

Aging and the effect of target-masker alignment

Karen S. Helfer,^{a)} Gabrielle R. Merchant, and Richard L. Freyman

Department of Communication Disorders, University of Massachusetts Amherst, 358 North Pleasant Street, Amherst, Massachusetts 01003, USA

(Received 11 April 2016; revised 5 October 2016; accepted 25 October 2016; published online 21 November 2016)

Similarity between target and competing speech messages plays a large role in how easy or difficult it is to understand messages of interest. Much research on informational masking has used highly aligned target and masking utterances that are very similar semantically and syntactically. However, listeners rarely encounter situations in real life where they must understand one sentence in the presence of another (or more than one) highly aligned, syntactically similar competing sentence(s). The purpose of the present study was to examine the effect of syntactic/semantic similarity of target and masking speech in different spatial conditions among younger, middle-aged, and older adults. The results of this experiment indicate that differences in speech recognition between older and younger participants were largest when the masker surrounded the target and was more similar to the target, especially at more adverse signal-to-noise ratios. Differences among listeners and the effect of similarity were much less robust, and all listeners were relatively resistant to masking, when maskers were located on one side of the target message. The present results suggest that previous studies using highly aligned stimuli may have overestimated age-related speech recognition problems. © 2016 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4967297>]

[MSS]

Pages: 3844–3853

I. INTRODUCTION

The similarity between a target speech message and competing speech maskers plays a large role in how easy or difficult it is to understand the target utterance. Listeners can use features such as fundamental frequency, level, spatial location, language, and timing to help identify and understand a speech message embedded in a background of other speech messages. In general, the more similar the target and masker, the greater the amount of both energetic and informational masking (e.g., Brungart, 2001; Freyman *et al.*, 2001; Darwin, 2008).

Many previous studies of speech-on-speech masking have maximized informational masking by using target and masking speech that is highly similar in terms of both timing and syntax (for example, the CRM sentences: Bolia *et al.*, 2000). This research has provided strong evidence of age-related decline in speech understanding under these conditions (e.g., Humes *et al.*, 2006; Marrone *et al.*, 2008; Singh *et al.*, 2008; Humes and Coughlin, 2009; Neher *et al.*, 2009; Rossi-Katz and Arehart, 2009; Helfer *et al.*, 2010; Lee and Humes, 2012; Singh *et al.*, 2013; Helfer and Freyman, 2014; Xia *et al.*, 2015). Although using this type of experimental paradigm allows tight control over the stimuli, the listening environment it creates is artificial. Individuals rarely encounter situations in real life in which they must understand one sentence in the presence of another (or more than one) highly aligned, syntactically similar competing sentence(s).

Similarity may take on even more importance for older adults. Data from some studies suggest that they are more

greatly affected than younger adults when target and masking speech are highly confusable. For example, older individuals show a greater difference in performance (as compared to young, normally hearing adults) between time-reversed and normally presented maskers (e.g., Rossi-Katz and Arehart, 2009; Helfer *et al.*, 2010) or between maskers spoken in a foreign language versus those spoken in the listener's native language (Tun *et al.*, 2002). This leads to the motivation for the present work: to measure how syntactic and semantic target/masker alignment affects speech understanding in listeners spanning the adult age range. The experiment described in this paper compared speech recognition when the target and masker were similar in terms of both syntax and semantics to conditions where the masker was manipulated so that word order differed between masker and target sentences. The importance of examining this issue is supported by results of a study in our laboratory (Helfer and Freyman, 2014), which found that self-assessed listening problems were significantly correlated with objective speech perception in the presence of running speech, but not with scores obtained in the presence of a temporally aligned speech masker.

Since individuals rarely listen to co-located voices, the present experiment used two types of spatially separated conditions: competing speech messages on one side of the target talker and speech maskers that surround the target message. Angular separation between target and masking speech is important to consider because it provides a strong cue for reducing both energetic and informational masking (see Bronkhorst, 2015, for a review of relevant studies). When the target speech is in front of the listener and the masker is to one side, spatial release from masking (SRM)

^{a)}Electronic mail: khelfer@comdis.umass.edu

occurs, in part due to listeners' ability to use the ear on the opposite side of the masker, which has a better signal-to-noise ratio (SNR). Although age-related change in these types of listening situations has been the focus of a considerable number of studies, there is little consensus about whether the observed reduction in the ability to use spatial cues is due to age-related hearing loss or to other factors related to aging (e.g., Duquesnoy, 1983; Dubno *et al.*, 2008; Helfer and Freyman, 2008; Helfer *et al.*, 2010).

When speech maskers are placed symmetrically around the target, SRM can still ensue even though neither ear has a long-term SNR advantage (e.g., Bronkhorst and Plomp, 1992; Peissig and Kollmeier, 1997; Hawley *et al.*, 1999; Marrone *et al.*, 2008; Brungart and Iyer, 2012). SRM in these situations most likely occurs, at least in part, because uncorrelated fluctuations from two different maskers lead to moments where the SNR is better in the left ear and other times when the SNR is better in the right ear. Listeners can use these brief, unpredictable better-ear glimpses to understand speech within the masker complex (Brungart and Iyer, 2012; Best *et al.*, 2015). Older adults with hearing loss appear to have reduced SRM when maskers are symmetrically placed around the target (e.g., Marrone *et al.*, 2008; Dawes *et al.*, 2013; Glyde *et al.*, 2013b; Besser *et al.*, 2014). Although results generally support the contention that age-related hearing loss drives this reduction in SRM with age (Marrone *et al.*, 2008; Neher *et al.*, 2009; Glyde *et al.*, 2013b), there is reason to believe that it cannot entirely explain differences between younger and older adults in the use of spatial cues in symmetric masking conditions. For example, using an *n*-back task, Schurman *et al.* (2014) found that older adults with clinically normal hearing showed reduced SRM in a symmetric masking condition, relative to younger adults. Results of Best *et al.* (2015) indicated that younger adults, both with and without hearing loss, can effectively use better-ear glimpsing in symmetric masking conditions, suggesting that hearing loss *per se* is not the cause of the decline in SRM in older adults. Additional research supports the idea that the ability to use brief, rapidly changing glimpses may be affected by age-related variations other than hearing loss, including changes in temporal processing or general slowing of processing speed (e.g., Neher *et al.*, 2009; Dawes *et al.*, 2013; Gallun *et al.*, 2013). These findings collectively suggest that factors beyond hearing loss limit older adults' use of spatial cues when the masker surrounds the target.

The experiment described in this paper was designed to identify how age-related changes in hearing and selected cognitive abilities modify the impact of syntactic/semantic alignment cues in two types of spatially separated conditions. As stated earlier, it is possible that speech perception differences between older and younger listeners are exaggerated when the target and masking speech are aligned (and therefore highly confusable). We speculate that alignment effects will be particularly apparent when the masker surrounds the target, since listeners may need to more heavily rely on misalignment as a cue to segregate target from masking speech in the symmetric masking conditions. Also of interest was examining how cognition influences speech

recognition. Segregating a target speech signal that is embedded in maskers from both sides is likely to be more difficult and require more cognitive processing, as compared to teasing out a target message from maskers that are located only on one side. We therefore anticipated that cognitive skills measured in the present study would be more important for explaining performance when the maskers are located to the right and left of the target, versus when they come from only one side.

II. METHOD

A. Participants

Three groups of listeners ($n = 17/\text{group}$) participated in this study. Younger (20–25 yrs, mean 22 yrs) adults who were native speakers of English were recruited from the undergraduate population of the University of Massachusetts. They were required to have normal pure-tone thresholds and no history of significant otologic or neurologic disorder. Participants were also required to have bilaterally normal (type A) tympanograms. The high-frequency pure-tone average (HFPTA; the average of thresholds at 2 kHz, 3 kHz, 4 kHz, and 6 kHz) for these individuals ranged from -1 dB hearing level (HL) to 8 dB HL (mean = 5 dB HL).

Groups of middle-aged and older participants also were tested. These individuals had a negative history of significant otologic or neurological problems and were native English speakers. They also were required to have bilaterally normal (type A) tympanograms. In order to participate, middle-aged and older participants needed to score at least 26 out of 30 points on the Mini-Mental State Exam (Folstein *et al.*, 1975). By design, we accepted individuals with a range of pure-tone thresholds, although they were required to have HFPTA no greater than 60 dB HL and symmetric pure-tone thresholds. Participants in the middle-aged group were 41–59 yrs (mean 53 yrs), with HFPTA of 1 dB HL–28 dB HL (mean = 13 dB HL). Individuals in the older group were 61–85 yrs (mean = 69 yrs) with HFPTA of 8 dB HL–56 dB HL (mean = 31 dB HL). Composite audiograms for these two groups are shown in Fig. 1.

Each participant completed a battery of cognitive tasks. Selection of specific tasks was driven by previous studies documenting connections between cognitive skills and speech understanding in complex listening situations (e.g., Jesse and Janse, 2012; Koelweijn *et al.*, 2012; Neher *et al.*, 2012; Woods *et al.*, 2013; Helfer and Freyman, 2014; Fullgrabe *et al.*, 2015; Helfer and Jesse, 2015). The cognitive test battery included measurement of visual working memory (SICSPAN: Sorqvist *et al.*, 2010); inhibitory ability (a computerized Stroop task: Jesse and Janse, 2012); processing speed (Connections test: Salthouse *et al.*, 2000); and attention switching/executive function (Visual Elevator task from the Test of Everyday Attention: Robertson *et al.*, 1996). A complete description of each task can be found in previous publications (Helfer and Freyman, 2014; Helfer and Jesse, 2015). Table I shows group performance on these tasks. A one-way analysis of variance (ANOVA) for each cognitive test found significant group differences for the Visual

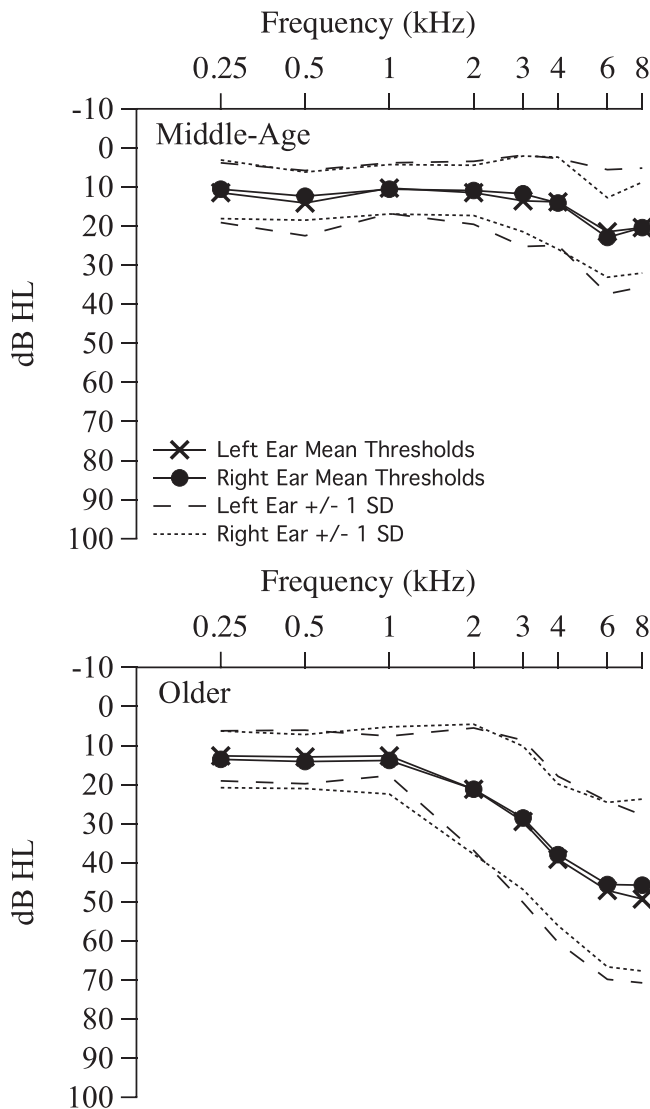


FIG. 1. Composite audiograms for the middle-aged and older groups. Dotted and dashed lines represent one standard deviation from the mean of obtained thresholds for the right and left ear, respectively.

Elevator task [$F(2,48) = 4.64, p = 0.014$] and for the SICSPAN [$F(2,48) = 3.68, p = 0.033$]. *Post hoc* Bonferroni contrasts (adjusted for multiple comparisons) indicated that the older participants had poorer performance than the

TABLE I. Cognitive task performance by each group. Values in parentheses represent the standard error. SICSPAN = number of items recalled in the correct order; Stroop = normalized Stroop effect, defined as the difference in mean response time in the incongruent and neutral condition divided by the mean response time in the neutral condition; Connections = ratio of mean performance (number of correct connections on the alternating trials divided by mean performance on simple trials); Visual elevator = average time per direction switch (in seconds). Values in parentheses are the standard error.

	Younger	Middle-aged	Older
SICSPAN	28.36 (1.44)	26.41 (1.98)	21.65 (1.92)
Stroop	-0.21 (0.02)	-0.20 (0.02)	-0.20 (0.02)
Connections	0.53 (0.03)	0.54 (0.03)	0.59 (0.04)
Visual elevator	3.43 (0.18)	3.93 (0.24)	4.52 (0.33)

younger participants on both of these tasks. Performance of the middle-aged individuals did not differ significantly from that of either the younger or the older participants on any of the cognitive measures.

B. Speech understanding

1. Stimuli

Speech perception was measured using a newly developed set of sentences that we call the TVM Colors (TVM-C) corpus. Each of these sentences begins with the cue name Theo, Victor, or Michael (hence “TVM”). The sentences take the form “Cue name found the color noun and the adjective noun here,” where the underlined words are used for scoring. An example is “Theo found the white pad and the helpful study here.” Colors were one of eight one-syllable words; nouns and adjectives were commonly encountered one- and two-syllable words, primarily from the Thorndike-Lorge lists (Thorndike and Lorge, 1952). Nouns and adjectives were never repeated within the corpus. Each sentence had 13 syllables. When developing these sentences, our goal was for each of the two nouns in the sentence to be feasible but not highly predictable based on the rest of the sentence context. In order to verify this lack of predictability, 38 college-aged adults (separate from those who participated in this study) completed a Cloze task (Taylor, 1953). Each of the 337 sentences was evaluated twice (once with the first noun removed, and a second time with the second noun removed). The participants read sentences with one of the nouns missing and were instructed to fill in a word that would best complete the sentence. Cloze probability was defined as the proportion of times the correct word was inserted into the sentence context. The average probability was 0.2% and ranged from 0% to 10.5%. As low probability is considered 0%–33% (Block and Baldwin, 2010), our goal of developing a set of sentences in which nouns were not highly predictable was achieved.

Each of the sentences (337 in total) was audio recorded from 6 talkers (3 male, 3 female) with no discernable regional dialect who were instructed to say the sentences in a natural (that is, not intentionally clear) manner. Some of the talkers noticeably paused between the first and second half of the sentence. In order to increase consistency, pauses at this juncture in the sentences were edited so that the time interval between the first noun and the word “and” was no more than 0.25 s. Each sentence had the same number of syllables, although there was some variation in sentence length (mean length of all recorded sentences = 3.06 s, standard deviation = 0.25 s). To verify that each sentence was intelligible, recordings from each talker were played in quiet to three young, normally hearing listeners. Sentences in which all five scoring words were not perceived with 100% accuracy were either re-recorded or eliminated.

2. Procedure

Each participant completed audiometric and cognitive testing just prior to measuring speech understanding. All participants completed testing in one visit lasting 1.5–2 h.

During speech perception testing, participants were seated in a double-walled IAC Acoustics test room (#1604A). On each trial, three sentences were presented simultaneously, all from three different talkers of the same gender. All five keywords (cue name, color, adjective, and both nouns) differed for each of the three sentences presented during a given trial. The target sentence was always presented from a loudspeaker directly in front of the listener, at a distance of 1.3 m and at approximately ear-level height. There were three spatial conditions: two masking sentences presented from a loudspeaker located 60 degrees to the right (RR); two masking sentences from a loudspeaker 60 degrees to the left (LL); and one masking sentence from each of these loudspeakers (RL).

There also were two masker alignment conditions, which are depicted in Fig. 2. In aligned conditions, the maskers were TVM-C sentences that began at the beginning of a sentence. It should be noted that, because of variations in where one- and two-syllable words occurred within sentences and minor differences in speaking rate among talkers, alignment was maximal at the beginning of trials (see left panel of Fig. 2). For non-aligned trials, the maskers (also TVM-C sentences) began at a random point within the sentence. The computer program controlling the experiment randomly chose a sampling point within the masker sentence file, started the sentence at this point, then appended the beginning of the sentence to the end of the sentence, with a 7 msec sigmoidal rise and fall time at the beginning and end of the sentence (to avoid clicks). One advantage of this method, rather than using a running speech masker, is that it allowed for easier examination of masker errors (that is, when the participant responded with a word from a masker rather than a word from the target) since maskers were always discrete sentences. It also allowed us to directly compare performance with the masker and target having the same (for aligned conditions) and different (for non-aligned conditions) syntactic order. It should be noted that there was an approximate 400 msec gap within the

non-aligned masker at the point where the beginning of the masking sentence was appended, as each sentence file included approximately 200 msec of silence at both the beginning and the end of the file. This led to the masker often beginning before the target in non-aligned trials, since the target (but not the maskers) began and ended with this brief period of silence. Another factor to be considered is that the masker often started mid-word at the beginning of non-aligned trials. Both of these aspects of the non-aligned trials could have reduced confusion between masker and target, thereby leading to less informational masking.

Target sentences were presented at a root-mean-square (RMS) of 68 dB sound pressure level (SPL) using 3 SNRs: -6 dB, -3 dB, and 0 dB, leading to 18 conditions (3 spatial conditions \times 2 alignment conditions \times 3 SNRs). The SNRs were expressed in relation to the total masker energy. For example, in the 0 dB SNR condition, the combined energy of the maskers was equal to the level of the target. Ten sentences (50 scoring words) were presented per condition. Participants were told to repeat the sentence coming from the front loudspeaker and to ignore the sentences coming from the side loudspeaker(s). Before the task began, participants completed a brief practice session to familiarize them with the procedure. Practice consisted of six trials in the presence of the non-aligned masker (two trials in each of the three spatial conditions). Correct-answer feedback was not provided, although occasionally participants needed reinforcement of instructions regarding the task during the practice trials. Participant responses were audio recorded and analyzed off-line.

III. RESULTS

A. Speech perception: Overall accuracy

Percent-correct performance on the speech recognition task, averaged across all five scored words in each sentence,

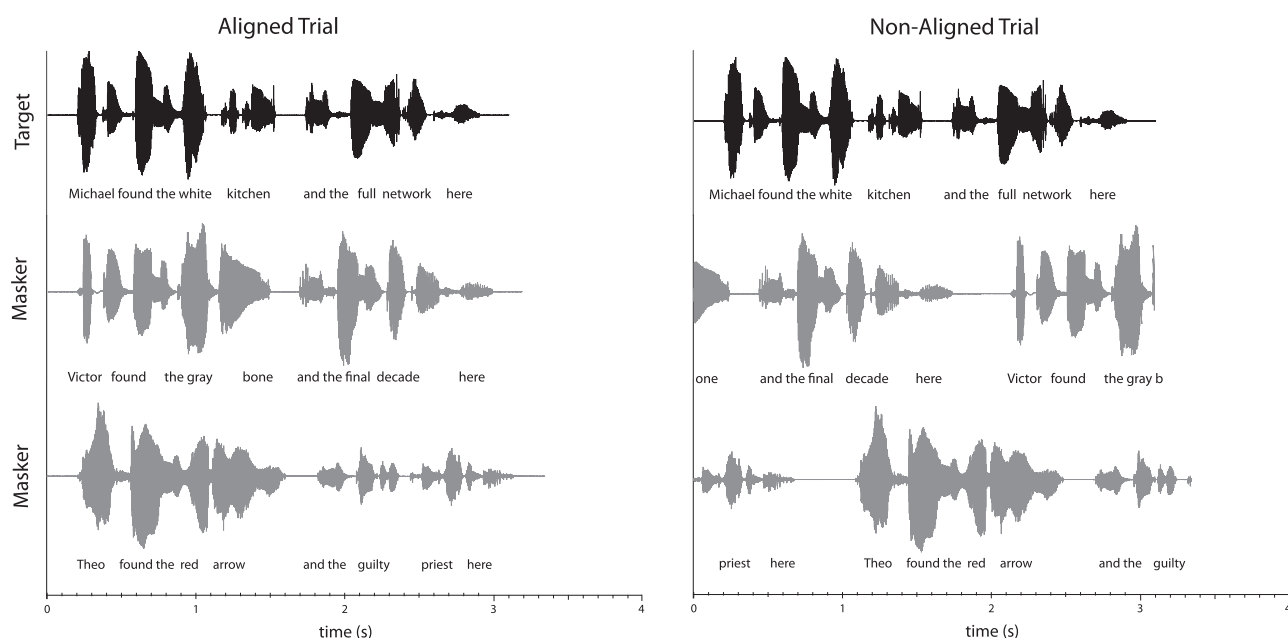


FIG. 2. Schematic depiction of the target and maskers during aligned and non-aligned trials.

is shown in Fig. 3. It is obvious that performance in the asymmetric condition, when both maskers were on one side (LL and RR; darker lines and solid symbols in Fig. 3) was quite high (ranging from 77% to 92%), especially for the younger and middle-aged participants. There were minimal differences between the aligned and non-aligned maskers in this spatial condition. In the symmetric condition, when the masker surrounded the target (RL; Fig. 3, lighter lines and open symbols), performance declined substantially for all three groups, dropping to as low as 31% correct in older participants. Moreover, unlike in the asymmetric spatial condition, there was a difference between performance in aligned and non-aligned trials in the presence of symmetric maskers.

The percent-correct scores measured in each condition were transformed to rationalized arcsine units (RAU; Studebaker, 1985) then subjected to a multivariate ANOVA with spatial condition (RR versus LL versus RL), masker alignment (aligned versus non-aligned) and SNR as within-subjects variables and group as a between-subjects variable. Results of this analysis showed significant main effects for all four variables as well as significant two-, three-, and four-way interactions. Of particular interest were the significant interactions involving group: Spatial Condition \times SNR \times Group [$F(8,96)=2.08, p=0.039$], Spatial Condition \times Alignment \times Group [$F(4,96)=3.20, p=0.016$], and the four-way interaction [$F(8,192)=2.44, p=0.016$]. *Post hoc* one-way ANOVAs (adjusted for multiple comparisons using Bonferroni corrections) found that when the maskers were on one side (LL or RR), the only significant group difference was between younger and older adults at -6 dB SNR for non-aligned maskers. When the maskers were symmetrically placed around the target, older adults performed significantly poorer than younger adults for all conditions, except for -6 dB SNR for non-aligned maskers. Scores obtained by middle-aged adults were significantly higher than those obtained by older adults for -6 dB SNR and 0 dB SNR for aligned maskers and at -3 dB SNR for non-aligned maskers (all in the symmetric masking condition). Performance did not differ significantly between younger and middle-aged participants in any condition.

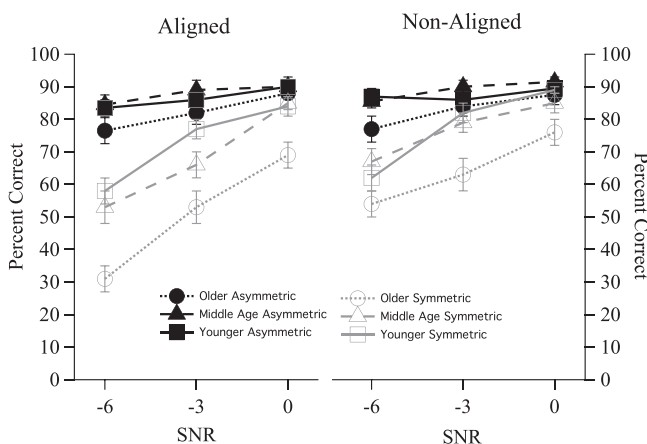


FIG. 3. Percent-correct performance on the speech recognition task averaged across the five scored words in each sentence. Error bars show the standard error.

B. Difference scores

Of primary interest in this study was comparing the difference in speech perception obtained with aligned versus non-aligned targets/maskers across groups, as well as examining performance differences when maskers were located on one side versus when they surrounded the target. We calculated simple difference scores for both of these factors (see Fig. 4). The top panel of Fig. 4 displays difference scores for alignment by SNR for each group. Alignment had little effect when the maskers were located on one side of the listener (RR and LL). The fact that performance was high in the aligned condition likely contributed to the lack of effect, as there was not much room for improvement in the non-aligned condition. When the maskers came from both sides (RL), alignment did influence speech recognition, especially at more adverse SNRs and particularly for older and middle-aged listeners. As can be observed in Fig. 4, older and middle-aged adults were at a greater disadvantage from aligned target/maskers, as compared to younger adults. Repeated-measures ANOVA on these difference scores with spatial condition and SNR as within-subjects factors and group as a between-subjects factor indicated significant main effects for all factors as well as significant interactions, including the three-way interaction [$F(4,96)=3.79, p=0.007$].

The difference between side and symmetrically placed maskers is shown in the bottom panel of Fig. 4. Although listening in a situation where speech maskers are on both sides was considerably more difficult than with single-side

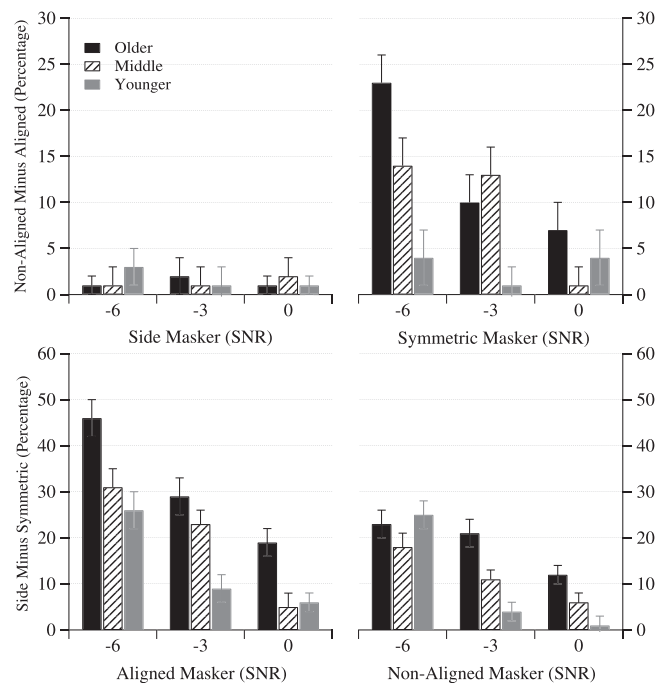


FIG. 4. (Top) Effect of alignment on speech recognition performance. Reported values were calculated as percent-correct in the non-aligned masker condition minus percent-correct in the aligned masker condition. Error bars show the standard error. (Bottom) Effect of spatial configuration of maskers on speech recognition performance. Reported values were calculated as percent-correct when maskers were on one side minus percent-correct when maskers were symmetric around the target. Error bars show the standard error.

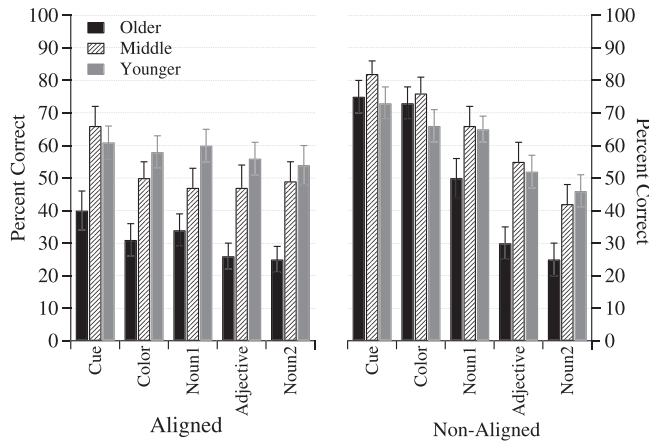


FIG. 5. Word-by-word analysis of responses for each word in the target sentences for data collected at -6 dB SNR in the symmetric (RL) masker condition. Error bars show the standard error.

maskers for all listeners, this difference was, in general, larger for older listeners than for younger listeners, with the exception of -6 dB SNR in the non-aligned masker. Moreover, for older and middle-aged participants, the effect of spatial configuration was generally larger for aligned maskers than for non-aligned maskers. Repeated-measures ANOVA on these difference scores led to similar results as the previous analysis on aligned/non-aligned difference scores, with significant main effects and interactions (three-way interaction: $[F(4,96) = 3.74, p = 0.007]$).

C. Word-by-word analysis

In order to more completely understand the percent-correct scores, we examined patterns of responses for each word in the target sentences for data collected at -6 dB SNR in the symmetric (RL) masker condition, as this was where the maximum number of errors occurred (see Fig. 5). There were substantial differences in patterns of performance between aligned and non-aligned maskers, especially for words at the beginning of the target sentence. Recall that the first two words in the target sentence were from a closed set of items (three possible choices for the cue name and eight possible options for the color). It might be expected that performance for these first two words would be better than that obtained for the last three words, which were not constrained

by set size. This pattern was only observed when the target and masker were not aligned. Identification of words within a sentence was fairly consistent across word position when the masker was aligned with the target. Hence, alignment of target and maskers was more detrimental to words at the beginning of a sentence than to words at the end of the sentence.

There are several possible contributors to this pattern of results. Masker alignment was greatest for words at the beginning of the sentences, as there were natural variations in timing of individual words in each sentence (as depicted in Fig. 2). Another possible contributing factor is that, during non-aligned trials, the maskers often began in the middle of a word. This cue could have helped the listener to differentiate the target words from the masker. Finally, since the stimuli were declarative sentences, words at the beginning of sentences were slightly more intense than words at the end of sentences. Analysis of a random sample of 30 sentences indicated that this difference in RMS amplitude was about 2 dB. During aligned trials, such variations in intensity would not have differentially affected words in different positions of the sentence, since the small decline in intensity would be, on average, the same for both the target and masker sentences. However, in non-aligned trials, masker words from the end of sentences (which were slightly less intense) often occurred during the early portion of the target sentence. This would be expected to lead to a slightly improved SNR for words at the beginning of the trial.

For non-aligned trials, differences in performance between older and all other participants were minimal for words at the beginning of the sentence and became larger as the sentence progressed. This pattern was not noted for aligned trials, where the deficit experienced by the older listeners was more consistent across word position.

We examined the types of errors made by listeners for aligned versus non-aligned trials in the symmetric spatial condition at -6 dB SNR for each word position except the cue name (since these were from a very closed set of three possible items). Each incorrect response was classified as one of four types of errors. Masker errors were when the participant reported either a word that was in one of the masker sentences or a word that was very similar phonetically to a masker word and very dissimilar to the target word; in Fig. 2, an example of a masker error would be if the

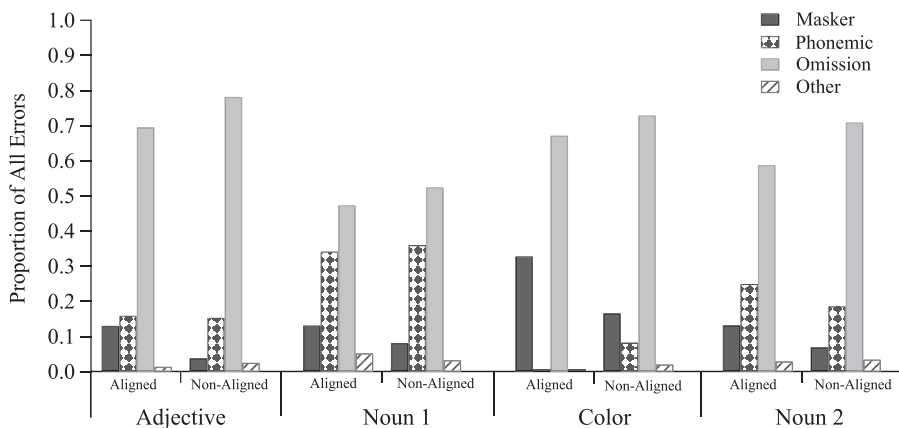


FIG. 6. Error patterns produced by participants in the symmetric masking condition at -6 dB SNR. Values represent the proportion of each error type relative to all errors.

TABLE II. Pearson r correlation coefficients for each of the cognitive tests. * = significant at the 0.05 level; ** = significant at the 0.01 level.

	Stroop	Connections	Visual elevator	SICSPAN
Stroop	—	0.18	-0.27	0.42**
Connections		—	0.02	0.11
Visual elevator			—	0.60**
SICSPAN				—

participant reported the word “decade.” Phonemic errors were incorrect approximations of the target word (in Fig. 2, an example would be the participant reporting “bull” instead of “full”). The other two error responses were omissions (where the participant did not attempt to guess) or unclassifiable responses that did not fall into any of the above categories.

Although, as described above, there were clear differences among groups in overall accuracy of performance, analysis of error patterns suggested only minimal group differences in terms of proportion of each type of error. We therefore averaged the error pattern data across all three listener groups. Figure 6 shows the proportion of each of these types of errors out of the total number of errors. It can be seen that there was a predominance of omission error responses for all words, with little effect of alignment on the proportion of these types of responses. There was a modest decrease in masker errors when the target and masker were not aligned, with a corresponding small increase in phonemic errors for non-aligned (versus aligned) trials, for all words except the last key word. Not surprisingly, masker errors were most common for color words, since these were drawn from a closed set of eight items.

D. Regression analyses

Stepwise regression analyses were used to predict percent-correct performance in each condition, averaged across all SNRs, using data from all 51 participants. We first examined our cognitive variables for inter-correlations.

These correlation coefficients are shown in Table II. Three of our four cognitive measures were significantly inter-correlated, leading us to compute a composite score by averaging the normalized (z -score) values for these three tasks (the score on the Connections test was retained as a separate variable). Predictor variables used in the regression analysis were the composite cognitive metric, the Connections score, age, and better-ear high-frequency average (HFPTA; average of thresholds between 2 kHz and 6 kHz). Results of these analyses are shown in Table III.

In the most difficult condition (symmetric/aligned maskers) both age and the composite cognitive score accounted for significant variance. Not surprisingly, performance in this condition was inversely related to participant age. Cognitive skills accounted for significant additional variance above and beyond age. Speech understanding with symmetric maskers that were not aligned was most strongly related to the cognitive composite score and to HFPTA. When the maskers were on one side, individuals with better cognitive task performance had higher speech perception ability, with neither age nor HFPTA contributing to the variance in these conditions. Performance on the Connections task did not explain additional variance in any of these analyses.

IV. DISCUSSION

The purpose of this work was to investigate age-related changes in the ability to use differences in syntactic/semantic alignment between the target and the masker to help overcome masking by competing speech. Much of the previous research in competing speech perception has utilized paradigms in which masking speech is highly similar, in terms of syntax and alignment, with the target speech. This raises the question of the generalizability of those results, as to-be-attended and to-be-ignored messages that are very similar in terms of timing and syntactic/semantic order is not representative of realistic listening situations. Indeed, when the masker surrounded the target spatially, differences in speech recognition between older and younger participants were greater for aligned trials, especially at more adverse SNRs.

TABLE III. Results of regression analysis on data from all 51 participants. Dependent variables were derived by averaging performance across all SNRs. Predictor variables entered into the equation were age, high-frequency pure-tone average (HFPTA), the composite score derived from three of the cognitive tasks (Cog), and performance on the Connections task (Con).

Variable	Significant predictors	r^2	β	t	Significance
Percent-correct in symmetric/ aligned masker	<i>Step 1</i>				
	Age	0.35	-0.59	-5.08	<0.001
	<i>Step 2</i>				
	Age		-0.48	-3.85	<0.001
Percent-correct in symmetric/ non-aligned masker	<i>Step 1</i>				
	Cog	0.29	0.54	4.43	<0.001
	<i>Step 2</i>				
	Cog		0.43	3.74	<0.001
Percent-correct in side/aligned masker	HFPTA	0.41	-0.37	-3.18	0.003
	<i>Step 1</i>				
Percent-correct in side/non-aligned masker	Cog	0.31	.55	4.65	<0.001
	<i>Step 1</i>				
	Cog	0.21	0.46	3.65	0.001

This suggests that previous studies using highly aligned stimuli (e.g., Humes *et al.*, 2006; Marrone *et al.*, 2008; Humes and Coughlin, 2009; Neher *et al.*, 2009; Rossi-Katz and Arehart, 2009; Helfer *et al.*, 2010; Lee and Humes, 2012; Singh *et al.*, 2013; Helfer and Freyman, 2014; Xia *et al.*, 2015) may have overestimated the difficulty experienced by older adults in competing speech situations. Regression analyses suggested that age and cognitive skills take on more importance than hearing loss in accounting for variance in speech perception in the presence of symmetric/aligned maskers, while amount of high-frequency hearing loss explained more variance when symmetric maskers are not aligned with the target. It could be that age-related cognitive skills were more important in perceptually segregating the aligned maskers because of greater potential confusion between target and masking speech in this condition. This suggests that there was more informational masking when the target and maskers were more highly aligned, consistent with others' work (Iyer *et al.*, 2010; Kidd *et al.*, 2014). Analysis of error patterns also supports the idea that there was more informational masking in aligned (versus non-aligned) trials. Masker errors occurred slightly less often for non-aligned conditions while the opposite was true for phonemic errors.

A reduction in informational masking in non-aligned conditions could have been caused by a number of factors. Our procedure for producing non-aligned maskers led to the first masker word in non-aligned trials often starting mid-word, meaning that it was likely unintelligible during these trials. As can be seen in Fig. 5, the largest differences in performance between aligned and non-aligned maskers in the symmetric masking condition can be seen for words at the beginning of the target sentence, consistent with the idea that there was less distraction or confusability at the beginning of non-aligned (versus aligned) trials. Moreover, there were word order differences between the target and maskers in non-aligned conditions, which likely reduced informational masking (Iyer *et al.*, 2010; Kidd *et al.*, 2014), and the masker syntax was less predictable, so listeners' expectations about syntax would not be as helpful in non-aligned conditions (Kidd *et al.*, 2014). The fact that differences between aligned and non-aligned trials diminished as the trial unfolded could be due to alignment "drifting" during aligned trials (as depicted in Fig. 2); even though each sentence had the same number of syllables, there were variations in the placement of one- versus two-syllable words within sentences, and there were small variations in sentence length and speaking rates. This led to alignment being maximal at the beginning of the aligned trials.

Examination of Fig. 5 shows that differences between the older participants and the other two groups were fairly consistent across word position for aligned maskers, but were considerably smaller for words at the beginning of the sentence when the masker and target were non-aligned. Hence, older adults were especially susceptible to whatever led to the increased difficulty with words beginning the sentence in the aligned condition. It could be that, for younger listeners, spatial cues were sufficient to segregate the speech streams, so alignment was less detrimental in terms of

distinguishing target from maskers. Perhaps spatial cues were less robust for older adults (e.g., Marrone *et al.*, 2008; Besser *et al.*, 2014; Schurman *et al.*, 2014), leading to increased disruption from alignment. It also is likely that energetic masking was greatest at the beginning of the sentence during aligned trials because of the drift in word alignment that occurred as the trial progressed. This might be expected to handicap our middle-aged and older listeners, who, as a group, had poorer hearing (and, therefore, were more susceptible to energetic masking, e.g., Best *et al.*, 2011; Best *et al.*, 2013). It should be noted that the current finding of maximum group differences in aligned trials for words at the beginning of the sentence is, to some extent, at odds with results of two studies that varied the onset between target and masker. These experiments found that younger listeners used this cue more effectively than did older participants, whether the target began before the masker (Lee and Humes, 2012) or after the masker (Ben-David *et al.*, 2012). In the present study, the masker often began before the target in non-aligned conditions, and differences between groups for words at the beginning of the sentence were smaller here than in aligned conditions. However, comparison of the present results with these two previous studies is not straightforward, since target and masking speech in those studies only differed in terms of onset, and stimuli were presented without spatial separation.

Consistent with previous research (e.g., Marrone *et al.*, 2008; Dawes *et al.*, 2013; Glyde *et al.*, 2013b; Besser *et al.*, 2014), older participants in the present study appeared to be at a particular disadvantage in the presence of symmetrically placed maskers. However, performance was higher overall with maskers on one side so any conclusions about the importance of symmetry versus asymmetry in the current study must be drawn with caution. True determination of group differences between single-side and symmetric maskers would need to be assessed at SNRs that lead to equivalent overall levels of performance in these two spatial configurations. It is interesting to note, however, that even though performance was quite good in the asymmetric condition, cognitive skills measured in this study accounted for a significant amount of variance in the speech perception scores obtained with this configuration.

Why might older adults have particular difficulty when maskers surround the target? Widened auditory filters, reduced temporal resolution, and lack of audibility all could contribute to decreased ability to use glimpses in a symmetric masking condition (e.g., Arbogast *et al.*, 2005; Hopkins and Moore, 2011; Glyde *et al.*, 2013a). Some research suggests that better-ear glimpsing cannot explain reduced SRM in young people with hearing loss (Glyde *et al.*, 2013b; Best *et al.*, 2015) and that cognitive mediation is particularly important in symmetric masking conditions (Neher *et al.*, 2012; Gallun *et al.*, 2013). Our regression results (which found that cognitive skills accounted for variance in both symmetric and asymmetric masking conditions) seem to be at odds with the idea that glimpsing in the symmetric maskers requires additional cognitive mediation, above and beyond that required for understanding speech in competing speech *per se*. Taken as a whole, existing data support the

idea that additional factors besides just age-related hearing loss limit the performance of individuals in competing speech situations.

Several of our previous studies that tested middle-aged adults found differences in speech recognition between these listeners and younger adults in the presence of competing speech (but not noise) maskers (Helfer and Vargo, 2009; Helfer and Jesse, 2015). In those studies, the target and competing sentences were aligned. The present results support our previous work: small but consistent differences between middle-aged and younger adults were found for aligned maskers in the symmetric masking condition. The underlying causes of this early aging effect have not yet been identified. Even minimal hearing loss can affect the ability to use spatial cues (e.g., Glyde *et al.*, 2013b). Moreover, temporal processing, which may be important for using the rapidly changing glimpses occurring in symmetric maskers, appears to be particularly vulnerable to early aging (e.g., Grose *et al.*, 2006; Helfer and Vargo, 2009; Grose and Mamo, 2010; Fullgrabe *et al.*, 2015). Notable in the present results, however, is the lack of differences between middle-aged and younger participants in the other conditions (with maskers that are on one side and with non-aligned maskers in the symmetric masking condition). Future studies will need to be conducted to better define acoustic differences in realistic complex sound fields that lead to early age-related changes in speech understanding.

ACKNOWLEDGMENTS

We thank Sarah Laakso, Kim Adamson-Bashaw, and Michael Rogers for their assistance with this project. This work was supported by The National Institute on Deafness and Other Communication Disorders (NIDCD) R01 012057 and NIDCD R01 01625.

Arbogast, T. L., Mason, C. R., and Kidd, G. (2005). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **117**, 2169–2180.

Ben-David, B. M., Yse, V. Y. Y., and Schneider, B. A. (2012). "Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker?," *Hear. Res.* **290**, 55–63.

Besser, J., Festen, J. M., Goverts, S. T., Kramer, S. E., and Pichora-Fuller, M. K. (2014). "Speech-in-speech listening on the LiSN-S Test by older adults with good audiograms depends on cognition and hearing acuity at high frequencies," *Ear. Hear.* **36**, 24–41.

Best, V., Mason, C. R., and Kidd, G., Jr. (2011). "Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the temporal overlap of competing talkers," *J. Acoust. Soc. Am.* **129**, 1616–1625.

Best, V., Mason, C. R., Kidd, G., Jr., Iyer, N., and Brungart, D. S. (2015). "Better-ear glimpsing in hearing-impaired listeners," *J. Acoust. Soc. Am.* **137**, EL213–EL219.

Best, V., Thompson, E. R., Mason, C. R., and Kidd, G., Jr. (2013). "Spatial release from masking as a function of the spectral overlap of competing talkers," *J. Acoust. Soc. Am.* **133**, 3677–3680.

Block, C. K., and Baldwin, C. L. (2010). "Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials," *Behav. Res. Meth.* **42**, 665–670.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). "A speech corpus for multitalker communications research," *J. Acoust. Soc. Am.* **107**, 1065–1066.

Bronkhorst, A. W. (2015). "The cocktail-party problem revisited: Early processing and selection of multi-talker speech," *Atten. Percept. Psychophys.* **77**, 1465–1487.

Bronkhorst, A. W., and Plomp, R. (1992). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," *J. Acoust. Soc. Am.* **92**, 3132–3139.

Brungart, D. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* **109**, 1101–1109.

Brungart, D. S., and Iyer, N. (2012). "Better-ear glimpsing efficiency with symmetrically-placed interfering talkers," *J. Acoust. Soc. Am.* **132**, 2545–2556.

Darwin, C. J. (2008). "Listening to speech in the presence of other sounds," *Philos. Trans. R. Soc. London. B Biol. Sci.* **363**, 1011–1021.

Dawes, P., Munro, K. J., Kalluri, S., and Edwards, B. (2013). "Unilateral and bilateral hearing aids, spatial release from masking and auditory acclimatization," *J. Acoust. Soc. Am.* **134**, 596–606.

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.

Duquesnoy, A. J. (1983). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," *J. Acoust. Soc. Am.* **74**, 739–743.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," *J. Psych. Res.* **12**, 189–198.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2001). "Spatial release from informational masking in speech recognition," *J. Acoust. Soc. Am.* **109**, 2112–2122.

Fullgrabe, C., Moore, B. C. J., and 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.

Gallun, F. J., Diedesch, A. C., Kampel, S. D., and Jakien, K. M. (2013). "Independent impacts of age and hearing loss on spatial release in a complex auditory environment," *Front. Neurosci.* **7**, 252.

Glyde, H., Buchholz, J., Dillon, H., Best, V., Hickson, L., and Cameron, S. (2013a). "The effect of better-ear glimpsing on spatial release from masking," *J. Acoust. Soc. Am.* **134**, 2937–2945.

Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (2013b). "The effects of hearing impairment and aging on spatial processing," *Ear. Hear.* **34**, 15–28.

Grose, J. H., Hall, J. W., and Buss, E. (2006). "Temporal processing deficits in the pre-senescent auditory system," *J. Acoust. Soc. Am.* **119**, 2305–2315.

Grose, J. H., and Mamo, S. K. (2010). "Processing of temporal fine structure as a function of age," *Ear. Hear.* **31**, 755–760.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (1999). "Speech intelligibility and localization in complex environments," *J. Acoust. Soc. Am.* **105**, 3436–3448.

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.

Helfer, K. S., and Freyman, R. L. (2014). "Stimulus and listener factors affecting age-related changes in competing speech perception," *J. Acoust. Soc. Am.* **136**, 748–759.

Helfer, K. S., and Jesse, A. (2015). "Lexical effects on competing speech perception," *J. Acoust. Soc. Am.* **138**, 363–376.

Helfer, K. S., and Vargo, M. (2009). "Speech recognition and temporal processing in middle-aged women," *J. Am. Acad. Aud.* **20**, 264–271.

Hopkins, K., and Moore, B. C. J. (2011). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," *J. Acoust. Soc. Am.* **130**, 334–349.

Humes, L. E., and Coughlin, M. P. (2009). "Aided speech-identification performance in single-talker competition by older adults with impaired hearing," *Scand. J. Psychol.* **50**, 485–494.

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.

Iyer, N., Brungart, D. S., and Simpson, B. D. (2010). "Effects of target-masker contextual similarity on the multimasker penalty in a three-talker diotic listening task," *J. Acoust. Soc. Am.* **128**, 2998–3010.

Jesse, A., and Janse, E. (2012). "Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners," *Lang. Cognit. Process.* **27**, 1167–1191.

- Kidd, G., Jr., Mason, C. R., and Best, V. (2014). "The role of syntax in maintaining the integrity of speech streams," *J. Acoust. Soc. Am.* **135**, 766–777.
- Koelweijn, T., Zekveld, A. A., Festen, J. M., and Kramer, K. (2012). "Pupil dilation uncovers extra listening effort in the presence of single-talker masker," *Ear Hear.* **33**, 291–300.
- Lee, J. H., and Humes, L. E. (2012). "Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background," *J. Acoust. Soc. Am.* **132**, 1700–1717.
- Marrone, N., Mason, C. R., and Kidd, G., Jr. (2008). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," *J. Acoust. Soc. Am.* **124**, 3064–3075.
- Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A. S., and van Schaik, A. (2009). "Benefit from spatial separation of multiple talkers in bilateral hearing aid users: Effects of hearing loss, age, and cognition," *Int. J. Aud.* **48**, 758–774.
- Neher, T., Lunner, T., Hopkins, K., and Moore, B. C. J. (2012). "Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners," *J. Acoust. Soc. Am.* **131**, 2561–2664.
- Peissig, J., and Kollmeier, B. (1997). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *J. Acoust. Soc. Am.* **101**, 1660–1670.
- Robertson, I. H., Ward, T., Ridgeway, V., and Nimmo-Smith, I. (1996). "The structure of normal human attention: The Test of Everyday Attention," *J. Int. Neuropsychol. Soc.* **6**, 525–534.
- Rossi-Katz, J., and Arehart, K. H. (2009). "Message and talker identification in older adults: Effects of task, distinctiveness of the talker's voices, and meaningfulness of the competing message," *J. Speech. Lang. Hear. Res.* **52**, 435–453.
- Salthouse, T. A., Toth, T., Daniels, K., Parks, C., Pak, R., Wolbrette, M., and Hocking, K. J. (2000). "Effects of aging on efficiency of task switching in a variant of the Trail Making Test," *Neuropsychol.* **14**, 102–111.
- Schurman, J., Brungard, D., and Gordon-Salant, S. (2014). "Effects of masker type, sentence context, and listener age on speech recognition performance in 1-back listening tasks," *J. Acoust. Soc. Am.* **136**, 3337–3349.
- Singh, G., Pichora-Fuller, M. 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.
- Singh, G., Pichora-Fuller, M. K., and Schneider, B. A. (2013). "Time course and cost of misdirecting auditory spatial attention in younger and older adults," *Ear Hear.* **34**, 711–721.
- Sorqvist, P., Ljungberg, J. K., and Ljung, R. (2010). "A sub-process view of working memory capacity: Evidence from effects of speech on prose memory," *Memory* **18**, 310–326.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech. Hear. Res.* **28**, 455–462.
- Taylor, W. L. (1953). "Cloze procedure: A new tool for measuring readability," *Journ. Q.* **30**, 415–433.
- Thorndike, K. I., and Lorge, I. (1952). *The Teacher's Word Book of 30,000 Words* (Columbia University Press, New York).
- Tun, P. A., O'Kane, G., and Wingfield, A. (2002). "Distraction by competing speech in young and older listeners," *Psychol. Aging* **17**, 453–457.
- Woods, W. S., Kalluri, S., Pentony, S., and Nooraei, N. (2013). "Predicting the effect of hearing loss and audibility on amplified speech perception in a multi-talker listening scenario," *J. Acoust. Soc. Am.* **133**, 4268–4278.
- Xia, J., Nooraei, N., Kalluri, S., and Edwards, B. (2015). "Spatial release of cognitive load measured in a dual-task paradigm in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **137**, 1888–1898.