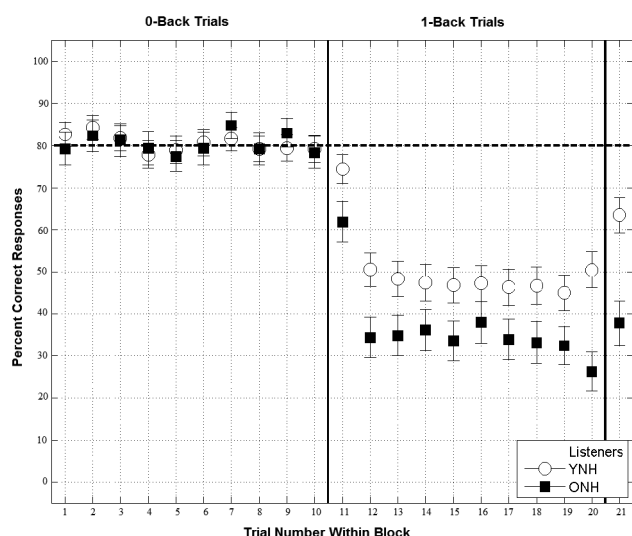


FIG. 6. Trial-by-trial percent-correct performance over the 21 trials within a test block for the two listener groups, averaged across the two contexts and four masker conditions. Error bars represent 95% confidence intervals for each trial position.

have a large impact on performance. This result is also consistent with studies of the time course of visual attention indicating that the central processes involved in allocating attention to a cued spatial location (in visual space) does not decline with aging (Gottlob and Madden, 1998).

IV. DISCUSSION

A. Age-related differences in speech perception in the 0-back and 1-back tasks

The goals of this study were to determine the impact of age, sentence context, and masker type on speech recognition and recall performance under conditions requiring the listeners to perform an immediate speech recognition task (0-back task) and a secondary WM task involving delayed recall (1-back task). As expected, the results show that older adults required a higher SNR to reach 80% correct in the 0-back task for all masker conditions. Difficulty understanding speech in noise is a common complaint for ONH listeners and has been well documented throughout the literature (Pichora-Fuller *et al.*, 1995; Tun and Wingfield, 1999). Therefore results showing that older listeners require a higher SNR to reach the same percent correct score as a younger listener confirm previous findings. Listener age also had a significant impact on performance in the 1-back task. Even when the SNR in the 0-back task was adjusted to produce equivalent performance, the older listeners consistently performed worse in the 1-back task than the younger listeners. The age-related difficulty in the 1-back task represented an approximate 15%–20% difference between groups in most maskers and in both contexts. It is noteworthy that this age effect was observed for listeners with normal hearing, for whom the signals were audible, underscoring the excessive difficulty that ONH listeners experience in listening conditions that combine background noise with WM demands. This result is consistent with the notion that cognitive resources needed for speech processing abilities decline with age (Committee on Hearing Bioacoustics and Biomechanics, 1988; Gordon-Salant and Fitzgibbons, 1997; Pichora-Fuller, 2003) and, to some extent, it can be explained directly by differences in the WM capacities of the ONH and YNH listeners. In addition, these results agree with the ELU model predictions that suggest individuals with decreased WM capacity (ONH), determined by LSPAN scores, will perform more poorly compared to participants with higher WM capacity (YNH) on difficult recall tasks (1-back) even when speech signals are audible.

B. Effect of context

The ELU model also suggests that individuals with decreased WM capacity will have increased difficulty on recall tasks with minimal contextual cues; this is demonstrated by the results of the current study. Sentence context was manipulated in this experiment by comparing recognition and recall of sentences containing several cues to word identity (R-SPIN HP sentences) and sentences that retained the syntactic structure of the HP sentences but contained no semantic contextual cues (AP sentences). This strategy was

implemented because recognition of each keyword in the sentence was scored rather than just the final test word (as in the standard R-SPIN test), and hence sufficient variety in the lexical items comprising the keywords in the low context sentences was required. Thus the AP sentences are somewhat different from the more typical “low probability” (LP) SPIN sentences used in other studies (e.g., Pichora-Fuller *et al.*, 1995). Findings from the 0-back task (Fig. 2) showed that for both listener groups and all masker conditions, the SRT required to achieve 80% correct recognition was lower for the HP sentences than for the AP sentences, as expected.

However, in the 1-back task, a significant interaction between context and listener group was observed; this was attributed to a greater group effect for HP than AP sentences. Essentially, the YNH listeners took more advantage of contextual cues than the ONH listeners when the recall task involved a WM component. This finding is somewhat surprising given that prior research suggests either that older listeners benefit more from contextual cues on a speech recognition task than younger listeners (Pichora-Fuller *et al.*, 1995) or to the same extent as younger listeners when listeners are tested in sufficient noise levels to avoid ceiling effects (Dubno *et al.*, 2000). These prior studies, however, involved word recognition and an immediate recall (0-back) task. The current observation that ONH listeners may not be able to capitalize on their knowledge of the language to enhance their performance on a sentence recall task involving a memory component has two important implications. The first is that older listeners may be at a greater disadvantage in everyday conversation than previously thought because their knowledge of the language does not aid speech understanding for longer conversational discourse or with multiple speaking partners (i.e., holding information in WM). The second implication, related to the first, is that taxes on WM render other cognitive abilities that support communication less efficient, especially among older people.

C. Effect of masker type in the 0-back and 1-back tasks

The largest SNR difference between groups in the 0-back task was found in conditions with the 1T masker and with the 2T spatial masker. Results showing that older listeners have excessive difficulty with a single co-located masker are similar to findings reported previously with the Synthetic Sentence Identification Test–Ipsilateral Competing Message (SSI-ICM: Gates *et al.*, 2008) and support the ELU model prediction that those with lower WM capacity will perform more poorly in situations with fluctuating noise compared to individuals with high WM capacity. The finding that older listeners have excessive difficulty with the 2T spatial masker is consistent with data suggesting that older listeners are less able to take advantage of spatial cues than younger listeners. The impact that age has on the ability to take advantage of spatial separation cues in a multitalker listening task is one area where there are conflicting findings in the literature. Singh *et al.* (2008) and Dubno *et al.* (2008) found that younger and older adults did not differ in the ability to use spatial cues to improve speech recognition. In contrast, other

findings indicate that although older normal hearing listeners benefited from spatial separation cues, older participants benefited less when compared to younger listeners (Dubno *et al.*, 2002; Helfer *et al.*, 2010) or benefited less from spatial separation cues than would be predicted (Dubno *et al.*, 2008). The results from the current study support those of Dubno *et al.* (2002) and Helfer *et al.* (2010) in suggesting that there may be an age-related decline in the benefit of a spatial release from masking.

Both the older and younger groups exhibited relatively greater difficulty in the 1-back task with the co-located speech maskers (1T and 2T) than with the noise condition. In part, this result could be related to the *irrelevant speech effect*, which is a well-known effect in the literature on WM in which a listener's ability to recall a list of visually presented words is compromised when the listener performs the WM task while listening to an irrelevant speech signal (as opposed to a non-speech masking noise or silence). To the extent that the presence of irrelevant speech impairs a listener's ability to store items in WM over time, one might expect it to generalize to cases where a listener is trying to remember an auditorily presented sentence rather than a visually presented list of words. However, there was very little evidence of an increased irrelevant-speech effect in the ONH group in this experiment in contrast to previous studies that have shown a greater impact of irrelevant (and meaningful) speech on the word recall abilities of older listeners (Tun *et al.*, 2002; Bell *et al.*, 2008).

D. Influence of WM (LSPAN)

LSPAN scores were analyzed to determine the relationship between performance during speech recognition/recall tasks and WM capacity. Mean LSPAN scores were lower for ONH listeners compared to YNH listeners, suggesting that the YNH listeners in this study have a larger WM capacity than the ONH listeners. Results from the LSPAN generally did not correlate with speech recognition performance on the 0-back task. For both groups, LSPAN did not correlate significantly with SRT performance in the 0-back task in any masker or sentence context with the exception of a relatively weak correlation with the 1T masker for YNH listeners. Also, in the LSPAN region where data were available for both ONH and YNH listeners (i.e., LSPAN = 3), it is clear that the ONH group required substantially higher SNRs to achieve the same level of performance. These results suggest that older listeners have some difficulties with speech recognition in noise that cannot be predicted simply from the reduced WM capacity in this group.

As would be expected from the nature of the task, performance in the 1-back task was much more correlated with WM capacity as measured by the LSPAN task than were the SRT₈₀ values in the 0-back condition. In fact, in the 2T spatial and noise masker conditions, the data in Fig. 5 suggest that most of the performance deficits observed in the 1-back task with the ONH listeners could be explained by the lower LSPAN values obtained in that group. The performance deficits on the 1-back task of ONH listeners with relatively low LSPAN scores are consistent with the ELU model prediction

that those with declines in WM capacity will have more difficulty on more challenging recall tasks. However, the results in the conditions where the target talker was masked by one or two co-located speech maskers (1T and 2T) suggest that the ONH listeners performed worse in the 1-back task than would be predicted for YNH listeners with similar LSPAN scores. This suggests that older listeners may have additional difficulties in tasks that combine a high level of demand on WM with the requirement to segregate speech signals from co-located speech maskers. This result may be related to earlier results suggesting that co-located speech maskers require a higher level of listening effort to achieve the same level of performance than other types of maskers. For example, in a previous study that used pupil dilation measures to examine listening effort across different types of maskers (Koelewijn *et al.*, 2012), the results showed that larger pupil size, which reflects increased listening effort, was greater in the single-talker masker condition compared to conditions with stationary or fluctuating noise maskers.

E. Listening effort in immediate and delayed recall tasks

Listening effort has been defined as the attention and cognitive resources required to understand speech (Tun *et al.*, 2009). Several indices of listening effort have been reported in the literature, including pupillometry (Zekveld *et al.*, 2010; Zekveld *et al.*, 2011), self-report via questionnaire (e.g., the NASA Task Load Index, Hart and Staveland, 1988), and use of a dual-task paradigm (e.g., Tun *et al.*, 2009; Gosselin and Gagne, 2011; Desjardins and Dougherty, 2013). Numerous studies generally confirm that listening effort (or processing load) increases with decreases in SNR (Kramer *et al.*, 1997), decreases with the use of noise reduction algorithms (Sarampalis *et al.*, 2009; Ng *et al.*, 2014), and is higher among older than younger listeners under equivalent performance level conditions (Desjardins and Dougherty, 2013).

In the current experiment, we showed that adding a memory component to the speech recall task (i.e., 1-back task) resulted in a substantial decrease in speech understanding scores relative to the immediate speech recognition task (i.e., 0-back task). That is, in the 0-back task, SNR scores were adjusted such that recognition performance of all listener groups was equated to 80% correct. The use of fixed SNRs yielding equivalent speech recognition scores and the addition of the 1-back task resulted in clear performance declines relative to the 0-back task by both age groups in high and low sentence contexts across four masker types. These results can be interpreted as reflecting an increase in listening effort with the addition of a secondary memory task during sentence recall. Thus the unique method described in this paper represents a novel technique for measuring listening effort. Of equal importance is that this novel method of adding a secondary WM task is closely linked to the processing demands placed on listeners in everyday communication in noise when the listener must hold a message in memory for later recall and an appropriate response.

V. SUMMARY

The current findings show that ONH listeners are more adversely affected than YNH listeners in a 1-back task that required them to remember the content of each sentence until after the next sentence in the sequence was presented. This was true even when the sentences contained contextual cues that made them easier to understand in noise and to hold in memory across stimulus presentations and when the overall SNR of the stimuli was adjusted to produce equivalent performance in the two groups in a simple 0-back speech recognition task. Age-related declines in WM contribute to speech perception difficulties experienced by older individuals in the delayed recall task. The addition of a secondary WM task underscored these difficulties experienced by ONH individuals. The results are consistent with the notion that age-related cognitive decline has a significant impact on speech understanding performance in challenging listening environments that include retaining sentence-length information for later recall as would be required in everyday conversation. The findings are also generally in agreement with predictions of the ELU Model (Ronnberg *et al.*, 2013) that those with low WM capacity perform more poorly than individuals with higher WM capacity in challenging speech recognition tasks in noise, including in fluctuating noise and low context conditions.

Unfortunately, the results of this study suggest that relative performance in complex, WM-intensive speech perception tasks may not always be predictable from the simplified immediate-recall tasks that are typically used to assess speech perception in laboratory and clinical environments. Even when the SNR values were adjusted to equalize performance in the 0-back conditions across all the combinations of listener age, context level, and masker type tested in this experiment, there were substantial differences in the performance levels obtained in the corresponding 1-back conditions. And while much of the difference in performance between the ONH and YNH listeners in the 1-back task could be explained by the LSPAN measure of WM in the listening conditions that produced a relatively low level of informational masking (noise and 2T spatial), it was apparent that the ONH listener group performed much more poorly than would be predicted from the LSPAN score alone in the co-located speech masking conditions that produced a high level of informational masking.

A. Clinical implications

The finding that ONH listeners performed more poorly in the delayed recall (1-back) task than could be predicted by performance in the immediate recall (0-task) task highlights an observation that represents both a problem and an opportunity for the assessment of speech perception among hearing-impaired listeners. Communication in real-world listening situations often involves remembering a spoken message over time and responding appropriately. Consequently, WM and other cognitive processes (e.g., selective attention) are often required in everyday listening tasks. However, most clinical speech perception tests use simple immediate recall tasks that do not capture the complexity of everyday

speech. This dichotomy may be one reason why many patients perform well on standard clinical speech perception tasks in noise but report difficulty in real-world situations. The results from this study suggest that the use of complex tasks like the 1-back task described here might provide more realistic listening challenges for evaluating speech recognition in noise and be more sensitive to the speech understanding problems of older listeners, both with and without hearing loss. These kinds of tasks could also be used in the evaluation of hearing aid algorithms and other rehabilitation strategies for hearing-impaired listeners where such tests might be successful in uncovering subtle differences in performance that would not be apparent in standard speech-in-noise testing but would have a significant impact on user satisfaction in the more effortful situations that listeners typically encounter in their everyday lives.

Also, while it is true that the 1-back task may represent a somewhat extreme WM task compared to what listeners would normally experience in their everyday lives, it is worth noting that it has some advantages over other types of tasks that have been used to assess the impact of listening effort on speech perception, in particular, those involving dual-task paradigms (Tun *et al.*, 2009; Gosselin and Gagne, 2011). Dual-task paradigms can be very effective at increasing the cognitive load of the listener, but their results can be difficult to interpret unless great care is taken to ensure that performance in the primary task remains constant across all the conditions tested. In contrast, the 1-back task provides a speech perception test that requires a high level of listening effort but produces only a single unified output variable that effectively incorporates both the primary task (the perception of speech in noise) and the secondary task (remembering the content of the speech across stimulus presentations). Moreover, the results shown here suggest that the interleaved 0-back and 1-back task used in this experiment could provide a relatively efficient way to identify small differences across stimulus conditions in the delayed response task. Thus if a stimulus manipulation such as a new hearing aid algorithm could be found that produced no improvement in speech reception threshold in a traditional 0-back task but produced a consistent improvement in performance in the 1-back task, we believe that a compelling argument could be made that this algorithm would be likely to have real-world benefit in the kinds of listening environments where ONH listeners often complain that they have difficulty. Further research is needed to determine which, if any, stimulus parameters could be manipulated to achieve this desirable result.

ACKNOWLEDGMENTS

Funding in part was provided by U.S. Army Public Health Command in support of the Army Hearing Program. The opinions and assertions presented are the private views of the authors and are not to be construed as official or as necessarily reflecting the views of the Department of the Army, Department of the Navy, Department of the Air Force, the Department of Defense, or the U.S. Government. Funding was also provided in part by the University of Maryland MCM Fund for Student Research Excellence. The

authors thank Chelsea Vogel for her assistance with recording of stimuli and data collection.

- Anderson, S., Parbery-Clark, A., White-Schwoch, T., and Kraus, N. (2012). "Aging affects neural ion of speech encoding," *J. Neurosci.* **32**, 14156–14164.
- ANSI (2010). ANSI S3.6-2010, "American National Standard Specification for Audiometers (Revision of ANSI S3.6-1996, 2004)," American National Standards Institute, New York.
- Baddeley, A. (1992). "Working memory," *Science* **255**, 556–559.
- Başkent, D., van Engelshoven, S., and Galvin, J. J. III (2014). "Susceptibility to interference by music and speech maskers in middle-aged adults," *J. Acoust. Soc. Am.* **135**, EL147–153.
- Bell, R., Buchner, A., and Mund, I. (2008). "Age-related differences in irrelevant-speech effects," *Psychol. Aging* **23**, 377–391.
- Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2013). "How linguistic closure and verbal working memory relate to speech recognition in noise—a review," *Trends Amplif.* **17**, 75–93.
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (1984). "Standardization of a test of speech perception in noise," *J. Speech Lang. Hear. Res.* **27**, 32–48.
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., and Noll, D. C. (1997). "A parametric study of prefrontal cortex involvement in human working memory," *Neuroimage* **5**, 49–62.
- Brungart, D. S., Romigh, G., and Simpson, B. D. (2011). "Rapid collection of head related transfer functions and comparison to free-field listening," in *Principles and Applications of Spatial Hearing*, edited by Y. Suzuki, D. S. Brungart, Y. Iwaya, K. Iida, D. Cabrera, and H. Kato (World Scientific, Hackensack, NJ), pp. 139–148.
- Brungart, D. S., and Simpson, D. B. (2002). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," *J. Acoust. Soc. Am.* **112**, 664–676.
- Brungart, D. S., Simpson, D. B., Ericson, A. M., and Scott, R. K. (2001). "Informational masking and energetic masking effects in the perception of multiple simultaneous talkers," *J. Acoust. Soc. Am.* **110**, 2527–2538.
- Committee on Hearing Bioacoustics and Biomechanics (1988). "Speech understanding and aging," *J. Acoust. Soc. Am.* **83**, 859–895.
- Daneman, M., and Carpenter, P. A. (1980). "Individual differences in working memory and reading," *J. Verb Learn. Verb Behav.* **19**, 450–466.
- Desjardins, J. L., and Doherty, K. A. (2013). "Age-related changes in listening effort for various types of masker noises," *Ear Hear.* **34**, 261–272.
- Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (2000). "Use of context by younger and older adults with normal hearing," *J. Acoust. Soc. Am.* **107**, 538–546.
- Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (2002). "Spectral contributions to the benefit from spatial separation of speech and noise," *J. Speech Lang. Hear. Res.* **45**, 1297–1310.
- Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (2008). "Binaural advantage for younger and older adults with normal hearing," *J. Speech Lang. Hear. Res.* **51**, 539–556.
- Dubno, J. R., Dirks, D. D., and Morgan, D. E. (1984). "Effects of age and mild hearing loss on speech recognition in noise," *J. Acoust. Soc. Am.* **76**, 87–96.
- Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.
- Gates, G. A., Feeney, M. P., and Mills, D. (2008). "Cross-sectional age-changes of hearing in the elderly," *Ear Hear.* **29**, 865–874.
- Goncalves, V. T., and Mansur, L. L. (2009). "N-back auditory test performance in normal individuals," *Dement. Neuropsychol.* **2**, 114–117.
- Gordon-Salant, S., and Fitzgibbons, P. J. (1997). "Selected cognitive factors and speech recognition performance among young and elderly listeners," *J. Speech Lang. Hear. Res.* **40**, 423–429.
- Gosselin, P. A., and Gagne, J.-P. (2011). "Older adults expend more listening effort than young adults recognizing speech in noise," *J. Speech Lang. Hear. Res.* **54**, 944–958.
- Gottlob, L. R., and Madden, D. J. (1998). "Time course of allocation of visual attention after equating for sensory differences: An age-related perspective," *Psychol. Aging* **13**, 138–149.
- Hart, S. G., and Staveland, L. E. (1988). "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Human Mental Workload*, edited by P. A. Hancock and N. Meshkati (North-Holland Press, Amsterdam).
- Helfer, K. S., Chevalier, J., and Freyman, R. L. (2010). "Aging, spatial cues, and single-versus dual-task performance in competing speech perception," *J. Acoust. Soc. Am.* **128**, 3625–3633.
- Helfer, K. S., and Freyman, R. L. (2008). "Aging and speech-on-speech-masking," *Ear Hear.* **29**, 87–98.
- Herman, R., and Pisoni, D. B. (2003). "Perception of 'elliptical speech' following cochlear implantation: Use of broad phonetic categories in speech perception," *Volta Rev.* **4**, 321–347.
- Humes, L. E., and Dubno, R. J. (2010). "Factors affecting speech understanding in older adults," in *The Aging Auditory System*, edited by S. Gordon-Salant, R. D. Frisina, A. N. Popper, and R. R. Fay (Springer, New York), Chap. 8, pp. 211–258.
- Humes, L. E., Lee, J. H., and Coughlin, M. P. (2006). "Auditory measures of selective and divided attention in young and older adults using single-talker competition," *J. Acoust. Soc. Am.* **120**, 2926–2937.
- Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2012). "Pupil dilation uncovers extra listening effort in the presence of a single-talker masker," *Ear Hear.* **33**, 291–300.
- Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2014). "The influence of informational masking on speech perception and pupil response in adults with hearing impairment," *J. Acoust. Soc. Am.* **135**, 1596–1606.
- Kramer, S., Kapteyn, T., Festen, J., and Kuik, D. (1997). "Assessing aspects of auditory handicap by means of pupil dilatation," *Audiology* **36**, 155–164.
- Lunner, T. (2003). "Cognitive function in relation to hearing aid use," *Int. J. Audiol.* **42**, 49–58.
- Ng, E. H., Rudner, M., Lunner, T., and Rönnberg, J. (2014). "Noise reduction improves memory for target language speech in competing native but not foreign language speech," *Ear Hear.* (in press).
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., and Smith, P. K. (2002). "Models of visuospatial and verbal memory across the adult life span," *Psychol. Aging* **17**, 299–320.
- Pichora-Fuller, M. K. (2003). "Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension," *Int. J. Audiol.* **42**, 59–67.
- Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). "How young and old adults listen to and remember speech in noise," *J. Acoust. Soc. Am.* **97**, 593–676.
- Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., and Rudner, M. (2013). "The Ease of Language Understanding (ELU) Model: Theoretical, empirical, and clinical advances," *Front. Neurosci.* **7**, 1–17.
- Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (2009). "Objective measures of listening effort: Effects of background noise and noise reduction," *J. Speech Lang. Hear. Res.* **52**, 1230–1240.
- Singh, G., Pichora-Fuller, K., and Schneider, B. A. (2008). "The effect of age on auditory spatial attention in conditions of real and simulated spatial separation," *J. Acoust. Soc. Am.* **124**, 1294–1305.
- Sörqvist, P., and Rönnberg, J. (2012). "Episodic LTM of spoken discourse masked by speech: What is the role for working memory capacity," *J. Speech Lang. Hear. Res.* **55**, 210–218.
- Tun, P. A., McCoy, S., and Wingfield, A. (2009). "Aging, hearing acuity, and the attentional costs of effortful listening," *Psychol. Aging* **24**, 761–766.
- Tun, P. A., O'Kane, G., and Wingfield, A. (2002). "Distraction by competing speech in young and older adult listeners," *Psychol. Aging* **17**, 453–467.
- Tun, P. A., and Wingfield, A. (1999). "One voice too many: Adult age differences in language processing with different types of distracting sounds," *J. Gerontol. B Psychol. Sci. Soc. Sci.* **54B**, P317–P327.
- Weeks, J. C., and Hasher, L. (2014). "The disruptive—and beneficial—effects of distraction on older adults' cognitive performance," *Front. Psychol.* **5**, 1–6.
- Wingfield, A., Poon, L. W., Lombardi, L., and Lowe, D. (1985). "Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time," *J. Gerontol.* **40**, 579–585.
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2010). "Pupil response as an indication of effortful listening: The influence of sentence intelligibility," *Ear Hear.* **31**, 480–490.
- Zekveld, A. A., Rudner, M., Johnsrude, I. S., Festen, J. M., Van Beek, J. H. M., and Rönnberg, J. (2011). "The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise," *Ear Hear.* **32**, E16–325.
- Zekveld, A., Rudner, M., Kramer, S. E., Lyzenga, J., and Rönnberg, J. (2014). "Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech," *Front. Neurosci.* **8**, 1–11.
- Zelazo, P. D., Craik, F. I., and Booth, L. (2004). "Executive function across the life span," *Acta Psychol.* **115**, 167–183.