J Psycholinguist Res (2014) 43:59–80
DOI 10.1007/s10936-013-9240-z

Tracking Reading: Dual Task Costs of Oral Reading for Young Versus Older Adults

Susan Kemper · Daniel Bontempo · RaLynn Schmalzried · Whitney McKedy · Bruno Tagliaferri · Doug Kieweg

Published online: 6 March 2013

Abstract A digital pursuit rotor was used to monitor oral reading costs by time-locking tracking performance to the auditory wave form produced as young and older adults were reading out short paragraphs. Multilevel modeling was used to determine how paragraph-level predictors of length, grammatical complexity, and readability and person-level predictors such as speaker age or working memory capacity predicted reading and tracking performance. In addition, sentence-by-sentence variation in tracking performance was examined during the production of individual sentences and during the pauses before upcoming sentences. The results suggest that dual tasking has a greater impact on older adults' reading comprehension and tracking performance. At the level of individual sentences, young and older adults adopt different strategies to deal with grammatically complex and propositionally dense sentences.

Keywords Aging · Linguistic processing · Dual task demands · Reading

Introduction

Older adults need to communicate with their families, friends, and neighbors, with their lawyers and physicians, through face-to-face interaction and over telephones, the internet and other devices. Successful communication involves reading and listening comprehension as well as oral and written production. Common challenges to communication include the declining sensory, cognitive, and physical abilities of older adults (Schneider and Pichora

S. Kemper (✉) · R. Schmalzried
Department of Psychology, Gerontology Center, University of Kansas, 3090 DHDC,
1000 Sunnyside Ave., Lawrence, KS 66045, USA
e-mail: SKemper@KU.edu

D. Bontempo · B. Tagliaferri · D. Kieweg
Center for Biobehavioral Neurosciences in Communication Disorders, University
of Kansas, Lawrence, KS 66045, USA

W. McKedy
Gerontology Center, University of Kansas, Lawrence, KS 66045, USA

Fuller 2000) as well as intergenerational differences in communicative goals and discourse styles (Hummert 2009). Although a variety of age-related impairments to communication have been documented (Thornton and Light 2006), understanding how aging affects linguistic processing, hence, how best to enhance older adults' communicative competence, has lagged.

In order to delineate how aging affects linguistic processing, Kemper et al. (2011) compared the performance of young versus older adults on a secondary task while they were speaking. The participants engaged in a well-practiced perceptual-motor task, pursuit rotor tracking, while responding orally to probe questions about their likes and dislikes. Their tracking performance was time-locked to their speech so that utterance-by-utterance variation in tracking performance could be assessed. When young and older adults were simultaneously talking and tracking a moving target, their tracking performance declined during the pauses before utterances containing many words or propositions or utterances that were propositionally dense, suggesting that planning long or propositionally dense utterances is costly, but equally so, for both young and older adults. Tracking performance also declined during the production of utterances containing many words or propositions, those that were propositionally dense, and those that were grammatically complex, and these production costs were greater for older adults. Tracking performance also declined during the pauses after utterances containing many words or propositions and those produced rapidly, suggesting that speakers must recover during the next pause after producing a difficult utterance. Further, these output costs were greater for older adults than for young adults. Thus, while speech planning appears to be no more costly for older adults than for young, producing and recovering from long, informative, and complex utterances is more costly for older adults than for young adults.

Although the costs of speech production and recovery appear to be greater for older adults than for young adults, both groups used a similar speech style when simultaneously engaged in pursuit rotor tracking. Both young and older adults tended to use slow, short, simple sentences although young adults in general use a faster, more complex speech style than older adults (Kemper 2006). Hence, one limitation of the Kemper et al. (2011) study is that neither young nor older adults spontaneously produced many grammatically complex or propositionally dense sentences when engaged simultaneously in pursuit rotor tracking. To further probe for age group differences in linguistic processing, the present experiment examined oral reading of paragraphs and sentences while the participants were also engaged in pursuit rotor tracking.

Ferreira (1991), following Selkirk (1986) and Levelt (1989), suggested that linguistic analysis is required for the oral repetition of sentences since the phonological form of a sentence is specified by its linguistic structure. Ferreira (1991) states "For any sentence to be spoken, it must be translated into a representation that can control the speech apparatus….The more syntactic nodes that must be translated, the longer the translation takes and so, the longer the initiation or pause time…". (p. 227). Following this assumption, reading sentences aloud should tap the same linguistic processes and cognitive resources used during spontaneous speech. In the present experiment, young and older adults were asked to read aloud paragraphs and sentences varying in grammatical complexity as well as on other linguistic dimensions. If Ferreira's assumption is correct, oral reading should, like spontaneous speaking, affect concurrent pursuit rotor tracking. However, unlike spontaneous speaking, this approach provides a way to experimentally manipulate the grammatical complexity or other psycholinguistic properties of the paragraphs or sentences that are read aloud.

In this experiment, young and older adults were trained on pursuit rotor tracking and then asked to track a moving target while simultaneously reading aloud. The continuous recording of tracking was time-locked to the digital recording of the oral reading. This approach permits two levels of analysis: at the first, or paragraph, level, reading comprehension, reading rate, and tracking performance as the paragraphs were read aloud compared to baseline conditions in which the participants read paragraphs aloud while ignoring a moving target or tracked the moving target while ignoring a paragraph; in addition, at a second, or sentence, level of analysis, individual sentences were examined by segmenting the continuous record tracking at the onset and offset of each sentence. At this level of analysis, we could examine tracking performance as individual sentences were read aloud as well as the duration of the pauses between sentences and tracking performance during these pauses. At both levels of analysis, decrements in tracking performance were assumed to reflect the processing costs of individual paragraphs or sentences.

Not only were we interested in comparing the performance of young versus older adults at both levels of analysis, we were also interested in investigating how their performance varied with properties of the individual paragraphs or sentences. Paragraphs were selected from a variety of sources so that varied in overall length as well as in their overall propositional density, grammatical complexity, and readability, all properties that have been shown to affect reading comprehension and reading rate (cf., Stine-Morrow et al. 2008). A variety of measures of length, complexity, and readability were used as predictors of reading comprehension, reading rate, and tracking performance to compare how they affected young and older adults. Individual sentences also varied in grammatical complexity and propositional density and these measures were used as predictors of tracking performance and pause duration.

Finally, we also investigated how age group and individual differences in cognition affected reading comprehension and tracking performance. Although vocabulary knowledge increases over the lifespan (Verhaeghen et al. 2003), most models of cognitive aging assume that working memory, processing speed, and inhibitory control decline (Park et al. 2002). Vocabulary knowledge (Roland et al. 2007; Stine-Morrow et al. 2008), working memory (Swets et al. 2007), processing speed (Rabaglia and Salthouse 2012; Stine et al. 1986), and inhibitory control (Connelly et al. 1991; Engelhardt et al. 2010; Hasher and Zacks 1988) have been shown to affect language processing. These individual and age group changes to cognition may also affect reading comprehension, reading rate, or tracking performance. A battery of cognitive tests was administered to the young and older adults and composite scores were used as predictors of the costs of reading comprehension, reading rate, and paragraph-level and sentence-level tracking performance to assess these hypotheses.

Method

Participants

Forty young and 40 older adults were tested; technical problems resulted in the loss of oral reading data for 2 older adults and 3 additional older adults withdrew citing problems reading the paragraphs on the computer monitor. The young adults were recruited by signs posted on campus and class announcements while the older adults were recruited from a database of prospective and previous research participants. The participants were paid for their participation. Participants were given a battery of tests of cognitive ability, including measures of working memory, processing speed, inhibition, and vocabulary. Table 1 summarizes the means, standard deviations, and age group comparisons for each observed measure; an alpha level of 0.05 was set for these and all subsequent tests.

Vocabulary was assessed by the Shipley (1940) Vocabulary Test, the North American Reading Test (AmNART; Grober and Sliwinski 1991), and educational attainment in years.

Table 1 Comparison of young and older participants

Characteristic	Young adults		Older adults	
	Mean	SD	Mean	SD
Age	21.0	2.4	75.0	7.6
Vocabulary composite	−0.40	0.56	0.38	1.09
Years of education**	14.7	1.8	15.9	2.5
North American reading test**	30.9	4.4	35.7	7.9
Shipley vocabulary**	31.1	4.9	34.8	4.2
Processing speed composite	0.68	0.47	−0.60	0.81
Stroop X**	84.8	12.1	68.6	14.1
Digit symbol**	31.9	4.9	23.6	6.0
Trail making A**	48.7	14.0	80.5	28.8
Working memory composite	0.27	0.82	−0.20	0.76
Digits forward*	9.6	2.4	8.9	2.4
Digits backward*	7.9	2.1	7.2	2.5
Daneman and carpenter**	3.8	1.0	3.0	0.5
Inhibition composite	0.03	0.17	0.03	0.17
Stroop words**	63.2	10.1	37.5	8.8
Stroop interference %**	−25.1	8.9	−45.9	11.2
Trail making B**	56.2	19.9	109.5	45.3
Trail making interference %*	−18.3	3.4	−38.9	4.1

* $p < .05$; ** $p < .01$.

Working memory was assessed by performance on the Digits Forward and Digits Backwards tests (Wechsler 1958) and the Daneman and Carpenter (1980) Reading Span Test. Processing speed was assessed using the Digit Symbol Test (Wechsler 1958), the baseline condition of the Stroop test (Stroop 1935), the Trails A portion of the Trail Making test (Reitan 1958), and the asymptotic rotor speed attained by the participant following practice (see below).

For working memory, processing speed, and vocabulary, a summary composite was formed. Factor loadings were obtained from Stata (Stata Corp 2009) using maximum likelihood estimation. Subsequently factor scores were generated for each participant. For each composite, the respective factor analysis found a single latent factor with moderately-high to high loadings for each indicator measure. The eigenvalues from the vocabulary, working memory, and processing speed factor models were, respectively, 1.91, 1.32, and 2.69. Loadings obtained for the vocabulary composite were: Shipley ($\lambda = 0.81$), AmNART ($\lambda = 0.90$), and educational attainment ($\lambda = 0.66$). Loadings obtained for the working memory composite were: Digits Forward ($\lambda = 0.62$), Digits Backward ($\lambda = 0.73$), and Reading Span ($\lambda = 0.63$). Loadings obtained for the processing speed composite were: Digit Symbol ($\lambda = 0.86$), Stroop baseline ($\lambda = 0.78$), Trails A total seconds ($\lambda = -0.87$), and asymptotic rotor speed ($\lambda = 0.76$).

Lastly, the Stroop and Trail Making Tests were also used to derive two measures of inhibition. A Stroop interference score was then calculated as Stroop Interference = (blocks of Xs—color names)/blocks of Xs. A Trail Making interference score was calculated as Trail Making Interference = (seconds Trail A—seconds Trail B)/seconds Trail A. Because only 2 measures of inhibition were available, the Stroop and Trail Making interference scores were averaged for each participant to create a summary measure.

Task and Design

Twelve paragraphs were selected from a variety of sources including on-line encyclopedia, high school and college textbooks, and newspaper articles so that their overall length in words and sentences, vocabulary in terms of word frequency, target audience, and writing style differed. Two additional paragraphs were used during training. The paragraphs were selected to cover a range of general-knowledge topics and writing styles. All were 10–20 sentences in length. A variety of measures of length, grammatical complexity, and content were assessed using procedures similar to those used by Kemper et al. (2010, 2011) to analyze oral language samples using Coh-Metrix (Graesser et al. 2004) and CPIDR-3 (Brown et al. 2008). Ten paragraph-level measures of length, grammatical complexity, and readability were obtained; correlations among these measures are reported in Table 2. Single indicators, rather than latent factor scores, were used as predictors of reading and tracking performance since the paragraphs were chosen to reflect a range of topics and writing styles rather to systematically explore their underlying factor structure.

Length: In addition to paragraph length in sentences and words, Mean Length of Utterance (MLU) in words was obtained automatically from the Coh-Metrix program (Graesser et al. 2004), and the number of propositions in the entire paragraph was obtained from the CPIDR-3 computer program (Brown et al. 2008). Sentences ranged from 9 to 20 ($M = 13.6$, SD $= 3.2$); words ranged from 126 to 299 ($M = 242.8$, SD $= 51.2$); MLUs ranged from 12.6 to 22.5 ($M = 18.0$, SD $= 2.5$); propositions ranged from 60 to 157 ($M = 121.3$, SD $= 26.2$). The number of words, sentences, and propositions are strongly correlated for this sample, all $r(11) > 0.80$, $p < .01$; average MLU is weakly correlated with the measures of paragraph length.

Grammatical complexity: Two measures of grammatical complexity were obtained for each paragraph. Developmental Level (DLevel) was scored based on a scale originally developed by Rosenberg and Abbeduto (1987). Grammatical complexity ranged from simple

Table 2 Correlations among the paragraph-level predictors of length, grammatical complexity, and readability

Measure	1	2	3	4	5	6	7	8	9	10
1. Words	–									
2. Sentences	0.84**	–								
3. Propositions	0.96**	0.81**	–							
4. MLU	0.24	−0.32*	0.21	–						
5. DLevel	−0.25	−0.42*	−0.29	0.36*	–					
6. GIndex	−0.50**	−0.36*	−0.15	−0.18	0.40*	–				
7. PDensity	0.17	−0.06	0.21	0.01	−0.15	−0.15	–			
8. CIndex	−0.71**	−0.59**	−0.77**	−0.13	−0.06	0.57**	−0.24	–		
9. TTR	0.58**	0.25	0.47**	0.53**	0.16	−0.29*	−0.28*	−0.62**	–	
10. Flesch	−0.13	0.36*	−0.17	−0.90**	−0.23	0.17	−0.24	0.10	−0.49**	–

MLU mean length of utterance, *DLevel* developmental level, *GIndex* grammatical index, *PDensity* propositional density, *CIndex* coherence index, *TTR* type token ratio, *Flesch* Flesch reading ease score
* $p < .05$; ** $p < .01$.

one-clause sentences (DLevel = 0) to complex sentences with multiple forms of embedding and subordination (DLevel = 7). Each sentence was scored and the average DLevel for each paragraph was then calculated. Second, Coh-Metrix provided the Grammatical Index (GIndex) of each paragraph as a sum of 3 counts per 10 words: the number of connectives such as "because", "and," or "if", the number of noun phrases, and the number of higher level constituents, such as noun phrase complements and relative clauses. Higher DLevel and GIndex scores indicate texts are more grammatically complex. DLevels ranged from 2.4 to 5.4 ($M = 3.4$, SD $= 0.8$); GIndex scores ranged from 52.8 to 114.9 ($M = 78.1$, SD $= 18.0$). DLevel and GIndex for this small sample correlate $r(11) = 0.40$, $p < .05$.

Readability: Propositional Density (PDensity) was calculated by the CPIDR-3 computer program (Brown et al. 2008); each sentence was decomposed into its constituent propositions that represent propositional ideas and the relations between them. PDensity was computed as the average number of propositions per 100 words. Higher PDensity scores indicate texts that are more dense. Second, Coh-Metrix provided a measure of coherence, the Coherence Index (CIndex), as the sum of 2 measures: (1) argument overlap or the proportional of adjacent sentences that share 1 or more nouns, pronouns, or noun phrases, and (2) LSA cohesion. LSA cohesion is based on latent semantic analysis (Landauer et al. 1998) which assesses the conceptual similarity of a text relative to that of other texts; in these analyses, the LSA cohesion score measured how conceptually similar each sentence was to all other sentences in the paragraph. Higher CIndex scores indicate more cohesive texts. Similarly is determined by the overlap of specific words, semantically related words, and words that commonly co-occur (e.g., "President" and "White House"). Coh-Metrix provided a Type-Token Ratio (TTR) to measure lexical diversity; lower TTRs indicate that many words are repeated throughout the paragraph and higher TTRs reflect the use of a greater diversity of words. Finally, Flesch reading ease (Flesch 1948) was determined by Coh-Metrix; the reading ease scores range from 0 to 10 with a higher score indicating greater reading ease. It reflects the average sentence length in words and the average number of syllables per word. (The reading ease score is often converted to a grade level readability score, ranging from 0 to 12, with lower numbers indicating greater readability or grade level suitability). PDensity ranged from 45.5 to 59.1 propositions per 100 words ($M = 50.0$, SD $= 3.70$); CIndex ranged from 0.9 to 1.7 ($M = 1.4$, SD $= 0.3$); TTRs ranged from 0.67 to 0.86 ($M = 0.78$, SD $= 0.29$); Flesch reading ease ranged from 23.7 to 83.9 ($M = 44.6$, SD $= 18.0$). Lexical diversity, assessed by TTR, is correlated negatively with propositional density, semantic coherence, and Flesch reading ease, all $r(11) \leq .28$, $p < .05$.

In addition to these 10 paragraph-level predictors, there were 2 sentence-level predictors available: The PDensity (number of propositions/number of words) of each sentence and the DLevel measure of grammatical complexity of each sentence. Finally, reading rate in words per min (wpm) was calculated for each paragraph in the baseline reading and dual task conditions; paragraph reading time was obtained from the synchronized tracking record which marked the onset and offset of the paragraph reading and converted to reading rate in words per min. Sentence reading rate in words per min was also calculated for each sentence in the baseline reading and dual task conditions.

Two 4-alternative choice questions were also prepared for each paragraph. The questions required inferential answers. To ensure that information obtained from the paragraph was necessary to answer the questions, a panel of 10 naïve judges attempted to the answer the questions without reading the paragraphs; they correctly choose the right answer only 22 % of the time. A second panel of 10 judges answered the questions after reading the paragraphs; they were correct 87 % of the time.

Rotor Training

Participants were initially trained on pursuit rotor tracking following the protocol in Kemper et al. (2010, 2011). Rotor training was conducted using a stand-alone version of the digital pursuit rotor developed by the Digital Electronics and Engineering Core of the Biobehavioral Neurosciences and Communication Disorders Center, a component of the Schiefelbusch Institute for Life Span Studies at the University of Kansas. The pursuit rotor features a target that rotates along a circular track. Participants use a trackball mouse to track the target, attempting to keep a pointer centered on the moving target. Rotor speed can be varied from approximately 0.2 to 2 revolutions per minute; the program samples the location of the pointer approximately every 16 ms, and determines its distance (in pixels) from the center of the target. A moving average of the pointer status (on/off target) is taken over 3 successive 100 ms intervals, and percentage time off target (TOT) is determined. In addition, tracking error (TE) or the distance in pixels from the center of the target to the pointer is used as a second measure of tracking performance; it is also averaged over 3 successive 100 ms intervals.

Participants were trained on the pursuit rotor task to an asymptotic performance level. Initial tracking speeds for young and older adults were set at 1.2 and 0.45 rev per min, respectively. Participants practiced tracking for 30 s and received feedback on their performance. A "2 up/1 down stair-case" training procedure was used to gradually increase tracking speed on successive 30 s trials: if average time off target was 20 % or less for a trial, the speed was increased by 10 % for the next trial; if time off target was greater than 20 %, the speed was decreased by 5 %. The stair-case procedure converged on an asymptotic rotor speed when the rotor speed oscillated around the same value, moving "up" and "down" past this value 3 times. Young adults ($M = 16.5$, SD $= 4.4$) required fewer trials to reach an asymptotic rotor speed than older adults ($M = 18.8$, SD $= 4.9$), $F(1, 78) = 4.86$, $p < .05$. Asymptotic tracking speed was greater for young adults ($M = 1.7$ rev/m, SD $= 0.3$) than for older adults ($M = 1.0$ rev/m, SD $= 0.3$), $F(1, 78) = 97.99$, $p < .01$. However, asymptotic TOT ($M = 18.4$ %, SD $= 3.8$) and TE ($M = 7.6$ pixels, SD $= 0.9$) were comparable for young and older adults, both $p > .50$.

Experimental Procedure

Following rotor training, two experimental tasks were administered; order was counterbalanced across participants. Both were administered using Paradigm (Tagliaferri 2005). In addition to the paragraph reading task, participants were also tested on a controlled sentence production task reported separately (Kemper et al. 2010).

To familiarize the participants with the paragraph reading task, 2 practice paragraphs were presented on the computer screen while they read the paragraphs aloud; following each paragraph, 2 comprehension questions were presented along with 4 alternatives to familiarize the participant with the use of trackball to indicate the correct answer. Following this practice, participants were tested on 3 conditions in a fixed order: First, in the baseline tracking condition, participants engaged in pursuit rotor tracking while ignoring a paragraph presented in the center of the circular rotor track. Second, in the baseline paragraph reading condition, participants read aloud a paragraph presented in the center of the rotor track while ignoring the moving target. Third, in the dual task condition, participants attempted to track the moving target while reading aloud a paragraph. Two trials were administered during each condition; paragraphs were counterbalanced across participants, trials, and the experimental conditions such that each participant ignored 2 randomly selected paragraphs in the baseline tracking condition, read aloud 2 randomly-selected paragraphs in the baseline reading condition, and read aloud 2 randomly-selected paragraphs while engaged in pursuit rotor tracking during the dual task condition. Two comprehension questions were presented immediately after each paragraph in all three (baseline tracking, baseline reading, and dual task) conditions.

A version of the pursuit rotor was embedded within Paradigm and tracking speed was set to the asymptotic speed achieved by the participant during training. The paragraphs were presented centered within the circular rotor track and did not obscure the track, the target, or the pointer. Each trial involved 3 phases:

1. Warm-up: The rotor track and bull's eye target were displayed and the target began to move after a 3 s delay and participants tracked it continuously for 20 s while a central fixation cross was presented. The rotor was reset at the beginning of each trial, repositioning the target to the "6 o'clock" starting position.
2. Paragraph Presentation: After the 20 s warm-up, the paragraph was presented centered within the track. It remained visible for 3 min in the baseline tracking condition or until the participant had finished reading it aloud in the baseline reading and dual task conditions.
3. Questions: Each comprehension question was presented along with 4 alternative choices. The participant used the trackball to point to the correct answer. The rotor track and target were not displayed during this phase.

Participants read the paragraphs aloud; their responses were recorded and the audio (WAV) files were synchronized with their tracking record in the dual task condition. A utility program, the Rotor On-line Speech Segmenter (ROSS), permits these time-locked records to be segmented into sentences and pauses. The audio file is replayed while a listener inserts cursors to mark the onset and offset of sentences; play-back speed can be adjusted, the location of the cursors can be manually fine-tuned. The ROSS utility produces a segmented wave form of sequentially ordered sentences and intra-sentential pauses. The ROSS utility then extracts measures of tracking performance corresponding to each sentence or pause. These include: TE, TOT, and variability in TE and TOT during the sentence or pause. The resulting segmented performance record was exported as a spreadsheet which was then annotated by inserting the sentence-level measures (number of words and propositions and sentence DLevel).

Using the ROSS utility, two trained coders analyzed 10 % of the paragraph audio files to assess reliability; the remaining samples were analyzed by a single coder. After practice, the two coders were able to accurately tag the onset and offset of sentences: the resulting segment durations were highly correlated, $r > 0.99$, and average disagreement as to the onset or offset of sentences was less than ± 20 ms.

Results

Three sets of analyses are reported. Paragraph-level analyses of reading performance in the baseline and dual task conditions are presented followed by the analysis of tracking performance during the baseline and dual task conditions. Following these analyses, sentence-level analyses of reading rate and tracking performance during the dual task condition are reported. The final analysis examined the duration of pauses between sentences during the dual task condition.

Paragraph-Level Analyzes

In these analyses, mixed effects regression using restricted maximum likelihood estimation was used to estimate effects of age group, condition, and their interaction on reading and tracking performance (Hoffman and Rovine 2007). The 4 person-level measures of working memory, processing speed, vocabulary, and inhibition were also examined as potential predictors of reading or tracking performance. Composites scores derived from the factor analysis were used as the predictors of working memory, processing speed, and vocabulary; the average of the two interference scores was used as the inhibition predictor. In addition, the effects of the 10 paragraph-level measures of length, grammatical complexity, and content were also examined as potential predictors of reading or tracking performance. All person-level measures of cognitive ability were group mean-centered prior to the analysis and the paragraph-level measures of length, grammatical complexity, and content were also mean-centered prior to the analysis. Unless noted below, young adults and either the reading or tracking baseline condition were used as the model reference; hence, estimates indicate the improvement or decline in performance for older adults or for the dual task condition. The mixed models included only a random intercept for subjects since the selection of 12 paragraphs was highly constrained. All analyses were performed using Stata's XTMixed procedure; significant parameter estimates (and SEs) are reported. Significance was tested by obtaining a z-score from ratio of the parameter estimate to its standard error, with an associated 2-tailed p value at $\alpha = 0.05$. Unless reported below, all other effects and interactions were not significant.

Reading Performance

Reading performance was assessed by 2 measures: comprehension accuracy assessed as percent correct on the 2 probe questions per paragraph and words per minute (wpm) reading rate. In the mixed effect models reported below, positive estimates indicate an increase in comprehension accuracy or reading rate, hence improvements in reading performance, whereas negative estimates indicate a decrease in reading performance.

Comprehension accuracy: Comprehension was assessed by the percentage of questions answered correctly (chance = 25 %) during the baseline tracking condition when the participants were instructed to ignore the paragraphs, baseline reading condition when participants were instructed to ignore the moving target, and the dual task condition. Mixed effects regression was used to estimate effects of age group, condition, and their interaction as well as the effects of the 4 person-level predictors of processing speed, inhibition, working memory and education and 10 paragraph-level measures of length, grammatical complexity, and content. Both groups were able to perform slightly above chance ($M_O = 29\,\%$, SD $= 26\,\%$; $M_Y = 37\,\%$, SD $= 22\,\%$) in the baseline tracking condition when they were instructed to ignore the paragraphs while tracking the target. Their comprehension accuracy improved during the baseline reading condition when they were instructed to read the paragraphs and ignore the moving target ($M_O = 56\,\%$, SD $= 23\,\%$; $M_Y = 55\,\%$, SD $= 22\,\%$), resulting in a significant estimate (est.) for the baseline reading condition (est. $= 26.9\,\%$, SE $= 5.2\,\%$, $p < .01$). Comprehension accuracy was also higher in the dual task condition ($M_O = 49\,\%$, SD $= 28\,\%$; $M_Y = 59\,\%$, SD $= 24\,\%$) than in the baseline tracking condition, resulting in a significant estimate for the dual task condition (est. $= 29.4\,\%$, SE $= 5.6\,\%$, $p < .01$). The comprehension of older adults was similar to that of young adults overall, as indicated by the nonsignificant estimate for age group (est. $= 7.8\,\%$, SE $= 5.6\,\%$, $p = .162$), and older adults' improvement in the baseline

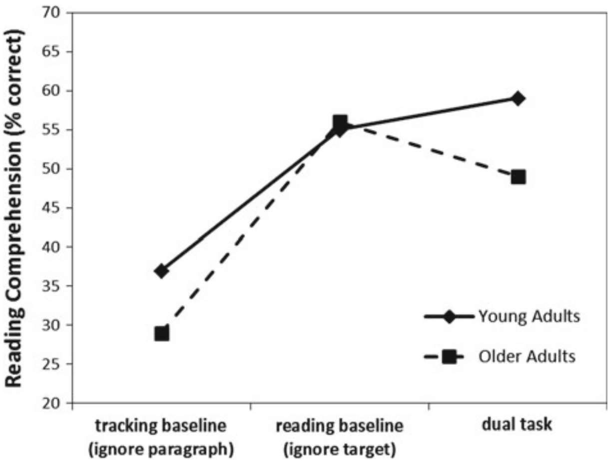

Fig. 1 Improvement in Comprehension Accuracy for Young and Older Adults during the Reading Baseline and Dual Task Conditions based on Mixed Model Estimates (with 95 % Confidence Intervals) for Age Group, Condition, and their Interaction

reading condition was comparable to young adults', as indicated by the nonsignificant estimate for the age by baseline reading interaction (est. = −9.0 %, SE = 7.68, p = .240). However, compared to the young adults, the older adults were less successful in the dual task condition, resulting in a significant age group by dual task condition estimate (est. = −17.2 %, SE = 7.7, p = .025). See Fig. 1.

Comprehension accuracy in all 3 conditions was similar for all participants regardless of individual differences in inhibition, working memory, processing speed, and vocabulary. Comprehension accuracy[1] did vary with the number of words (est. = 0.27, SE = 0.06), sentences (est. = 1.01, SE = 0.26), propositions (est. = 0.39, SE = 0.08), MLU (est. = 0.62, SE = 0.18), DLevel (est. = −0.36, SE = 0.16), GIndex (est. = −0.04, SE = 0.01), PDensity (est. = −0.74, SE = .17), Flesch reading ease (est. = 0.33, SE = .13), TTR (est. = 0.77, SE = .37), and CIndex (est. = −0.95, SE = 0.74), all $p < .01$. Adding information (words, sentences, or propositions) improved comprehension as did reducing DLevel and propositional density, increasing Flesch reading ease, and increasing lexical diversity. Interestingly, higher GIndex scores and CIndex scores were associated with worse comprehension; it may be that for these short paragraphs, GIndex and CIndex do not provide reliable measures of grammatical complexity and semantic cohesion, respectively. The effects of paragraph length, grammatical complexity, propositional density, and semantic cohesion were similar in the baseline reading condition and the dual task condition.

Reading Rate

Reading rate in words per min (wpm) was assessed during the baseline reading task when participants were instructed to ignore the moving target and during the dual task

[1] Estimates indicate the increase or decrease in the number of correctly answered questions per each 1 unit increase in the predictor. For example, the estimate for propositions of 0.39 indicates that adding an additional proposition will increase comprehension by approximately 0.4 questions and the estimate for PDensity of −0.74 indicates that increasing PDensity by 1 proposition per 100 words will reduce the number of correctly answered questions by 0.74.

condition. Mixed effects regression was used to estimate effects of age group, condition, and their interaction as well as the effects of the 4 person-level predictors of processing speed, inhibition, working memory and education and 10 paragraph-level measures of length, grammatical complexity, and content. Compared to the baseline reading condition in which participants ignored the moving target while reading ($M_O = 131$ wpm, SD $=$ 31 wpm; $M_Y = 158$ wpm, SD $=$ 21 wpm), reading rates in the dual task condition ($M_O = 107$wpm, SD $=$ 28wpm; $M_Y = 150$ wpm, SD $=$ 27 wpm) were slower (est. $= -7.3$ wpm, SE $=$ 3.1 wpm, $p < .02$); Compared to young adults, older adults read more slowly (est. $= -26.4$ wpm, SE $=$ 5.3 wpm, $p < .01$), especially in the dual task condition (est. $= -17.2$ wpm, SE $=$ 4.5 wpm, $p < .001$).

Reading rates in both conditions were similar for all participants regardless of individual differences in inhibition, working memory, and vocabulary although those individuals who were faster on the processing speed measures, relative to their age group mean, also read more rapidly (est. = 19.9 wpm, SE = 2.8 wpm, $p < .01$). Reading rate was affected by length (sentences: est. = 13.7, SE = 6.3; words: est. = −5.1, SE = 1.4; propositions: est. = 7.7, SE = 2.1), by grammatical complexity (DLevel: est. = −9.3, SE = 4.1; GIndex: est. = −0.7, SE = 0.3), and by readability (PDensity: est. = −17.2, SE = 4.3; Flesch: est. = 10.0, SE = 3.2), all $p < .01$.[2] Reading rate did not vary with MLU, CIndex, or TTR. Reading rates were faster for paragraphs with more sentences and propositions and for those higher in Flesch reading ease (e.g., lower in grade level); reading rates were slower for paragraphs with more words, more complex sentences, and greater propositional density. The effects of paragraph length, grammatical complexity, propositional density, and semantic cohesion on reading rate were similar in both the reading baseline condition and the dual task condition.

Reading Summary

Although they read more slowly than young adults, older adults had equally good comprehension of the paragraphs in the baseline reading condition when they were instructed to ignore the moving target. However, the older adults were unable to match the young adults' comprehension of the paragraphs in the dual task condition, despite a further decline in their reading rate. Comprehension and reading rate varied with the length, grammatical complexity, and readability of the paragraphs in both conditions.

Tracking Performance

Tracking performance in the baseline tracking condition in which the participants were instructed to ignore the paragraphs while tracking the moving target was compared to performance during the dual task condition. Tracking performance was assessed by 4 measures, TE or tracking error in pixels and TOT or time on target (percent), as well as the variability of each measure. Separate models were estimated for the two phases: the initial warm-up phase prior to the onset of the paragraph, and the paragraph presentation phase, examining the effects of age group, condition, and their interaction as well as the 4 person-level predictors of processing speed, inhibition, working memory, and education. The effects of the 10 paragraph-level predictors of length, grammatical complexity, and content were examined

[2] Estimates indicate the increase or decrease in reading rate per each 1 unit increase in the predictor. For example, the estimate for length in sentences of 13.7 indicates that adding an additional sentence will increase reading rate by approximate 14 wpm and the estimate for PDensity of −17.2 indicates that increasing PDensity by 1 proposition per 100 words will reduce reading rate approximately 17 wpm.

Table 3 Tracking performance by young and older adults during the warm-up and paragraph presentation phases in the baseline tracking condition and the dual task condition

Measure	Young adults				Older adults			
	Warm-up phase		Paragraph presentation		Warm-up phase		Paragraph presentation	
	Baseline	Dual task	Baseline	Dual task	Baseline	Dual task	Baseline	Dual task
TE	15.1 (10.8)	11.2 (4.8)	8.8 (1.2)	16.0 (2.7)	23.5 (19.8)	19.6 (15.4)	10.9 (2.6)	27.6 (13.5)
TE SD	17.5 (10.7)	13.1 (9.1)	5.5 (1.4)	10.9 (3.1)	17.1 (11.6)	17.1 (10.8)	8.6 (4.7)	19.4 (7.8)
TOT	87.0 (.5)	88.6 (6.9)	90.4 (4.6)	59.5 (10.8)	80.0 (12.9)	80.0 (12.7)	83.3 (7.4)	43.0 (16.2)
TOT SD	25.6 (6.0)	23.7 (6.3)	21.5 (5.7)	30.0 (0.3)	26.5 (6.0)	26.5 (6.4)	27.9 (3.0)	29.7 (1.5)

Means and SD (in parenthesis) for Tracking Error (TE) and Time on Task (TOT) are reported

in the models of the paragraph presentation phase only. Tracking performance by young and older adults during the warm-up and paragraph presentation phases is summarized in Table 3.

Warm-up Phase: Tracking performance during the warm-up phase was somewhat less accurate and more variable in the baseline tracking condition when the participants ignored the paragraphs than in the dual task condition as indicated by significant estimates for condition for TE (est. = −3.9, SE = 1.9) TE SD (est. = −4.4, SE = 1.5), and TOT SD (est. = −1.9, SE = 0.8), all $p < .05$. But this effect of condition was similar for both young and older adults with the exception TE SD (condition by age group est. = 4.34, SE = 2.2, $p < .05$). Overall, older adults during the warm-up phase were somewhat worse at tracking than young adults, as indicated by significant age group estimates for TE (est. = 8.4, SE = 2.4) and TOT (est. = −7.0, SE = 1.9), both $p < .01$.

Individuals who were faster, relative to their age group mean, on the processing speed measures had an overall advantage for tracking during either condition as indicated by significant estimates for TE (est. = −3.5, SE = 0.6), TE SD (est. = 3.0, SE = 0.4), TOT (est. = 6.2, SE = 1.1) and TOT SD (est. = −1.3, SE = 0.3), all $p < .01$. Those with better working memory were also somewhat better at tracking (TE est. = −2.70, SE = 1.2; TE SD est. = −2.59, SE = 0.95, TOT SD est. = −1.48, SE = 0.63), all $p < .05$. Individual differences in inhibition and vocabulary did not affect tracking performance in either condition. Tracking performance was similar for all paragraphs regardless of their differences in length, grammatical complexity, and readability with 3 exceptions: TOT variability was slightly reduced when the participants were reading longer paragraphs in words (est. = −0.6, SE = 0.3) and TOT variability slightly increased when the paragraphs contained more propositions (est. = 0.9, SE = 0.4) and for paragraphs higher in Flesch reading ease (est. = 1.3, SE = 0.6), all $p < .01$; these effects on TOT variability were similar for both the baseline tracking condition and the dual task condition.

Paragraph Presentation Phase: Table 4 summarizes the results for the mixed effects analysis of tracking performance examining effects for age group, condition, and their interaction. Tracking performance during the dual task condition was compared to tracking performance during the tracking baseline condition when the participants were instructed to ignore the paragraphs. Tracking was less accurate and more variable during the dual task condition than during the baseline condition as indicated by the significant estimates for condition for TE, TE SD, TOT, and TOT SD. Older adults' tracking was worse than young adults' tracking, as indicated by the significant estimates for age group for TE SD, TOT, and TOT SD. The impact of paragraph reading on tracking performance was greater for older adults than for young adults, as indicated by significant estimates for the condition by age group interactions.

Table 4 Results of the mixed regression models of tracking performance during the paragraph presentation phase

Measure	Condition			Age group			Condition by age group		
	Est.	SE	p	Est.	SE	p	Est.	SE	p
TE	7.3	0.9	<.01	2.1	1.3	=.09	9.4	1.3	<.01
TE SD	5.4	0.6	<.01	3.1	0.9	<.01	5.5	0.9	<.01
TOT	−30.9	1.2	<.01	−7.1	2.1	<.01	−9.4	1.8	<.01
TOT SD	8.5	0.5	<.01	6.4	0.6	<.01	−6.4	0.7	<.01

Parameter estimates (Est.) and Standard errors (SE) are reported for the effects of condition, age group, and their interaction

Individuals who were faster, relative to their age group mean, on the processing speed measures, had an overall advantage for tracking during either condition as indicated by significant estimates for TE (est. = −3.5, SE = 0.6), TE SD (est. = 3.0, SE = 0.4), TOT (est. = 6.2, SE = 1.1) and TOT SD (est. = −1.3, SE = 0.3), all $p < .01$. Individual differences in inhibition, working memory, and vocabulary did not affect tracking performance in either condition. Tracking performance was similar for all paragraphs regardless of their differences in length, grammatical complexity, and readability with 3 exceptions: TOT variability was slightly reduced when the participants were reading longer paragraphs in words (est. = −0.6, SE = 0.3) and TOT variability slightly increased when the paragraphs contained more propositions (est. = 0.9, SE = 0.4) and for paragraphs higher in Flesch reading ease (est. = 1.3, SE = 0.6), all $p < .01$; these effects on TOT variability were similar for both the baseline tracking condition and the dual task condition and for young versus older adults.

Tracking Summary

Older adults could not match the tracking performance of young adults in either the warm-up or paragraph presentation phases and the impact of reading on tracking was greater for older adults than for young adults. However, tracking performance did not, in general, vary with the length, grammatical complexity, or content of the paragraphs.

Sentence-Level Analyses

In these analyses, crossed random effects regression using restricted maximum likelihood estimation was used to estimate fixed effects of age group on reading and tracking performance for individual sentences nested within paragraphs during the dual task condition. The 4 person-level measures of working memory, processing speed, vocabulary, and inhibition were then examined as potential predictors of individual reading or tracking performance. Sentence-level measures of propositional density and grammatical complexity were also examined as potential predictors of reading or tracking performance. All person-level measures of cognitive ability were group mean-centered prior to the analysis and the sentence-level measures were also mean-centered prior to the analysis. Young adults were used as the model reference; hence, estimates indicate the improvement or decline in performance for older adults. All analyses were performed using Stata's XTMixed procedure; significant parameter estimates (and SEs) are reported. Significance was tested by obtaining a z-score from ratio of the parameter estimate to its standard error, with an associated 2-tailed p value at $\alpha = 0.05$. Unless reported below, all other effects and interactions were not significant.

Reading Rate

Reading rate at the level of individual sentences during the dual task condition was analyzed with mixed effects regression to examine the effects of age group, the 4 person-level predictors of processing speed, inhibition, working memory, and education, and the 2 sentence-level predictors of DLevel and PDensity. Overall, young adults read aloud more rapidly than older adults ($M_O = 123$ wpm, SD $= 45$ wpm; $M_Y = 169$ wpm, SD $= 53$ wpm), as indicated by the significant estimate for age group (est. = −55.2, SE = 10.7, $p < .001$). Individuals who were faster, relative to their age group mean, on the processing speed measures, had an overall advantage for tracking during either condition as indicated by a significant estimate for sentence reading rates (est. = 24.6, SE = 3.6, $p < .001$). Individual differences in inhibition, working memory, and education did not affect oral reading rates at the level of individual sentences.

Sentence reading rate was also affected by the propositional density of the individual sentences (est. = −77.7, SE = 15.8, $p < .001$). Although the overall effect of DLevel was not significant, PDensity did interact with DLevel (est. = 8.4, SE = 3.4, $p = .015$) but this interaction was similar for young and older adults. Figure 2 illustrates this interaction. Sentence reading rate declined with the propositional density of individual sentences; this effect was greatest for simple 1-clause sentences at DLevel = 0 and gradually dissipated as DLevel increased, so that sentence reading rates for very complex sentences at DLevel = 7 were unaffected by propositional density. This pattern suggests that young and older adults' oral reading rates vary with the difficulty of extracting propositional information, as long as the sentences are grammatically easy to parse.

Tracking Performance

Tracking performance at the level of individual sentences during the dual task condition was assessed by 4 measures, TE or tracking error in pixels and TOT or time on target (percent), as well as the variability of each measure. For each measure of tracking performance crossed random effects regression was used to examine the effects of age group, the 4 person-level predictors of processing speed, inhibition, working memory, and education, and the 2 sentence-level predictors of DLevel and PDensity. Table 5 summarizes the results. At the sentence level, older adults' tracking was worse than young adults' tracking, as indicated by the significant estimate for age group for TOT (est. = −14.6, SE = 6.2, $p = .018$). Faster individuals had an overall advantage for tracking during either condition as indicated by significant estimates for TE (est. = −5.6, SE = 0.9), TE SD (est. = −2.9, SE = 0.4), and TOT (est. = 10.3, SE = 1.8), all $p < .01$. Individual differences in inhibition, working memory, and education did not affect tracking performance at the level of individual sentences.

TOT tracking performance was also affected by the DLevel of individual sentences (est. = 1.9, SE = 1.0, $p = .049$) and this effect of DLevel was different for older versus young adults, as indicated by the age group by DLevel interaction (est. = −3.0, SE = 1.4, $p = .035$). Although the overall effect of PDensity was not significant, PDensity did interact with DLevel (est. = 5.6, SE = 2.9, $p = .048$) and this interaction varied with age group as indicated by the significant estimate for the age group by DLevel by PDensity interaction (est. = −3.8, SE = 2.0, $p = .052$). Figure 3 illustrates this interaction. For young adults, as DLevel increased, the effect of PDensity also increased such that propositional density had no effect on TOT for simple sentences at DLevels ≤ 2 but TOT declined with propositional density for complex sentences at DLevels > 2. However, a different pattern is evident for older adults: propositional density had little effect on tracking performance regardless of sentence DLevel.

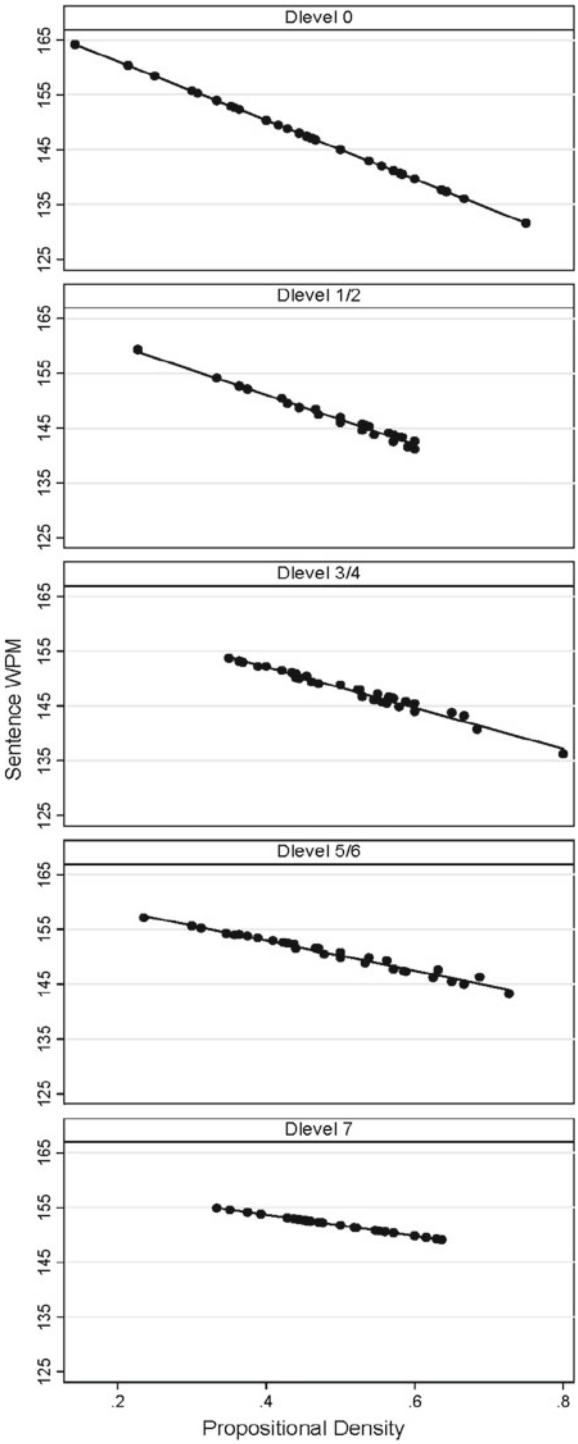

Fig. 2 Interaction of Grammatical Complexity (DLevel) and Propositional Density (PDensity) on Reading Rates. Estimates were Derived for Individual Sentences; For Convenience, DLevels 1 and 2, 3 and 4, and 5 and 6 were Collapsed

Table 5 Tracking performance during the dual task condition by young and older adults during the oral reading of individual sentences as well as tracking performance during the pauses preceding upcoming sentences

Measure	Young adults		Older adults	
	During sentences	Pauses preceding sentences	During sentences	Pauses preceding sentences
TE	16.0 (5.3)	15.5 (8.7)	25.0 (13.8)	23.8 (15.5)
TE SD	9.0 (4.0)	4.5 (3.5)	13.9 (7.4)	5.8 (5.3)
TOT	59.1 (19.8)	60.8 (35.8)	42.7 (24.1)	42.3 (38.9)
TOT SD	35.2 (8.8)	14.5 (12.5)	36.1 (12.4)	13.1 (13.3)

Means and SD (in parenthesis) are given

Sentence-level Summary

Reading rate and tracking performance for individual sentences varied with their grammatical complexity and propositional density. For both young and older adults, reading rate declined as propositional density increased, with the magnitude of the decline lessening with increasing grammatical complexity. Young adults could maintain good tracking accuracy when the sentences were simple and propositional density was low but their tracking accuracy declined with propositional density when the sentences were grammatically complex. On the other hand, older adults' tracking performance was poor and did not vary with either the propositional density or grammatical complexity of the sentences they were reading.

Individual Pauses

Readers pause between sentences when reading aloud and these pauses may reflect the costs of either planning the production of upcoming sentences or recovery from the production of preceding sentences. The duration of the pauses between sentences was analyzed with crossed random effects regression to examine the effects of age group, the 4 person-level predictors of processing speed, inhibition, working memory, and education, and the 2 sentence-level predictors of DLevel and PDensity, on one hand looking forward to the production of upcoming sentences, on the other looking backward at recovery from preceding sentences.

Overall, young adults' pauses were shorter than older adults' pauses ($M_O = .94$ s, SD $=$ 1.17; $M_Y = .50$ s, SD $= 0.40$), as indicated by the significant estimate for age group (est. = 0.4, SE = 0.1, $p = .005$). And faster participants paused more briefly between sentences than slower participants, resulting in a significant estimate for processing speed (est. = −0.2, SE = 0.1, $p < .001$).

Pause duration was also affected by planning the production of upcoming sentences such that both the DLevel (est. = 0.38, SE = 0.14, $p = .001$) and PDensity (est. = 0.38, SE = 0.14, $p = .045$) of upcoming sentences contributed to pause duration. For both groups, pause duration increased with propositional density, and with DLevel. These effects of sentence grammatical complexity and propositional density were greater for older adults, as indicated by the age group by DLevel interaction (est. = 0.2, SE = 0.1, $p = .015$) and the age group by PDensity interaction (est. = 1.0, SE= 0.5, $p = .038$). See Fig. 4. Recovery costs were negligible as pause duration was not affected by either the DLevel or PDensity of preceding sentences.

To investigate whether tracking performance during pauses also reflected planning or producing sentences, four measures of tracking performance were examined: TE or tracking error in pixels and TOT or time on target (percent), as well as the variability of each measure.

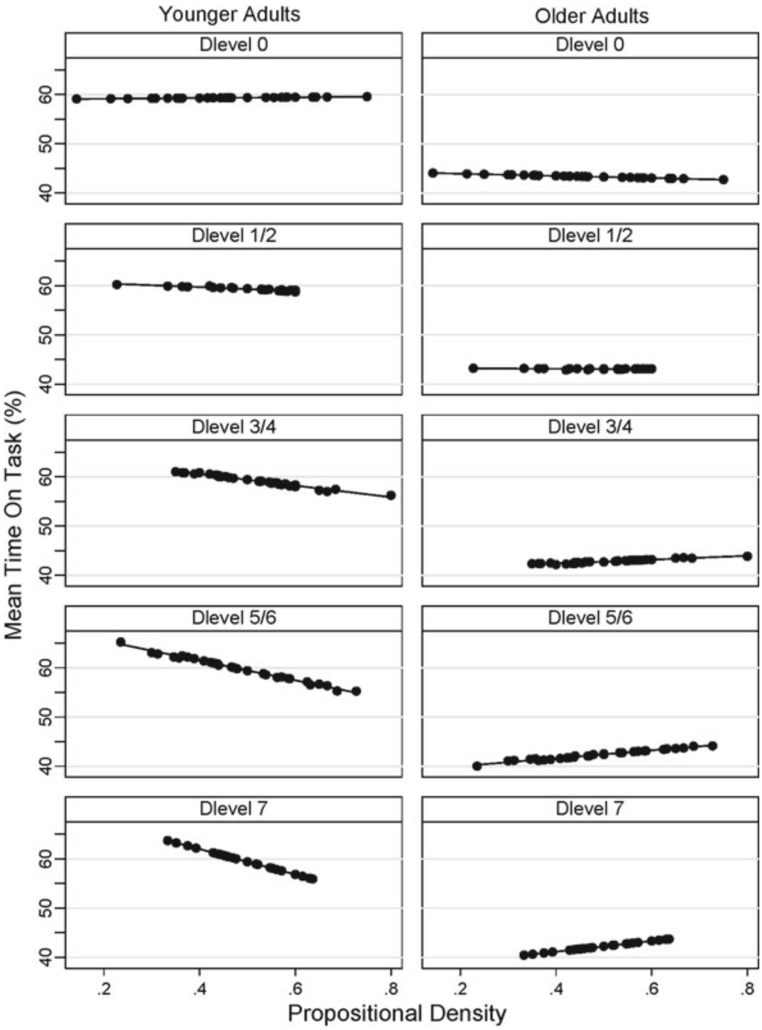

Fig. 3 Interaction of Grammatical Complexity (DLevel) and Propositional Density (PDensity) for Young versus Older Adults' Tracking TOT. Estimates were Derived for Individual Sentences; For Convenience, DLevels 1 and 2, 3 and 4, and 5 and 6 were Collapsed

For each measure of tracking performance during the pauses, crossed random effects regression was used to examine the effects of age group, the 4 person-level predictors of processing speed, inhibition, working memory, and education, and the 2 sentence-level predictors of DLevel and PDensity.

Table 5 summarizes the results. During pauses, older adults' tracking was worse than young adults' tracking, as indicated by the significant estimates for age group for TE (est. = 10.1, SE = 2.9), TE SD (est. = 2.4, SE = 1.4), TOT (est. = −19.4, SE = 8.8), and TOT SD (est. = −0.4, SE = 4.1), all $p < .05$. Individual differences in inhibition, working memory, and education did not affect tracking performance during the pauses but faster participants had an advantage over slower participants, TE (est. = −4.9, SE = 0.8, TE SD (est. = −0.5,

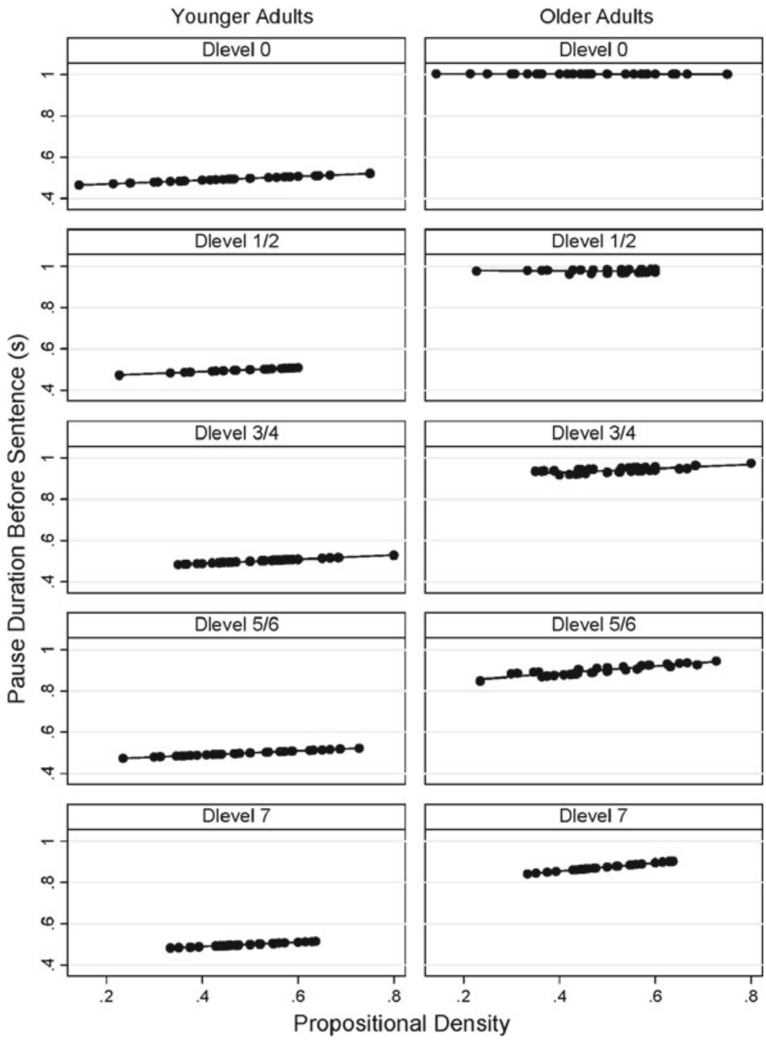

Fig. 4 Interaction of Grammatical Complexity (DLevel) and Propositional Density (PDensity) of Upcoming Sentences for Young versus Older Adults' Pause Durations. Estimates were Derived for Individual Sentences; For Convenience, DLevels 1 and 2, 3 and 4, and 5 and 6 were Collapsed

SE = 0.2), TOT (est. = 11.3, SE = 2.0), and TOT SD (est. = 1.4, SE = 0.6), all $p < .05$. However, TE and TOT, and their SDs, were not sensitive to the costs of planning or producing sentences varying in DLevel or PDensity, perhaps because the pauses were so short for these measures to be reliably calculated.

Pause Summary

Although tracking performance during pauses did not vary, pause duration varied with the grammatical complexity and propositional density of upcoming sentences. Readers, especially the older adults, paused longer before reading grammatically complex and propositionally dense sentences.

Discussion

A variety of experimental paradigms have been used to investigate how aging affects the processing of individual words, sentences, or paragraphs (Stine-Morrow et al. 2008). These paradigms typically rely on the analysis of time-comparing response times of young versus older adults to different psycholinguistic manipulations or in different experimental conditions. Pre-existing age group differences in the speed of processing, response time distributions, and intraindividual variability complicate the interpretation of age by condition interactions (Faust et al. 1999; Myerson et al. 2003; Ratcliff et al. 2000). Other paradigms manipulate the linguistic input by adding noise or through time-compression (Stine et al. 1986; Tun 1998) but they are also subject to criticism since the manipulations may differentially impact aging sensory systems, inducing ad hoc accommodations and processing strategies (Schneider et al. 2005). The current approach combines oral reading with pursuit rotor tracking to investigate age differences in linguistic processing.

Combining oral reading with pursuit rotor tracking provides 2 ways of looking at how aging affects linguistic processing. First, by looking at reading and tracking performance at the paragraph level, we see that older adults cannot match the performance of young adults, showing deficits of reading rate, reading comprehension, and tracking due to the demands of simultaneously reading and tracking the moving target. Regardless of condition, comprehension accuracy and reading rates improved when the paragraphs were longer, grammatically simpler, and more readable, suggesting that providing more information and increasing the ease at which that information can be processed benefits readers regardless of whether their attention is divided between reading and a secondary task or not. However, at this level of analysis, tracking performance, in general, did not vary with the overall length, grammatical complexity, or readability of the paragraphs, indicating that readers were able to maintain a consistent level of tracking performance even when they were reading long paragraphs filled with complex, propositionally dense sentences.

At this point we might conclude that aging results in a general deficit, affecting reading rate, reading comprehension, and dual task performance. We might also conclude that oral reading is sensitive to some aspects of language processing since reading rate does vary with psycholinguistic properties of the paragraphs such as grammatical complexity and propositional density. And we might conclude that combining pursuit rotor tracking with oral reading does not provide critical new insights into aging and linguistic processing.

When we examine performance at the level of individual sentences, a different picture emerges. We observe that both young and older adults modulated their oral reading rate with the ease of processing the sentences, slowing down as sentence propositional density increased. This modulation of reading rate gradually dissipated as the sentences became increasing complex grammatically, suggesting that just slowing down was no longer sufficient to overcome the processing demands imposed by the need to analyze complex syntactic structures while also unpacking a lot of propositional information.

This level of analysis also allows us to look at the pauses before upcoming sentences as a reflection of speech planning costs. When we do, we see that readers strategically lengthen their pauses before they read aloud difficult sentences, taking extra time to plan out how to articulate these sentences. The pattern suggests that readers were attempting to "buy time" for tracking while they were planning how to articulate demanding sentences. This finding contrasts with the reports by Kemper et al. (2010) using a controlled production task; after producing a long or grammatically complex sentence, speakers required more time to recover than after producing a shorter or simpler sentence and that these recovery costs were greater for older than young adults. Our results are, however, consistent with studies of the eye-voice span (Levin and Buckler-Addis 1979) which suggested that grammatical complexity affects how far ahead readers look when reading aloud. Kemper et al. (2011) reported that tracking performance during spontaneous production varied with the difficulty of preceding utterances and these recovery costs were also greater for older than young adults. We did not find that tracking performance during pauses varied with the difficulty of either upcoming or preceding sentences; perhaps because pause durations were so short, averaging less than 1 s, that our tracking measures, aggregated over 3 successive 100 ms segments, were insensitive to either planning or recovery costs.

It is at this sentence-level of analysis that we see marked differences in how young and older adults respond to the dual task demands. We see that young adults' tracking performance declined with the propositional density of the sentences, especially when the sentences were also grammatically complex. Hence, just reading more slowly was not sufficient to enable the young adults to fully process propositionally dense and grammatically complex sentences; they also shifted attention away from the demands of pursuit rotor tracking in order to do so. However, older adults' tracking performance did not vary with sentence difficulty. Older adults attempted to 'buy time' by reading more slowly as the sentences increased in difficulty and by paused longer before grammatically complex and propositional dense sentences. However, controlling the temporal dynamics of reading was not sufficient as the older adults' were unable to maintain their baseline level of tracking just by reading more slowly and pausing longer.

Prior research combining pursuit rotor tracking with spontaneous speech (Kemper et al. 2010) suggested that planning long, complex sentences was equally costly for young and older adults but production costs were greater for older adults. However, both young and older adults could choose their words and sentences, and both groups tended to use slow, short, simple sentences. In this study, young and older adults were forced to read aloud sentences varying in length, grammatical complexity, and readability. Both young and older adults experienced difficulty as they struggled to engage in pursuit tracking while reading aloud sentences. Greater planning costs for older adults emerged as they prepared to read aloud propositionally dense and grammatically complex sentences, ones that they are unlikely to produce spontaneously. Indeed, these planning costs for older adults were so severe that their tracking performance dropped from a baseline level of 83 % time on target to only 43 % time on target, even for simple sentences.

Prior studies of spontaneous speech suggested that older adults adopt a simplified speech style (Kemper 2006; Kemper et al. 2010), one characterized by the use of short and grammatically simple sentences. However, a variety of pragmatic and stylistic preferences may contribute to older adults' use of this simplified style. This approach of combining oral reading with pursuit rotor tracking has confirmed prior findings that aging affects the ability to plan grammatically complex and propositionally dense sentences. Even when older adults need only read aloud long, complexity, and propositionally dense sentences, they struggle to do so, speaking slowly and pausing longer between sentences. Although working memory capacity limitations have been implicated in prior studies of the effects of aging on linguistic processing (Kemper 2006), the present results suggest that individual differences in inhibition, working memory, and education do not affect dual task costs. Faster individuals do have an advantage but only in that they read more rapidly and pause more briefly between sentences. Thus, by examining dual task costs of reading aloud individual sentences, this study suggests that aging leads to a general loss of the ability to plan grammatically complex and propositionally dense sentences. Older adults rarely spontaneously produce such sentences and when forced to read them aloud, they are unable to simultaneously perform a simple visual-motor tracking task.

Acknowledgments This research was supported in part by grants from the NIH to the University of Kansas through the Mental Retardation and Developmental Disabilities Research Center, grant number P30 HD-002528, and the Center for Biobehavioral Neurosciences in Communication Disorders (BNCD), grant number P30 DC-005803 as well as by grant RO1 AG-025906 from the National Institute on Aging to Susan Kemper. Its contents are solely the responsibility of the author and do not necessarily represent the official views of the NIH. We thank Ruth Herman for her assistance with data collection and analysis. A suite of digital pursuit rotor applications is available upon request.

References

Brown, C., Snodgrass, T., Covington, M. A., Herman, R., & Kemper, S. (2008). Measuring propositional idea density through part-of-speech tagging. *Behavior Research Methods*, *40*, 540–545.

Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: The impact of distraction. *Psychology and Aging*, *6*, 533–541.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Engelhardt, P. E., Corley, M., Nigg, J. T., & Ferreira, F. (2010). The role of inhibition in the production of disfluencies. *Memory and Cognition*, *38*, 617–628.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, *125*, 777–799.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*, 210–233.

Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*, 221–233.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Reseach Methods, Instruments and Computers*, *36*, 193–202.

Grober, E., & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology*, *13*, 933–949.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 20, pp. 193–226). New York: Academic Press.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: foundations and illustrative examples. *Behavior Research Methods*, *39*, 110–117.

Hummert, M. L. (2009). Group identity, age stereotypes, and communication in a life span context. In H. Giles, S. Reid, & J. Harwood (Eds.), *Dynamics of intergroup communication* (pp. 41–53). New York: Peter Lang Publishers.

Kemper, S. (2006). Language in adulthood. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 223–238). Oxford: Oxford University Press.

Kemper, S., Bontempo, D., McKedy, W., Schmalzried, R., Tagliaferri, B., & Kieweg, D. (2010). Tracking sentence planning and production. *The Journals of Gerontology: Psychological Science*, *66*, 160–168.

Kemper, S., Hoffman, L., Schmalzried, R., Herman, R., Kieweg, D. (2011). Tracking talking: Dual task costs of planning and producing speech for young versus older adults. *Aging, Neuropsychology and Cognition, 18*, 257–279.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.

Levin, H., & Buckler-Addis, A. (1979). *The eye-voice span*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Myerson, J., Adams, D. R., Hale, S., & Jenkins, L. (2003). Analysis of group differences in processing speed: Brinley plots, Q-Q plots, and other conspiracies. *Psychonomic Bulletin and Review*, *10*, 224–237.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the life span. *Psychology and Aging*, *17*, 299–320.

Rabaglia, C. D., & Salthouse, T. A. (2012). Natural and constrained language production as a function of age and cognition. *Language and Cognitive Processes*, *26*, 1505–1531.
Ratcliff, R., Spieler, D., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin and Review*, *7*, 1–25.
Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271–276.
Roland, D., Dick, F., & Elman, J. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory & Language*, *57*, 348–379.
Rosenberg, S., & Abbeduto, L. (1987). Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults. *Applied Psycholinguistics*, *8*, 19–32.
Schneider, B. A., Daneman, M., & Murphy, D. R. (2005). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? *Psychology and Aging, 20*, 261–271.
Schneider, B. A., & Pichora Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 155–219). Mahwah, NJ: Lawrence Erlbaum Associates.
Selkirk, E. O. (1986). Derived domains in sentence phonology. *Phonology Yearbook*, *3*, 371–405.
Shipley, W. C. (1940). A self-administered scale for measuring intellectual impairment and deterioration. *Journal of Psychology*, *9*, 371–377.
Stata Corp. (2009). *Stata statistical software: Release 11*. College Station, TX: StataCorp LP.
Stine, E. A. L., Wingfield, A., & Poon, L. W. (1986). How much and how fast: Rapid processing of spoken language in later adulthood. *Psychology and Aging*, *1*, 303–311.
Stine-Morrow, E. A. L., & Soederberg Miller, L. M. (2008). Self-regulated reading in adulthood. *Psychology and Aging*, *23*, 131–153.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.
Swets, B., Desmet, T., Hambrick, D.Z., Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136*, 64–81.
Tagliaferri, B. (2005). Paradigm. www.perceptionresearchsystems.com.
Thornton, R., & Light, L. (2006). Language comprehension and production in normal aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (6th ed., pp. 261–287). Amsterdam: Elsevier.
Tun, P. A. (1998). Fast Noisy Speech: Age differences in processing rapid speech with background noise. *Psychology and Aging*, *13*, 424–434.
Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., & Cerella, J. (2003). Aging and dual-task performance: A meta-analysis. *Psychology and Aging*, *18*, 443–460.
Wechsler, D. (1958). *The measurement and appraisal of adult intelligence*. Baltimore: Williams and Wilkins.