2011

Working memory in simultaneous interpreters: Effects of task and age

Teresa Signorelli

Henk Haarmann

Loraine Obler
CUNY Graduate Center

Article

The International Journal of Bilingualism

International Journal of Bilingualism
16(2) 198–212

DOI: 10.1177/1367006911403200
ljb.sagepub.com
SAGE

Working memory in simultaneous interpreters: Effects of task and age

Teresa M. Signorelli
Marymount Manhattan College, USA

Henk J. Haarmann
University of Maryland, USA

Loraine K. Obler
City University of New York Graduate Center, USA

Abstract

This study examines whether interpreters have better working memory (WM) than non-interpreters, taking into account different WM components and the potential modulatory influence of age. Younger and older interpreters and non-interpreters were tested on reading span, nonword repetition, and order- and category-cued recall, using English, second-language materials. Articulation rate was also assessed. Interpreters outperformed non-interpreters in reading span and nonword repetition, but not cued recall and articulation rate. These results suggest that interpreters have better ability to manipulate information in working memory and process or store sub-lexical phonological representations, but have no advantage in short-term retention of words and their meaning. Compared to the other tested groups, younger interpreters were marginally better in nonword repetition and cued recall, suggesting that future studies on WM advantages in interpreters should consider the age factor.

Keywords

aging, articulation rate, cued recall, interpreting, nonword repetition, reading span, working memory

1 Introduction

Since simultaneous interpreting involves oral translation of ongoing spoken language in real time, it places a high demand on working memory (WM). That is, interpreting requires short-term

Corresponding author:
Teresa Signorelli, Department of Communication Sciences and Disorders, Marymount Manhattan College, New York, New York 10021, USA
Email: tsignorelli@mmm.edu

retention of words and their meanings in the source and target languages, while simultaneously processing this information for understanding, translation, and production (Cowan, 2000–1; Darò, 1994). Given the crucial role of working memory in simultaneous interpreting, one might expect a working memory advantage in simultaneous interpreters compared to non-interpreters, reflecting, perhaps, an aptitude factor or the effects of extensive practice (Darò, 1994). Several studies have tested the hypothesis of an interpreter advantage in WM using a variety of tasks and materials.

An interpreter advantage in WM was reported for word span with written presentation (Christoffels, de Groot, & Kroll, 2006) and for digit span with spoken presentation (Padilla, Bajo, Cañas, & Padilla, 1995; Tzou, Eslami, Chen, & Vaid, this issue). Such span tasks typically involve presenting words or digits in increasing increments for participants to repeat immediately. However, Köpke and Nespoulous (2006) did not find such an advantage for word or digit span with spoken presentation, while Chincotta and Underwood (1998) did not find it for digit span with written presentation.

An interpreter advantage was also reported in word list recall under conditions of articulation suppression with written presentation in immediate and delayed contexts (Padilla et al., 1995). An advantage for novice, though not professional, interpreters was reported with aural presentation (Köpke & Nespoulous, 2006). Such recall tasks typically involve listening to or reading a list of words, perhaps 12 to 16 words in length, while repeating a nonsense syllable such as ‘bla’ and then recalling what was presented. No interpreter advantage was found in probed recognition (Köpke & Nespoulous, 2006). In this task, participants listened to lists of four to 12 words and recalled the information following a phonological (rhyme) or semantic (category) probe.

All of the above studies assessed the storage function of verbal WM, that is, the maintenance of words and their meanings in a heightened state of accessibility. Still further studies comparing WM in interpreters and non-interpreters assessed the combined storage-plus-processing function of WM, using either reading span (Christoffels et al., 2006; Padilla et al., 1995; Tzou et al., this issue) or listening span (Köpke & Nespoulous, 2006; Liu, Schallert, & Carroll, 2004). In the reading span task, an individual must read aloud an increasingly larger set of sentences (processing) and maintain each sentence-final word in memory (storage) for subsequent recall (Daneman & Carpenter, 1980). The listening span task also often requires recall of sentence-final words, except that they are part of spoken sentences that the participant must process for meaning (e.g., semantic anomaly detection). Of the five studies employing storage-plus-processing tasks, four studies reported an interpreter advantage in WM (Christoffels et al., 2006; Köpke & Nespoulous, 2006; Padilla et al., 1995; Tzou et al., this issue), while one did not (Liu et al., 2004). It should be noted, though, that Köpke and Nespoulous found the advantage only for their novice, not professional, interpreters.

As this review of studies indicates, the evidence for an interpreter advantage in WM is mixed, albeit a majority of studies report it: an advantage was found in five out of seven studies reported here. An advantage manifested in seven of the 15 tasks involving assessment of the storage-only function of WM and in four of five tasks involving assessment of the storage-plus-processing function.

Although an interpreter advantage in WM was obtained in a majority of studies, one might conjecture about reasons why such an advantage was not consistently obtained. One possibility is that professional interpreters have a better ability to cope with response interference during recall compared to non-interpreters. This interpretation of the existing data might explain why professional interpreters tend to show an advantage when storage of words is assessed with word span (Christoffels et al., 2006; Padilla et al., 1995), but not when it is assessed with probed recognition (Köpke & Nespoulous, 2006). The reason is that response interference plays a larger role when individual test trials require recall of all list words, as in word span, compared to only a single list word, as in cued recall (Haarmann & Usher, 2001), which is similar to probed recognition.

This interpretation of existing data is speculative and, even if correct, does not account for all discrepancies in the data. Using a similar storage-only task, that is, word/digit span, an interpreter advantage in WM was reported by Padilla et al. (1995) with spoken digits, but not by Köpke and Nespoulous (2006) with spoken digits and words, nor by Chincotta and Underwood (1998) with written digits. These inconsistent results cannot be explained by a simple materials (words vs. digits) or modality (spoken vs. written presentation) factor. Likewise, using a similar storage-plus-processing task, an interpreter advantage in WM for listening span was obtained by Köpke and Nespoulous (2006) with novices, but not by Liu et al. (2004). Taken together, these discrepancies in the data suggest that there might be factors at work that were not controlled in previous studies that determine whether or not an interpreter advantage in WM is obtained.

One such factor could be the relative degree to which individuals rely on phonological and semantic short-term memory (STM).[1,2] There is convergent evidence from behavioral and neurophysiological data of individuals with and without neurological impairment that verbal WM consists not only of a phonological STM that stores word sound (Baddeley, 1986, 2000) but also of a semantic STM that stores word meaning (Cameron, Haarmann, Grafman, & Ruchkin, 2005; Davelaar, Haarmann, Goshen-Gottstein, & Usher, 2006; Haarmann, Cameron, & Ruchkin, 2002, 2003; Haarmann, Davelaar, & Usher, 2003; Martin, Shelton, & Yaffee, 1994). There is, furthermore, evidence that word span performance receives contributions from both STM components. Thus, individuals without interpreting experience might compensate for a relative deficit in one of these two STM components through greater reliance on the other component, potentially masking any interpreter advantage in WM in word span.

This consideration suggests that it is important to test interpreters and non-interpreters with WM storage tasks that differ in their relative reliance on phonological and semantic STM. The present study aimed to provide such a test through several design characteristics. First, interpreters and non-interpreters were tested with an order-cued recall task and category-cued recall task to obtain relative indices of phonological and semantic STM, respectively. In each case, presentation of a list of written words was followed by presentation of a single recall cue. In order-cued recall, individuals had to recall the word that immediately followed the cued word in the word list (Carter & Haarmann, 2001). In category-cued recall, they had to recall the one word that belonged to the cued semantic category (Carter & Haarmann, 2001; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Haarmann & Usher, 2001; Martin et al., 1994). The studies using these tasks cited above reported evidence that order- and category-cued recall involve greater reliance on phonological and semantic STM, respectively. To help evaluate this assumption, we included a word length manipulation. In a previous study, we found a greater word length effect in order- than category-cued recall, consistent with the assumption that such effects arise in phonological STM (Baddeley, 1986, 2000) and that order-cued recall induces greater reliance on phonological STM than category-cued recall (Carter & Haarmann, 2001). In another study, we found evidence that category-cued recall (but not word span) predicts on-line semantic integration, consistent with the interpretation of this cued recall task as providing a relative index of semantic STM (Haarmann, Davelaar, et al., 2003).

Two further properties of our study that helped to test whether an interpreter advantage in WM resides in phonological STM involved the inclusion of a nonword repetition test and measure of digit articulation rate. Nonword repetition provides an alternative measure of phonological STM (Gathercole, Willis, Baddeley, & Emslie, 1994; Klein, Watkins, Zatorre, & Milner, 2006) that complements order-cued recall. On the one hand, order-cued recall (but not nonword repetition) has the appealing property that it matched category-cued recall in stimulus presentation (identical) and response demands (one word per trial). On the other hand, nonword repetition is much less likely than order-cued word recall to include any reliance on semantic STM because nonwords do not have lexical-semantic representations. Differences in digit articulation rate are likely to be associated with faster sub-vocal rehearsal rates and, therefore, with better prevention of temporal decay of speech sounds in phonological STM (Ellis & Hennelly, 1980). If obtained, an interpreter-related advantage in digit articulation rate in conjunction with better phonological STM performance could therefore indicate that interpreters are better than non-interpreters in rehearsing speech sounds in phonological STM.

Another uncontrolled factor that might have contributed to discrepancies in the results of studies of WM differences between interpreters and non-interpreters is participants' age. Age-related declines in lexical retrieval (Nicholas, Obler, Albert, & Goodglass, 1985) and working memory, particularly in storage-plus-processing tasks, such as reading span, and in semantic STM, have been well documented in healthy adults (for review see Haarmann, Ashling, Davelaar, & Usher, 2005). One might expect that the daily work of interpreters permits substantial practice with WM that could buffer them from some or all of the WM decline in healthy aging. As a result, one might expect an interpreter advantage in WM to arise or be more prominent in older than younger adults, particularly in working memory tasks for which age-related declines have been observed. The ages of the participants in most of the studies surveyed above were not reported adequately to evaluate the extent to which age may contribute to contradictory findings vis-a-vis the question of whether there is a WM advantage for interpreters. We, therefore, decided to explore the joint effects of profession (interpreter vs. non-interpreter) and age on WM performance in this study and included reading span among the WM tasks. Other considerations for the discrepancies in the literature can be found in Köpke and Signorelli (this issue).

In summary, we sought to test differences in WM between interpreters and non-interpreters in tasks that highlighted various components of WM. Specifically, we asked whether interpreters show better performance than matched non-interpreters in the following skill areas:

- Sub-vocal articulation skills
- Nonword repetition skills
- Cued recall for phonological and semantic information

As well, we asked whether interpreters have greater reading spans than non-interpreters.

Should WM differences obtain between the two groups, we asked:

- How much of the difference is accounted for by semantic vs. phonological STM?
- How much does age contribute to those differences?

2 Methods

2.1 Participants: General information

The participants for the study included 47 multilingual adults who were selected for one of four groups, namely: (1) 12 younger interpreters (YI) (8 female) ranging in age from 30 to 40 years with a mean age of 34.5 ($SD = 3.5$); (2) 11 younger non-interpreters (YN) (6 female) ranging in age from 26 to 41 years with a mean age of 31.8 ($SD = 5.0$); (3) 13 older interpreters (OI) (9 female) ranging in age from 46 to 67 with a mean age of 56.2 ($SD = 7.3$); and (4) 11 older non-interpreters (ON) (6 female) ranging in age from 48 to 81 with a mean age of 63.6 ($SD = 11.6$). Non-interpreters had the same number of languages (4–5) as the interpreters, but had no training in interpretation and none of them had jobs that involved interpretation or translation.

All participants had normal speech, language, cognitive, hearing, and visual functioning as judged by a hearing screening, participant report, and the primary experimenter's professional judgment as a speech-language pathologist. Participants had comparable education levels save OI9 who had no formal education post high school. The interpreters had all been working in the field for more than one year, with the exception of YI3 who had eight months' professional experience. On average the YI group had 4.72 years' experience ($SD = 4.05$) and the OI group had 21.5 years' experience ($SD = 11.8$). All but two had had formal training; those two were older interpreters who were self-taught. All but four of the participants had also spent at least one year in an English-speaking country. On average, the YI group had spent 4.5 years ($SD = 5.5$), the YN group had spent 7.5 years ($SD = 4.9$), the OI group had spent 18.9 years ($SD = 16.7$), and the ON group had spent 25.7 years ($SD = 12.7$) living in an English-speaking country.

The groups, on average, began learning English at comparable ages. The younger interpreters had a mean age of acquisition (AOA) for English of 10.0 years ($SD = 4.9$), and the younger non-interpreters had a mean AOA of 12.5 years ($SD = 3.8$). The older interpreters had a mean AOA for English of 9.8 years ($SD = 3.6$), the older non-interpreters had a mean AOA of 11.8 years ($SD = 3.0$). The AOA across groups did not differ statistically. Univariate analyses of variance indicated that the age and the profession groups did not differ regarding when they began learning English, $F(1, 43) = 0.119$, p values of .731, MSe 1.82, and $F(1, 43) = 3.74$, .06, MSe 57.28, respectively.

Participants rated their English skills on a scale from 1 to 7. Only the endpoints of the scale were defined (a score of '1' indicated that an individual had limited knowledge of a given language and a score of '7' indicated having native-like knowledge in a given language). Participants rated them on five different parameters: overall skill, speaking ability, listening comprehension, reading comprehension, and writing. The groups were all comparable in their estimation of their language skills in English. Younger interpreters gave themselves an average overall rating of 5.8 ($SD = 0.7$). Younger non-interpreters gave themselves an average overall rating of 6.3 ($SD = 0.5$). Older interpreters gave themselves an average overall score of 6.3 ($SD = 0.7$). Older non-interpreters gave themselves an average overall rating of 6.3 ($SD = 0.6$). The English proficiency overall rating scores across groups did not differ statistically. The p values were .271 for the Age group comparison, .344 for the Profession comparison, and .221 for the Age × Profession interaction.

Testing took place over a single two- to two-and-a-half-hour experimental session. The articulation rate, nonword repetition, and reading span tasks were added at different points in the data collection process; as a result most (i.e., 62%), but not all participants participated in all four tasks. Demographic variables discussed above did not differ significantly across the tests. Table 1 lists the demographic data for each group of participants for each task.

2.2 Articulation rate

2.2.1 Participants. Thirty-three participants took part in the articulation-rate task. This included 11 younger interpreters, six younger non-interpreters, seven older interpreters, and nine older non-interpreters.

2.2.2 Procedures. Participants counted from one to ten, three times in a row, without pause between sets as quickly and as accurately as possible and their responses were recorded. Nine participants' measurements had to be made from a second or third production attempt as their initial attempt was corrupted by laughter, omitted numerals, or counting at a normal/slow rate. Spectographic analysis Cool Edit Pro software permitted calculation of rate measuring the length of sound wave in milliseconds at the start and close of speech production of the entire production string.

Table 1. Demographic data of participants for each task.

Participant group		Nonword repetition			Reading span		
		Age	Age of acquisition	English skill rating	Age	Age of acquisition	English skill rating
YI	Mean	34.67	10.56	6.11	34.73	9.82	5.73
	SD	3.12	4.25	0.93	3.58	5.06	0.90
YN	Mean	32.25	10.50	6.38	32.25	10.50	6.38
	SD	5.04	0.93	0.52	5.04	0.93	0.52
OI	Mean	56.75	9.25	6.25	59.33	9.33	6.67
	SD	7.85	3.28	0.71	7.03	3.67	0.52
ON	Mean	63.64	11.86	6.27	63.64	11.86	6.27
	SD	11.65	3.12	0.65	11.65	3.12	0.65
Participant group		Articulation rate			Cued recall		
		Age	Age of acquisition	English skill rating	Age	Age of acquisition	English skill rating
YI	Mean	34.40	10.20	5.80	34.50	10.00	5.83
	SD	3.60	5.27	0.92	3.50	4.86	0.94
YN	Mean	30.33	10.33	6.17	31.70	12.50	6.30
	SD	3.78	0.82	0.41	5.12	4.03	0.48
OI	Mean	56.43	10.71	6.43	56.33	9.67	6.33
	SD	7.93	0.76	0.53	7.66	3.75	0.65
ON	Mean	63.75	11.13	6.38	64.10	11.85	6.30
	SD	13.75	2.95	0.74	12.17	3.28	0.67

Y = Younger, O = Older, I = Interpreter, N = Non-interpreter.

2.3 Nonword repetition

2.3.1 Participants. Thirty-nine participants engaged in the Comprehensive Test of Phonological Processing (CTOPP) nonword repetition subtest (Wagner, Torgesen, & Rashotte, 1999). This included 11 younger interpreters, eight younger non-interpreters, 11 older interpreters, and nine older non-interpreters.

2.3.2 Stimuli and procedures. The subtest consists of 18 professionally recorded nonwords that progress in length from one to seven syllables. The stimuli were presented one at a time over headphones at a comfortable hearing level. Participants repeated each nonword after hearing it. Responses were recorded for off-line analysis. Scoring was strict in that no allowances were made for regional English or foreign accents. This protocol was employed to be consistent with the administration procedures delineated in the CTOPP manual calling for participants to repeat nonwords exactly as they are presented. Participants received a point for each nonword correctly articulated. Group averages of the individual participants' raw scores were compared.

2.4 Cued recall

2.4.1 Participants. Data from the cued recall task were analyzed for 46 participants as one, OI11, was only able to complete part of the cued recall task due to time constraints. Her data, consequently, could not be included in the analysis for this task.

2.4.2 Stimuli. The stimuli were those of Carter and Haarmann (2001) and included 96 lowercase printed words that belong to six semantic categories. The words in different categories were matched for their frequency of occurrence according to Kučera and Francis (1967) and for concreteness, familiarity, and imageability characteristics according to Coltheart (1981). Matching the stimuli in this way ensured comparable recall potential across the word pool. The matching criterion was a p value of .10 in a t-test. The stimuli were comprised of either short (one to two syllables) or long (three to five syllables) words to test for word length effects (i.e., better recall of shorter than longer words). Words from the word pool were pseudo-randomly assigned to each of two lists of 72 trials, one for the category-cued recall task and one for the order-cued recall task. Each trial consisted of a memory list of six words and a cue word. The word materials within each trial met the following constraints: each word belonged to a different semantic category, no words were repeated, and no words began with the same phoneme. For each of the two trial lists the presentation order of trials was randomized and cues were selected so that each serial position was cued a total of six times across trials at each word length. As explained below, the cue was either an order cue or a category cue depending on the recall task.

2.4.3 Procedures. The task began with an oral pre-reading phase. Participants saw the stimulus lists grouped by semantic category. They then read through the entire stimulus set twice while thinking about how each word fit into its category.

In the experimental task, participants saw lists of lower case words presented one at a time on a computer screen at a rate of one word every second. The last stimulus word in a trial list was immediately followed by a cue word in capital letters and a question mark (e.g., FOOD?). The stimulus lists for each task condition were identical with the exception of the cue word. The cue word in the phonological-order condition was one of the words from the presented stimulus list. Participants were asked to recall out loud the word in the list that occurred immediately after the cue word. For example, given the trial list, ‘apple, miner, wolf, dew, ache, flute’ and the cue, ‘APPLE?’, participants should recall the word ‘miner’ because ‘miner’ followed ‘apple’ in the stimulus list. The cue ‘FIRST?’ prompted recall of words in list-initial position. By contrast, the cue word in the semantic category condition was the label of the semantic category to which one of the words from the trial lists belonged. For example, the participant may have seen the words, ‘apple, miner, wolf, dew, ache, flute,’ and the cue ‘FOOD?’ and should recall the word ‘apple’ because an apple is a type of food item. Task order was randomized such that half of the participants in each group started with the semantic condition and the other half with the order condition. There were also no limits imposed on response time. For each group, the percentage of correct recall was determined for each participant for each cell of the serial-position (1 to 6) by word length (short, long) design. A post-test interview determined whether participants knew each word probed for recall.

2.5 Reading span

2.5.1 Participants. Thirty-eight participants were tested in the reading span task. This included 12 younger interpreters, eight younger non-interpreters, seven older interpreters, and 11 older non-interpreters. Two participants, YI8 and OI9, had scores that qualified them as outliers, being two standard deviations or more below the means for their respective groups, so their data were removed for group analysis.

2.5.2 Stimuli. This task replicated Christoffels et al.’s (2006) English reading span task. Christoffels created 42 English sentences, 11 to 13 words in length, that were partly derived from those used in

Harrington and Sawyer (1992). The final words across the sentences were selected to be balanced for length and frequency and not to rhyme. Two additional comparable practice sentences were generated for the current experiment to provide further training in the task. The stimuli were organized into three sets of two to five sentences. Stimulus presentation was incremental. First, three sets of two sentences were presented followed by three sets of three sentences, and so on up to three sets of five sentences.

2.5.3 Procedures. The sentences were presented one at a time on a computer screen. Each sentence was preceded by a warning tone occurring with a blank screen for 500 ms. The sentence was then presented in the center of the screen and participants read it aloud at a normal pace. Once a sentence was read, the experimenter triggered the next sentence by pressing a key. Following the last sentence in a set the screen read, '*****Please recall the final words*****,' to prompt the participant to recall all the final words from the set in any order. The experiment began with two practice sets of two sentences each, followed by the 42 experimental stimuli. Participants received one point for each word recalled. Raw scores were converted into averages for each participant. Group averages were calculated from these measures.

2.6 Analysis

For each of the tasks, scores were averaged per participant per condition. Participants whose average performance on a task was more than two standard deviations below their group means were eliminated from analysis of that task. The data points of the remaining participants were in by-subject ANOVAs that crossed the between-subject factors of Profession (interpreter, non-interpreter) and Age (younger, older). In the analysis of the cued recall data, these same ANOVAs also included examination of the within-subject factors of Cued recall task (phonological order, semantic category), Word length (short, long), and Serial position (1 through 6), which were crossed with each other and with Profession and Age. Interaction effects were examined by breaking them down into their simple main effects, using one-way ANOVAs. In the analysis of the cued recall data, a small number of trials were eliminated from data analysis in instances when the participant did not know the word targeted for recall (i.e., they had not heard of the word in English), as determined in their post-testing report. No participants had more than five percent of their trials discarded from data analysis. Pearson correlations were calculated between the overall performance measures of all tasks.

3 Results

Table 2 provides descriptive statistics for performance on all tasks for each of the four participant groups, including the group mean, standard error, number of participants included in the analysis, and number of participants excluded from the analysis due to outlier performance.

3.1 Articulation data

The analysis of the articulation rate data revealed no significant effects of profession or age.

3.2 Reading span and nonword repetition

Effects of profession, suggesting superior performance in interpreters, were found for reading span and nonword repetition. As shown in Figure 1, interpreters had better reading span (score of 86) than non-interpreters (score of 81), $F(1, 32) = 4.98$, $p = .033$, MSe = 44.25. Likewise, interpreters

Table 2. Descriptive statistics for performance on all tasks.

	YI group	YN group	OI group	ON group
Articulation rate	6.39 (1.06) [$N1$ = 10, $N2$ = 1]	6.68 (1.30) [$N1$ = 6, $N2$ = 0]	6.19 (1.03) [$N1$ = 7, $N2$ = 0]	6.75 (.86) [$N1$ = 8, $N2$ = 1]
Nonword repetition	60.49 (7.58) [$N1$ = 9, $N2$ = 1]	42.36 (11.48) [$N1$ = 8, $N2$ = 0]	47.22 (8.91) [$N1$ = 8, $N2$ = 1]	43.49 (16.88) [$N1$ = 11, $N2$ = 0]
Phonological cued recall	55.33 (14.57) [$N1$ = 12, $N2$ = 0]	47.80 (13.97) [$N1$ = 10, $N2$ = 0]	40.52 (19.66) [$N1$ = 12, $N2$ = 0]	50.19 (16.86) [$N1$ = 10, $N2$ = 0]
Semantic cued recall	70.24 (11.52) [$N1$ = 12, $N2$ = 0]	65.70 (14.95) [$N1$ = 10, $N2$ = 0]	61.34 (19.71) [$N1$ = 12, $N2$ = 0]	69.48 (14.43) [$N1$ = 10, $N2$ = 0]
Reading span	87 (5.00) [$N1$ = 11, $N2$ = 1]	81 (11.00) [$N1$ = 8, $N2$ = 0]	85 (4.00) [$N1$ = 6, $N2$ = 1]	80 (5.00) [$N1$ = 11, $N2$ = 0]

N1 is the number of participants included in the analysis. $N2$ is the number of participants excluded from analysis due to outlier performance.
Y = Younger, O = Older, I = Interpreter, N = Non-interpreter.
Articulation rate is expressed in average number of seconds. The other tasks reflect average number of words or nonwords recalled.

showed better nonword repetition (X = 54.25%, SD = 10.49) than non-interpreters (X = 43.02%, SD = 14.48), $F(1, 32) = 7.069$, $p = .012$, MSe = 149.56 (see Figure 2). This latter effect of profession was qualified by a trend toward an interaction with Age ($p = .089$) due to younger, but not older, interpreters outperforming non-interpreters in nonword repetition. The younger interpreters and non-interpreters recalled 10.89 and 8.14 nonwords, respectively, and the older interpreters and non-interpreters recalled 8.5 and 8.4 nonwords, respectively. This trend deserves further research on larger sample sizes if possible.

The reader will remember that there was a marginal effect of profession on age of acquisition of English (AOA) ($p = .06$), creating the possibility that the effect of profession on nonword repetition and reading span might be confounded with profession differences in AOA. However, the following additional results rule out this possibility. The effect of profession on nonword repetition and reading span remained significant in ANCOVAs (analyses of covariance) which statistically controlled for individual variation in AOA by entering AOA as a covariate: nonword repetition, $F(1, 31) = 6.82$, $p = .014$, MSe = 154.17, and reading span, $F(1, 31) = 4.79$, $p = .036$, MSe = 45.62. In addition, Pearson correlations showed that AOA was not reliably associated with the dependent variables in these analyses of the effects of profession. The correlation of AOA with nonword repetition was $r = -.006$ ($p = .97$) and the correlation of AOA with reading span was $r = -.06$ ($p = .71$).

3.3 Cued recall

In cued recall, there were no main effects of profession. However, younger interpreters showed a trend toward better performance than the other three groups across the two cued recall tasks, as indicated by a marginally significant interaction of Age and Profession ($p = .072$). In addition, performance on the phonological-order and semantic-category cued recall tasks was differentiated in terms of serial position and word length.

In phonological order cued recall, there was a main effect of Word length, $F(1, 40) = 63.39$, $p < .001$, MSe = 31,889, due to better recall for short than long words and Serial position, $F(5, 200) = 21.93$, $p < .001$, MSe = 14,785, due to standard primacy and recency effects (i.e., better recall for items in list-initial and list-final positions, respectively, compared to list-medial positions). The

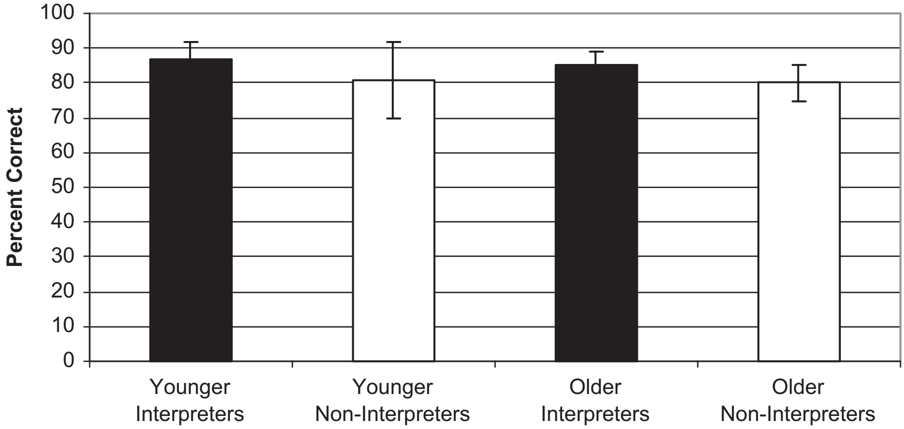

Figure 1. Reading span.
Error bars indicate standard deviation from the mean.

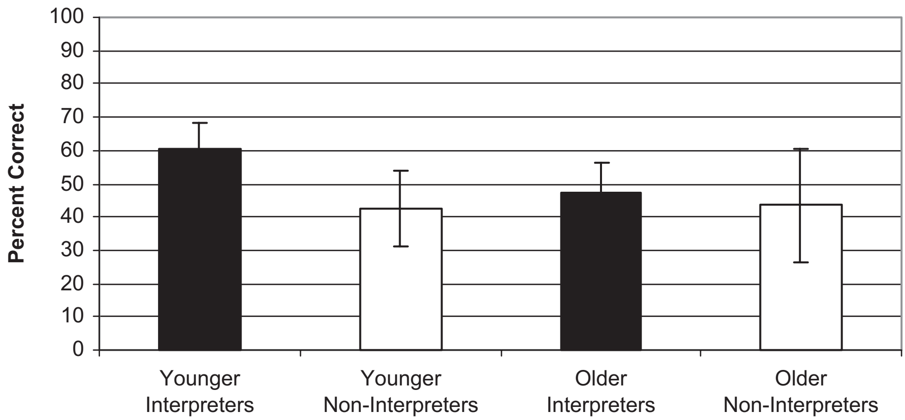

Figure 2. Nonword repetition (CTOPP).
Error bars indicate standard error of the mean.

word length effect was significant for the first, but not the second half of the list, explaining the interaction between Word length and Serial position, $F(5, 200) = 2.58$, $p = .028$, MSe = 935. The serial position effect also interacted with age, with younger individuals outperforming older individuals at serial positions with the poorest recall (i.e., list-medial positions 3 and 4), $F(5, 200) = 2.89$, $p = .015$, MSe = 1948. This interaction modulated the main effect of Age, $F(1, 40) = 139.60$, $p = .039$, MSe = 92,371.01.

In semantic-category cued recall, there were main effects of Word length, $F(1, 40)$ 11.65, $p = .0015$, MSe = 3794, and Serial position, $F(5, 200) = 35.88$, $p < .001$, MSe = 18,712, and an interaction of Word length and Serial position, $F(5, 200) = 5.33$, $p = .00012$, MSe = 1539. Short words were recalled better than long words and the serial position curve exhibited a recency, but not primacy effect: recall was not statistically different for positions in the first list half and increased across positions in the second last half. This recency-only finding in semantic-category cued recall contrasts with the primacy-plus-recency finding in phonological-order cued recall, resulting in an interaction of Task and Serial position, $F(5, 200) = 19.09$, $p < .001$, MSe = 9183.30. Within the recency part of the list, long words showed a steeper recency effect than short words, reflected by a significant word length effect at serial position 5 and explaining the interaction of Word length and Serial position.

The pattern of word length effects in the two cued recall tasks differed by age, as seen in the three-way interaction of Task, Word length, and Age, $F(1, 40) = 5.20$, $p = .028$, MSe = 1715.95. In semantic-category cued recall, older, but not younger individuals showed a word length effect, whereas in phonological-order cued recall both younger and older adults showed a word length effect. In terms of overall performance, the semantic-category cued recall task was easier than the phonological-order cued recall task, $F(1, 40) = 114.78$, $p <.001$, MSe = 86,344.53.

3.4 Task correlations

There was a strong and significant positive relationship between performance on the two cued recall tasks ($r = .78$, $p < .01$, $n = 44$). No other significant correlations among the various task pairings were evident.

3.5 Supplementary analyses

The statistical analyses were re-run on the three tasks for which participants with outlying performance had been omitted (i.e., articulation rate, nonword repetition, and reading span). Since the group numbers are small, the second analysis used a less rigid exclusion criterion of +/– 2.5 *SD* to see the potential and possible effect of increasing group numbers over those used when we ran the original analyses using +/– 2 *SD* as our cut-off for outlier status. The articulation rate results remained the same with the new analysis. For nonword repetition and reading span, the descriptive data pattern remained the same, in the direction of an interpreter advantage. Nonword repetition ($p = .13$) and reading span ($p = .12$) were no longer significant with these outliers added in.

4 Discussion

This study was aimed at testing the hypothesis that interpreters have better WM than non-interpreters. To help resolve discrepancies in the existing evidence for this hypothesis, this study included different WM tasks and evaluated the impact of age. We found evidence for an interpreter advantage in WM. Interpreters had better reading span and nonword repetition than non-interpreters. No group differences were obtained for order- and category-cued recall and articulation rate.

The finding of an interpreter advantage in reading span replicates the results in the studies of Tzou et al. (this issue), Christoffels et al. (2006) and Padilla et al. (1995). An interpreter advantage in listening span has been reported by Köpke and Nespoulous (2006) for their novice interpreters, but not by Liu et al. (2004). Thus, across the present and previous studies examining the combined storage and processing function of WM, a clear majority of studies (i.e., five out of six studies) find that interpreters outperform non-interpreters. This finding could indicate that interpreters have better executive control of language than non-interpreters. It would be of interest to test whether this advantage extends to better general executive control, regardless of whether linguistic materials are involved or not, in view of reports that experience with (early) bilingualism confers such an advantage compared to monolinguals (for a review see Bialystok, Craik, & Ryan, 2006).

We also obtained an interpreter advantage in nonword repetition. The nature of this effect may be interpreted best in the context of additional results in this study, in particular, the lack of an interpreter advantage in order-cued recall and articulation rate as well as the finding that interpreters and non-interpreters showed identical effects of word length in order-cued recall. The interpreter advantage in nonword repetition is likely to reflect better sub-lexical storage of speech sounds in WM, possibly due to better phonological, phonetic, or motor processing of unfamiliar word sequences (Klein et al., 2006). A lexical-level advantage in the WM storage of the phonological form of words seems to be excluded, since interpreters and non-interpreters performed similarly in terms of their performance in order-cued word recall. The assumption that this latter task assesses phonological STM is supported by the finding of robust word length effects, replicating Carter and Haarmann (2001). An interpreter advantage for sub-vocal rehearsal of words seems excluded, because the size of the word length effect was the same in interpreters and non-interpreters and because both groups were matched for digit articulation rate. Perhaps interpreters do not get to practice sub-vocal rehearsal of words more than non-interpreters because the real-time nature of simultaneous interpreting may not permit much (or any) time for it. This assumption remains to be tested. Our assessment of articulation rate involved digits, which are over-practiced. An interpreter advantage in articulation rate might emerge with use of less practiced materials.

Our finding of no interpreter advantage in order-cued recall and category-cued recall is consistent with previous results in a similar probed recognition task by Köpke and Nespoulous (2006) for professional interpreters. This finding could indicate that interpreters have identical phonological and semantic storage of words in WM, indexed by order-cued recall and category-cued recall, respectively. This interpretation raises the question why a majority of studies assessing word and digit recall with word span did find evidence for a WM advantage in interpreters. One possible answer was suggested in the introduction. At the time of recall, word span involves much greater response competition than cued recall, because a list of words has to be retrieved on each trial in word span, whereas only a single word has to be retrieved on each trial in cued recall. The retrieval of each word at each serial position in the response competes with other list words and non-list words for output and that response competition must be managed through executive control (Davelaar et al., 2005). This analysis of task demands suggests that interpreters might have a better ability to cope with response competition during recall of items from verbal WM than non-interpreters. It is of interest to note that the other tasks for which we obtained an interpreter advantage in WM, nonword repetition and reading span, might involve a high degree of response competition at the time of recall. In nonword recall, syllables from the previous trials and the current trial are likely to compete for output at each of the different syllable positions, especially as the number of syllables increases. In reading span, all sentence-final words have to be recalled during a trial and therefore are likely to compete for output during recall with one another and with similar, previously presented words. Testing this interpretation of the data will require parametric manipulation of the degree of response competition in the same WM task (i.e., same materials and task demands except for the degree of response competition at recall) in a study with interpreters and non-interpreters.

The finding of a marginally significant interaction of age and profession for nonword repetition shows a tendency for younger interpreters to perform best of the four groups. Further studies with more statistical power are needed to establish whether a more statistically reliable interaction can be obtained. Such studies are of theoretical and practical interest because an interaction could indicate that repetition of unfamiliar speech motor sequences conveys a sub-lexical phonological working memory advantage in interpreters that deteriorates with age. One might posit either that people with particularly good ability to keep words they do not understand in WM are likely to successfully complete the rigorous training interpretation generally requires, or that training enhances ability to keep speech in mind for those relatively recent graduates, who, with further practice in the field, permit this particular skill to deteriorate while the ability to keep meaningful materials in WM remains.

The apparent failure to find an age effect in reading span is of interest as well, since in L1 one typically finds such an effect (for a review see Meguro et al., 2000). However, here we tested multilinguals in a non-L1 language. That multilingual experience interacts with cognition in aging has now been reported from more than one source (Bialystok et al., 2006; Kavé, Eyal, Shorek, & Cohen-Mansfield, 2008). Clearly, further theoretical and empirical follow-up is warranted to assess the joint impact of multilingualism and aging on WM for L1 and L2 materials.

A final result concerning age was the finding of word length effects in category-cued recall in older but not younger adults. This finding together with previous evidence for an age-related decline in category-cued recall (Haarmann et al., 2005) provides further evidence for an age-related decline in semantic short-term memory in older adults (Haarmann et al., 2005). To compensate for such a decline, older adults in our study may have relied to a greater extent on phonological than semantic STM during category-cued recall, explaining the emergence of a word length effect in this task in older adults. A word length effect (i.e., better recall for short than longer words) was obtained in both younger and older adults for order-cued recall, supporting the assumption that these effects reflect the contribution of phonological STM. The finding of primacy and recency effects in order-cued recall and only recency in category-cued recall (cf. Haarmann & Usher, 2001) provides further support for our assumption that these two recall tasks engage different verbal retention processes.

Taken together, the findings of the present study and previous studies on WM in interpreters and non-interpreters are strongly indicative of a WM advantage in interpreters despite this study's relatively small group numbers and inconsistent subject participation at times across tasks. Future studies should test our hypothesis that this advantage is related to better processing of unfamiliar phonological sequences or executive control during retention and recall in verbal WM. Such studies would use multilingualism as a fruitful test bed for theories about the decomposition of verbal WM and are likely to result in insights that will benefit the training and job performance of professional interpreters. Future studies might also consider running more participants and testing for resistance to articulation suppression to explore the locus of interpreter WM advantages.

Acknowledgements

The authors wish to acknowledge Barbara Moser Mercer and Kilian G. Seeber for their invaluable assistance during the data collection process.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. See Cowan (2000–1) and Davelaar et al. (2005) for reviews of the construct of short-term memory (STM) as a system for maintaining long-term memory representations in a highly accessible active state in working memory.
2. Semantic STM might support the context- and time-sensitive integration of multi-modal information in working memory, a function achieved by the episodic buffer in Baddeley's (2000) theory of working memory.

References

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press/Clarendon Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.

Bialystok, E., Craik, F. I. M., & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1341–1354.

Cameron, K. A., Haarmann, H. J., Grafman, J., & Ruchkin, D. S. (2005). Long-term memory is the representational basis for semantic verbal short-term memory. *Psychophysiology, 42*, 643–653.

Carter S., & Haarmann, H. J. (2001). *Verbal short term memory: Evidence for separate phonological and semantic components*. Unpublished master's thesis, Department of Hearing and Speech Sciences, University of Maryland.

Chincotta, D., & Underwood, G. (1998). Simultaneous interpreters and the effect of concurrent articulation on immediate memory. *Interpreting, 3*, 1–20.

Christoffels, I., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language, 54*, 324–345.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33A*, 497–505.

Cowan, N. (2000–1). Processing limits of selective attention and working memory: Potential implications for interpreting. *Interpreting, 5*, 117–146.

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

Darò, V (1994). Non-linguistic factors influencing simultaneous interpretation. In S. Lambert & B Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation.* Philadelphia, PA: John Benjamins.

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112*, 3–42.

Davelaar, E. J., Haarmann, H. J., Goshen-Gottstein, Y., & Usher, M. (2006). Semantic similarity dissociates short- from long-term recency effects: Testing a neurocomputational model of list memory. *Memory and Cognition, 34*, 323–334.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71*, 43–51.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2*, 103–127.

Haarmann, H. J., Ashling, G. E., Davelaar, E. J., & Usher, M. (2005). Age-related declines in context maintenance and semantic short-term memory. *The Quarterly Journal of Experimental Psychology, 55A*, 34–53.

Haarmann, H. J., Cameron, K. A., & Ruchkin, D. S. (2002). Neural synchronization mediates on-line sentence processing: EEG coherence evidence from filler-gap constructions. *Psychophysiology, 39*, 820–825.

Haarmann, H. J., Cameron, K. A., & Ruchkin, D. S. (2003). Short-term semantic retention during on-line sentence comprehension: Brain potential evidence from filler-gap constructions. *Brain Research: Cognitive Brain Research, 15*, 178–190.

Haarmann, H. J., Davelaar, E. J., & Usher, M. (2003). Individual differences in semantic short-term memory capacity and reading comprehension. *Journal of Memory and Language, 48*, 320–345.

Haarmann, H. J., & Usher, M. (2001). Maintenance of semantic information in capacity-limited item short-term memory. *Psychonomic Bulletin & Review, 8*, 568–578.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25–38.

Kavé, G., Eyal, N., Shorek, A., & Cohen-Mansfield, J. (2008). Multilingualism and cognitive state in the oldest old. *Psychology and Aging, 23*, 70–78.

Klein, D., Watkins, K. E., Zatorre, R. J., & Milner, B. (2006). Word and nonword repetition in bilingual subjects: A PET study. *Human Brain Mapping, 27*, 153–161.

Köpke, B., & Nespoulous, J.-L. (2006). Working memory performance in expert and novice interpreters. *Interpreting, 8*, 1–23.

Köpke, B., & Signorelli, T. M. (this issue). Methodological aspects of working memory assessment in simultaneous interpreters. *International Journal of Bilingualism*.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Liu, M., Schallert, D. L., & Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting, 6*, 19–42.

Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language, 33*, 83–111.

Meguro, Y., Fujii, T., Yamadori, A., Tsukiura, T., Suzuki, K., Okuda, J., et al. (2000). The nature of age-related decline on the reading span task. *Journal of Clinical Experimental Neuropsychology, 22*, 391–398.

Nicholas, M., Obler, L., Albert, M., & Goodglass, H. (1985). Lexical retrieval in healthy aging. *Cortex, 21*, 595–606.

Padilla, P., Bajo, M. T., Cañas, J. J., & Padilla, F. (1995) Cognitive processes of memory in simultaneous interpretation. In J. Tommola (Ed.), *Topics in interpreting research* (pp. 61–71). Turku: Centre for Translation and Interpreting, University of Turku.

Tzou, Y., Eslami, Z. R., Chen, H., & Vaid, J. (this issue). Effect of language proficiency and degree of formal training in simultaneous interpreting on working memory and interpreting performance: Evidence from Mandarin–English speakers. *International Journal of Bilingualism.*

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *The comprehensive test of phonological processing.* Austin, TX: Pro-Ed.

About the authors

Teresa M. Signorelli is a bilingual speech-language pathologist. She is the Director of the Smadbeck Communication and Learning Center in the Communication Sciences and Disorders Department at Marymount Manhattan College. Her research interests include how the brain changes through normal development, in response to disorder or injury and following cognitive-linguistic, motor-speech and related therapies.

Henk J. Haarmann directs the cognitive neuroscience area at the University of Maryland Center for Advanced Study of Language (CASL). He obtained his PhD in cognitive science in 1993 from Radboud University Nijmegen in The Netherlands. His research studies address the nature of working memory and its role in language cognition, using psychological experimentation, correlation studies, electro-encephalographic recording, and computational modeling.

Loraine K. Obler is Distinguished Professor of Speech-Language-Hearing Sciences and Linguistics at the City University of New York Graduate Center. Her research on bilingualism yielded articles and *The Bilingual Brain: Neuropsychological and Neurolinguistic Aspects of Bilingualism* (with Martin Albert), and *Bilingualism across the Lifespan* (with Kenneth Hyltenstam). Her book *Language and the Brain* (with Kris Gjerlow) covers neurolinguistics more generally.