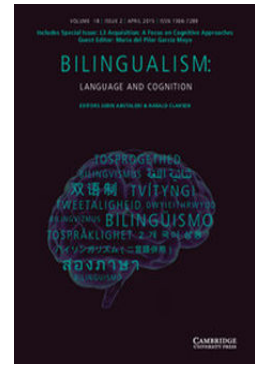

The relation between the working memory skills of sign language interpreters and the quality of their interpretations

RICK VAN DIJK, INGRID CHRISTOFFELS, ALBERT POSTMA and DAAN HERMANS

Bilingualism: Language and Cognition / Volume 15 / Issue 02 / April 2012, pp 340 - 350
DOI: 10.1017/S1366728911000198, Published online: 04 November 2011

Bilingualism: Language and Cognition 15 (2), 2012, 340–350 © Cambridge University Press 2011 doi:10.1017/S1366728911000198

The relation between the working memory skills of sign language interpreters and the quality of their interpretations*

RICK VAN DIJK
Helmholtz Institute, Experimental Psychology, Utrecht University, Utrecht, the Netherlands and University of Applied Sciences, Utrecht the Netherlands
INGRID CHRISTOFFELS
Leiden Institute of Brain and Cognition Department of Psychology, Leiden University
ALBERT POSTMA
Helmholtz Institute, Experimental Psychology, Utrecht University, Utrecht, the Netherlands
DAAN HERMANS
Royal Dutch Kentalis, Sint Michielsgestel and Behavioural Science Institute, Radboud University Nijmegen

(Received: January 7, 2010; final revision received: March 23, 2011; accepted: March 25, 2011; first published online November 4, 2011)

In two experiments we investigated the relationship between the working memory skills of sign language interpreters and the quality of their interpretations. In Experiment 1, we found that scores on 3-back tasks with signs and words were not related to the quality of interpreted narratives. In Experiment 2, we found that memory span scores for words and signs under oral articulatory suppression were related to the quality of interpreted narratives. We argue that the insensitivity to articulatory suppression in memory span tasks reflects the interpreters' ability to bind information from multiple sources in episodic memory. This enhanced ability leads to less reliance on the retention of information from the source language in memory during interpreting, and will positively affect the quality of interpretations (Padilla, Bajo & Macizo, 2005). Finally, in contrast to previous studies on the memory spans for signs and words (Hall & Bavelier, 2010), we found that the memory spans scores for spoken words and signs were equally large. We argue that the use of a large set of phonologically complex stimuli in the present study may have stimulated participants to use a speech-based code to store and retain the signs in short-term memory.

Keywords: sign language interpreters, working memory skills, quality of interpretations

Introduction

Many deaf adults and children make use of sign language interpreters to bridge the gap between them and the hearing world. Sign language interpreters can provide access to information for many deaf adults and children at work and at school. The desire for sign language interpreters in many countries is increasing as the educational spectrum has been broadened considerably in the last couple of decades, and more deaf students have moved from residential schools for the deaf to mainstream classes (Billies, Buchkoski, Kolvitz, Sanderson & Walter, 2003; Power & Hyde, 2002; Powers, 2002). At the same time, it has recently become quite clear that sign language interpreters who work with deaf children, adolescents and adults are not always equipped to provide them with full access to information (Schick, Williams & Kupermintz, 2006).

The present study builds upon a study we recently conducted on the quality of interpreted narratives produced by twenty-five experienced sign language interpreters (van Dijk, Boers, Christoffels & Hermans, 2011). In the present study our aim was to investigate to what extent individual differences we observed in the quality of narratives interpreted by experienced sign language interpreters are related to individual differences in their working memory skills. In the van Dijk et al. (2011) study, participants were instructed to interpret six narratives (with a mean duration of approximately 10 minutes) from spoken Dutch to Sign Language of the Netherlands (SLN), from spoken Dutch to Sign Supported Dutch (SSD) or from SLN to spoken Dutch. Note that SLN is the natural language of deaf people in the Netherlands, with its own grammar. Sign Supported Dutch is based on spoken Dutch and the grammar of spoken Dutch is dominant. SSD is usually expressed without the use of voice. The signs used in SSD are derived from SLN. In contrast to spoken Dutch or SLN, SSD is not a language of its own because it does not have a grammar of its own.

All participants were registered as qualified interpreters. Their ages ranged from twenty-nine to sixty-five years (average 41.9; standard deviation 8.8). Ten of them had deaf parents, and were raised with SLN. At the

* The authors thank the editor and two anonymous reviewers for their helpful and constructive comments.

Address for correspondence:
Rick van Dijk, Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, the Netherlands
rick.vandijk@hu.nl

time of the experiments, the participants were working between two and forty hours a week as sign language interpreters, with an average of 22.6 hours. Most of them had been working as sign language interpreters for more than ten years. Eleven interpreters used sign language in their home environments at the time of the experiment.

The interpreted narratives were judged on quality on two scales by five certified experienced interpreters who were working at the Institute for Sign Language and Deaf Studies in Utrecht. First, the judges were asked to rate the quality of the interpreted narratives on a 7-point-scale (1 – very poor, 7 – excellent). Second, the propositional accuracy of twenty pre-selected fragments in the narratives was assessed. The judges determined to what extent the twenty fragments from the source narrative were realized in the interpreted narratives. Subjects received null points if a fragment was not expressed, one point if the fragment was expressed but contained a mistake or omission, and two points if the fragment was fully expressed. Van Dijk et al. (2011) found that the quality of interpreted narratives produced by experienced interpreters was considerably higher in the Dutch-to-SLN and Dutch-to-SSD interpreting directions in comparison to the SLN-to-Dutch interpreting direction. To illustrate, the propositional accuracy of the SLN-to-Dutch (87.6%) and SLN-to-SSD (91.6%) interpretations were considerably higher than the propositional accuracy of the SLN-to-Dutch interpretations (49.5%).

Of most interest for the purpose of the present study, we found that the variation in the quality scores of the interpreted narratives in the three interpreting directions was quite large. To illustrate, the propositional accuracy of the interpreted narratives ranged from 22.5% to 77.5% for interpreting from SLN to Dutch, from 50.0% to 100% for interpreting from Dutch to SLN and from 71.0% to 100% for interpreting from Dutch to SSD.

As interpreting is one of the cognitively most complicated skills (Christoffels & de Groot, 2005), individual variation in the quality of the interpretations of experienced interpreters is likely to be related to individual differences in their cognitive skills. A prime candidate might involve working memory efficiency. The present study therefore focused on the working memory skills of sign language interpreters, and more specifically on the relation between sign language interpreters' working memory skills and their interpreting skills. It is likely that the processes that occur when interpreting from Dutch to SSD deviate strongly from the processes that occur during either Dutch-to-SLN or SLN-to-Dutch interpreting.[1] Therefore, in the present study we focused on the relation between sign language interpreters' working memory skills and the quality of their interpretations in the Dutch-to-SLN and SLN-to-Dutch interpreting directions only.

In Experiment 1, the working memory skills of sign language interpreters were assessed using the 3-back task. Three-back tasks are assumed to measure working memory capacity as they require the storage, the retention and the update of a dynamic rehearsal set. In Experiment 2, a series of classical memory span tasks was administered in which interpreters had to recall series of signs or words that were to be remembered under different conditions (oral articulatory suppression, manual articulatory suppression, no suppression). In both experiments, the accuracy scores of the interpreters were compared to the quality scores of the narratives they interpreted in the SLN-to-Dutch and Dutch-to-SLN interpreting directions (van Dijk et al., 2011).

As pointed out by Unsworth and Engle (2007), simple and complex memory tasks often measure the same processes (storage, maintenance and updating), but differ in the extent to which these processes are engaged. Consequently performance might depend more on a particular process in one task than in the other. To illustrate, La Pointe and Engle (1990) demonstrated that oral articulatory suppression affects simple memory tasks (e.g., word span tasks) more strongly than complex memory tasks (e.g., operation span tasks). This finding suggests that phonological processes such as active rehearsal operate to a greater extent in simple memory tasks than in complex memory tasks. Thus, the administration of different working memory tasks in the present study allows us to investigate the role of working memory in producing quality interpretations in more detail.

Experiment 1: working memory skills of sign language interpreters

In Experiment 1, we investigated the working memory skills of experienced sign language interpreters in relation to the quality of their interpretations. Gerver (1976), Moser (1978) and Gile (1997) have all emphasized the role of working memory in producing quality interpretations. To illustrate, Gile (1997) has argued that

[1] When interpreting from Dutch to SLN and from SLN to Dutch, interpreters have to comprehend the information from the source language and subsequently reformulate this information in the target language (e.g., constructing a syntactic frame and retrieving lexical items). In contrast, when interpreting from Dutch to SSD, interpreters have to comprehend the information from the source language. They subsequently have to (subvocally) reproduce the information from the source language and add lexical items from SLN to the (reproduced) spoken Dutch sentence. Thus, the processes that occur during interpreting from Dutch to SSD are fundamentally different from the processes that occur during interpreting between two languages.

simultaneous interpreting is a very complex cognitive skill, which requires the effective use of working memory resources. Consistent with this assumption, Christoffels, de Groot and Kroll (2006) found that experienced interpreters who work with two spoken languages (unimodal interpreters) scored higher on a reading span task, a listening span task and a word span task in comparison to highly proficient unimodal bilinguals without interpreting experience. Christoffels et al. (2006) concluded that professional interpreting experience is associated with elevated working memory capacity (see also Tzou, 2008). However, several other studies have failed to find consistent differences between professional unimodal interpreters, student unimodal interpreters and/or unimodal bilinguals without interpreting experience (Kopke & Nespoulos, 2006; Liu, Schaller & Carroll, 2004). For instance, Liu et al. (2004) found that the general working memory skills of professional unimodal interpreters were equal to the general working memory skills of beginning student unimodal interpreters and advanced student unimodal interpreters. Thus, the role of working memory capacity in producing quality interpretations in unimodal interpreters is not uncontroversial (Timarova, 2008). In most of these studies, the working memory skills of experienced interpreters were compared to the working memory skills of non-interpreters or student interpreters (but see Liu et al., 2004). However, to investigate the role of working memory in interpreting, it is necessary to directly relate working memory skills to interpreting performance across individual interpreters (Timarova, 2008). This approach was also taken in the present study.

In Experiment 1, two 3-back tasks (for a review, see Owen, McMillan, Laird & Bullmore, 2005) were administered to sign language interpreters, one with signs and one with spoken words. In the 3-back task, sequences of verbal/non-verbal stimuli are presented. For each item in the sequence, participants judge whether it matches the item presented three items earlier in the sequence.

Of special interest to us are the so-called lure trials in these 3-back tasks; trials on which the target is the same as the target on the 2-back or 1-back trial, but different to the target on the 3-back trial. Conway, Kane and Engle (2003) have argued that it is difficult to correctly reject such targets in 3-back tasks, and that performance on lure trials in particular may be strongly related to working memory performance.

The experienced interpreters who took part in the van Dijk et al. study (2011) also participated in the present study. To investigate whether or not working memory skills contribute to successful interpreting, the interpreters' scores in the 3-back tasks were related to the quality scores ascribed by judges to their interpreted narratives in the SLN-to-Dutch and Dutch-to-SLN interpreting directions (van Dijk et al., 2011).

Method

Participants

Twenty-five experienced interpreters, 20 females and 5 males, participated in the experiment. Their average age was 41;9 years (range 29 to 65). They worked as interpreters between 2 and 40 hours a week (average 22.6 hours a week), and were paid for their participation. Fifteen interpreters had hearing parents. The other 10 interpreters had deaf parents, and were raised with spoken Dutch and SLN.[2]

Materials

Twelve spoken Dutch words and 12 SLN signs were selected for the 3-back tasks. The 12 spoken words and 12 signs were translation equivalents. The words and signs were taken from 12 different semantic categories. The words were highly frequent in Dutch, and were judged to be known by the interpreters in SLN and in Dutch. For each task, a sequence of 16 practice trials and a sequence of 48 experimental trials were generated. Both sequences consisted of 33.3% matches (accept), 20.8% lure trials (reject) and 45.8% control trials (reject). For the spoken words, the onset of the words was defined as the onset of voicing. For the signs, the onset and offset of the signs were defined as the resting position of both hands, rather than the actual location of articulation, which is often difficult to define.

Design

The order in which the two tasks were administered to the group of interpreters was varied. Half of them first received the 3-back task in Dutch and then in SLN.

Procedure

Participants were tested in the sign language laboratory of the Institute for Sign Language and Deaf Studies in Utrecht, with the exception of one experienced interpreter who was tested at home. This sign language laboratory has 16 booths, all equipped with the audiovisual equipment that is needed to train sign language interpreters. In most of the cases, one or two interpreters were tested during the same day in this laboratory. We instructed participants to decide as quickly and accurately as possible whether or not a spoken Dutch word (3-back task Dutch) or sign (3-back task SLN) matched the spoken word or sign presented three trials earlier by pressing a "yes" or "no" button. The instruction was followed by the practice session consisting

[2] Note that no differences were observed in Experiments 1 and 2 between the interpreters with deaf parents and interpreters with hearing parents. Although this finding may be surprising, we think this issue is beyond the scope of this article. We therefore do not report the results as a function of the parental hearing status of the interpreters.

Table 1. *Average accuracy scores and standard deviations as a function of trial type in both languages in Experiment 1.*

	SLN		Dutch	
	Accuracy (Standard deviation)		Accuracy (Standard deviation)	
All	82.90	(8.58)	83.17	(8.42)
Reject (non-lure)	97.02	(4.96)	95.99	(5.56)
Reject (lure)	70.87	(16.31)	80.00	(18.17)
Accept	73.64	(23.91)	69.94	(13.96)

Table 2. *Correlations between a combined measure of quality of the interpreted stories (van Dijk et al., 2011) and the accuracy scores in the 3-back tasks in Dutch and SLN (** = $p < .05$).*

3-back		Interpreting condition	
Language	Trial	SLN to Dutch	Dutch to SLN
SLN	All	.05	.18
	Lure	.08	.24
Dutch	All	.25	.26
	Lure	.28	.29

of 16 trials. Next, the experimental session consisting of 48 trials was administered. A trial consisted of the following sequence: First, a fixation point was presented on the computer screen for 500 ms. Next, the spoken word (through headphones) or a video of the sign (on the computer screen) was presented. No time-out was set. Inter-stimulus interval was set at 1,000 ms. In all, the session lasted for approximately 15 minutes.

Results

The results (as a function of trial type) are listed in Table 1. Analyses of variance were conducted on the accuracy scores in all conditions and in the lure-conditions only with Language (SLN and Dutch) as the within-subject variable. The main effect of Language failed to reach significance in the analyses of the accuracy scores ($F(1,24) < 1$) of all trials. However, the main effect of Language reached significance in the analyses of accuracy on the lure-trials only ($F(1,24) = 6.08$, $MSE = 113.1$, $p < .05$), such that accuracy scores were higher in Dutch in comparison to SLN.

Correlational analyses

To compare the performance of sign language interpreters on the 3-back tasks with their interpreting skills, correlations were computed between the accuracy scores and the quality measures of the interpreted narratives in the Dutch-to-SLN and SLN-to-Dutch interpreting directions. In the van Dijk et al. study (2011), two measures of the quality of interpretations were assessed: a subjective scale in which the narratives were rated on quality on a 7-point scale (1 = very poor, 7 = excellent), and (2) an objective scale in which the propositional accuracy of the interpreted narratives was judged. The scores on both measures were combined into a third measure for the purpose of the present study. This third measure was computed as follows. First, the subjective scores on the 7-point scale were converted to a score between 0 and 100 ((score-1) * (100 / 6)). Second, the raw propositional accuracy scores (minimum 0, maximum 40) were converted into propositional accuracy percentage scores. Then, the third measure was computed by averaging across the propositional accuracy percentage scores and the reverted subjective scores. Next, correlations were computed between the quality of the interpreted narratives in the two interpreting directions (SLN-to-Dutch and Dutch-to-SLN) as obtained in the study by van Dijk et al. (2011) and the accuracy scores of the interpreters in SLN and Dutch for all trials and lure trials only. The results are listed in Table 2.

The correlational analyses revealed that accuracy scores in the 3-back task on all trials and the lure trials only did not correlate significantly with the quality of interpreted narratives in the Dutch-to-SLN and SLN-to-Dutch interpreting directions.

Discussion

Experiment 1 yielded the following results. First, no main effect of Language was observed in the overall analyses of the accuracy scores. This finding suggests that 3-back tasks in SLN and Dutch are equally difficult for sign language interpreters. Yet the analyses of the lure trials indicated that correctly rejecting a lure trial in SLN is more difficult than correctly rejecting a lure trial in Dutch. Since lure trials in *n*-back tasks are arguably the most sensitive measure of working memory skills (Conway et al., 2003), overall we may conclude that the working memory capacity of these interpreters is slightly better in Dutch than in SLN.

One possible explanation for this finding is that working memory for speech is more efficient in retaining serial order information in comparison to working memory for signs (e.g., Boutla, Suppalla, Newport & Davelier, 2004), leading to an advantage for the 3-back tasks in Dutch on the lure-trials compared to the SLN condition.

Alternatively, the difference between the accuracy scores in Dutch and SLN in the lure conditions may simply be caused by a difference in the length of the stimuli. On average, signs are considerably longer in duration than spoken words (Bellugi & Fischer, 1972). In tasks that assess working memory skills, differences in the duration of stimuli can indeed affect the working memory scores (Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005). In other words, the differences in the accuracy scores in SLN and Dutch on the lure trials may be caused by differences in the duration of the SLN signs and the spoken Dutch words.

Second and most importantly, we found that the interpreters' accuracy scores in the 3-back tasks did not correlate significantly with the quality of interpreted narratives in both interpreting directions. Thus, differences in the working memory skills of experienced sign language interpreters are not related to differences in the quality of their interpretations. As we pointed out in the 'Introduction', models of interpreting generally emphasize the role of working memory in producing quality interpretations (Gerver, 1976; Gile, 1997; Moser, 1978), but empirical studies on the working memory skills of interpreters have yielded inconsistent results (for a review, see Timarova, 2008). We speculate here that *n*-back tasks are not the most ideal tasks to measure working memory skills (Conway et al., 2005; Kane, Conway, Miura & Colflesh, 2007). Although *n*-back tasks have become the gold standard in neuroimaging research investigating the neurological substrates of working memory (Owen et al., 2005), the validity of *n*-back tasks as a working memory measure is not uncontroversial (Kane et al., 2007; but see Schmiedek, Hildebrandt, Lövdén, Wilhelm & Lindenberger, 2009). To illustrate, several recent studies have found either that *n*-back tasks correlate more strongly with simple span tasks than with complex span tasks (Dobbs & Rule, 1989; Roberts & Gibson, 2002) or that *n*-back tasks correlate only moderately with complex span tasks (Conway et al., 2005; Kane et al., 2007; Miller, Price, Okun, Montijo & Bowers, 2009). Conway et al. (2005) argued that *n*-back studies may be a more appropriate indicator of the construct measured by simple word span tasks (short-term memory capacity) rather than the construct measured by complex span tasks (working memory skills). In other words, the absence of a significant positive correlation between the interpreters' scores in the 3-back tasks and the quality of their interpreted narratives in the SLN-to-Dutch and Dutch-to-SLN interpreting conditions may also be caused by limitations of the 3-back task as a valid measure of working memory skills.[3]

[3] Quite remarkably, the accuracy scores on the lure trials in the 3-back tasks correlated with the quality of interpreted narratives in the Dutch-to-SSD condition. We have no explanation for this finding.

Experiment 2: sensitivity to oral and manual articulatory suppression

In Experiment 2, we further investigated to what extent working memory skills of sign language interpreters are related to the quality of their interpretations by assessing sensitivity to articulatory suppression in serial order memory span tasks. Padilla, Bajo, Cañas and Padilla (1995) compared the effects of oral articulatory suppression on the (free) recall of unrelated words among experienced interpreters, second- and third-year student interpreters, and highly proficient bilinguals without interpreting experience. Padilla et al. (1995) found that student interpreters and highly proficient bilinguals showed significant decrements in recall under oral articulatory suppression. In contrast, recall of unrelated words by professional interpreters was hardly affected by oral articulatory suppression. Padilla et al. (2005) replicated this finding, but they also obtained clear evidence that the availability of long-term language-specific knowledge underlies the relative insensitivity to oral articulatory suppression.[4] Thus, these studies suggest that unimodal interpreters are relatively insensitive to the detrimental effects of concurrent articulation on the retention of information in short-term memory.

In explaining why unimodal interpreters are relatively insensitive to oral articulatory suppression, Padilla et al. (2005) have argued that the most important source of difficulty in interpreting between two (spoken) languages is the simultaneity of production and comprehension (see also Christoffels, 2006). Although interpreters can exploit the natural pauses in the source language to facilitate their task (Barik, 1975), they must still speak and listen at the same time for about 70% of the time. The simultaneity of language comprehension and production will make it more difficult to retain information from the source language in memory through active rehearsal (Christoffels, 2006). Therefore, in order to effectively retain information from the source language in memory during interpreting, interpreters need to be less dependent upon the storage of information from the source language in short-term memory.

In this respect, Padilla et al. (2005) and Christoffels (2006) have pointed to the role of the episodic buffer (Baddeley, 2000, 2003). In this model, the episodic buffer is a limited capacity system that binds information from a range of sources (e.g., visual memory, long-term memory, short-term memory) into a single complex structure or episode. Padilla et al. (2005) argued that language-specific long-term knowledge is generally more easily available for interpreters (Bajo, Padilla & Padilla, 2000). This

[4] Similarly, Christoffels (2006) found that the memory spans for L1 Dutch words under oral articulatory suppression of L1 Dutch–L2 English university students without interpreting experience were significantly correlated with their L2-to-L1 interpreting performance.

will make it easier for them to construct episodes in the buffer and thus reduces the reliance upon short-term memory during interpreting (see, for a very similar suggestion, Christoffels, 2006). In other words, Padilla et al. (2005) have argued that the ability to bind information in episodic memory is elevated in interpreters. This would enable interpreters to retain information from the source language in memory during interpreting. The relative insensitivity for articulatory suppression in word span tasks reflects the interpreters' elevated ability to bind information in episodic memory.

In Experiment 2, we investigated to what extent the sensitivity of sign language interpreters to oral articulatory suppression and manual articulatory suppression in serial order memory span tasks are related to the quality of their interpretations. Wilson and Emmorey (1997a, 1997b, 2000) have shown that deaf native signers can store signs in short-term memory using a visual–spatial phonological code. For those deaf signers, the retention of signs in short-term memory is controlled by a manual rehearsal mechanism. This manual rehearsal mechanism can be disturbed by hand movements, resulting in a manual articulatory suppression effect. The rationale of Experiment 2 was the following: if the ability to bind information in episodic memory is important for sign language interpreters in producing high-quality interpretations, the sensitivity to oral articulatory suppression (for spoken words) and manual articulatory suppression (for signs) should be related to the quality of interpreted narratives in both interpreting directions.

Method

Participants

The participants were the same as in Experiment 1.

Materials

Forty-five spoken Dutch words and their SLN translation equivalents were selected for the short-term memory tasks. The words were selected from nine different semantic categories (e.g., fruits, colours, clothes, animals). The words were highly frequent in Dutch, and were judged to be known by the interpreters in SLN and in Dutch. For each task, 24 sequences of words/signs were constructed in 4 series of 3–8 words/signs. Care was taken that none of the 144 sequences were the same. The words or signs in each sequence were not semantically related. Furthermore, the words and signs in each sequence were not phonologically related in the language in which they were presented (respectively, spoken Dutch and SLN). Finally, we also minimized (as much as possible) the phonological relatedness of the translation equivalents of the spoken Dutch words and SLN signs in a sequence in, respectively, SLN and spoken Dutch.

Design

The order in which the six tasks were administered to the group of interpreters was counterbalanced. Half of them first received three short-term memory tasks in Dutch, and after a short pause the remaining three tasks in SLN. Within each of these two sessions, the order of the conditions were administered to the subjects (Control, Oral articulatory suppression, Manual articulatory suppression) was also varied.

Procedure

Participants were tested in the sign language laboratory of the Institute for Sign Language and Deaf Studies in Utrecht, with the exception of one interpreter who was tested at home. The responses of the interpreters were recorded on videotape. In most of the cases, one or two interpreters were tested during the same day in this laboratory.

In each session, participants received the instructions for the three tasks. In the control condition, they were told that a series of signs / spoken words would be presented on the computer screen / through headphones, and that they were required to remember the order of the signs / spoken words as accurately as possible. They were also instructed to repeat the sequence of SLN signs in SLN and the sequence of spoken Dutch words in Dutch immediately when the sequence was finished. In the oral articulatory suppression condition participants received the same instruction, with the exception that they were instructed to repeatedly produce the Dutch nonsense word "DA" while they watched the signs or listened to the words. In the manual articulatory suppression condition, the instruction participants received was to repeatedly produce an SLN nonsense sign, a two-handed alternating movement forward with a hand-internal movement. In both suppression conditions, this was exemplified by the test administrator.

For the short-term memory tasks with spoken words, a trial looked as follows. First, a fixation point was presented on the computer screen. Next, the sequence of (three to eight) spoken words was presented through a headphone with an inter-stimulus interval of 1,000 ms. Then, participants repeated the sequence, and pressed a symbol on the computer screen to go to the next trial. For the short-term memory tasks with signs, a trial looked as follows. Again, a fixation point was first presented on the computer screen. Then, the sequence of recorded signs was presented on the computer screen with a inter-stimulus interval of 1,000 ms. Next, participants repeated the sequence, and mouse-clicked a symbol on the computer screen to go to the next trial. The first four trials in each task (with a length of three signs or three words) were practice items, and were excluded from further analysis. In all, the session lasted for approximately 90 minutes, including a pause about halfway.

Table 3. *Average serial order memory span scores and standard deviations in the three conditions of the STM experiments in SLN and Dutch. NSC* refers to the Number of Series Correctly recalled * Series Length. MLC refers to the Maximum List length Correctly recalled.*

	SLN		Dutch	
Condition	NSC*	MLC	NSC*	MLC
Control	38.1 (13.9)	6.24 (0.72)	37.9 (11.6)	6.21 (0.93)
Articulatory suppression	17.4 (17.1)	4.92 (1.44)	25.4 (11.7)	5.52 (0.87)
Manual Suppression	37.9 (14.1)	6.13 (0.90)	36.2 (14.8)	6.02 (0.95)

Analyses

For a sequence in each task, it was determined whether or not interpreters had successfully managed to reproduce the whole sequence. For each whole sequence participants successfully repeated, the number of points equalling the length of the sequence was given. To illustrate, five points were given for a sequence of five words or signs which were repeated in the correct order. If participants made one or more errors in the repetition of a sequence (deleting an item, substituting two items in a sequence) they received no points. We will refer to these scores as "NSC*" (Number of Series Correctly recalled * Series Length) scores. In contrast to more traditional serial order span scores measures (e.g., the longest list length at which a correct serial recall is observed, WAIS: Wechsler, 1955), the NSC* measure of memory span also takes the consistency of serial order recall by participants into account. However, in order to compare the span scores of the sign language interpreters in the present study to the span scores of bimodal bilinguals for spoken words (average span scores 6−7 words) and signs (average span scores 4−5) reported in the literature (e.g., Boutla et al., 2004), span scores were also computed using the more traditional method, which will be referred to as "MLC" (Maximum List length Correctly recalled).

Results

The results are listed in Table 3. Analyses of variance were conducted with Language (SLN and Dutch) and Condition (Control – OAS (Oral articulatory suppression) and MAS (Manual articulatory suppression)) as within-subject variables on the NSC* scores (Number of Series Correctly recalled * Series Length). The main effect of Condition reached significance ($F(2,48) = 37.32$, $MSE = 116$, $p < .001$). Post-hoc Bonferoni tests revealed that, averaged across SLN and Dutch, participants could recall more words in the Control (38.0) and the MAS (37.0) conditions in comparison to the OAS condition (21.4). Furthermore, the main effect of Language failed to reach significance ($F(1,24) = 1.38$, $MSE = 114.4$, $p > .1$).

Finally, the two-way interaction between Language and Condition reached significance ($F(2,48) = 4.53$, $MSE = 75.4$, $p < .05$). Separate analyses per Language showed that Condition reached significance in SLN ($F(2,48) = 34.10$, $MSE = 103.6$, $p < .001$) and in Dutch ($F(2,48) = 13.02$, $MSE = 88.1$, $p < .001$). Post-hoc Bonferonni tests (with alpha = .05) revealed that, in both languages, participants could recall more words in the Control and MAS conditions in comparison to the OAS condition. Next, additional analyses were conducted on the effect of Language for the three conditions of the memory span tasks separately. These analyses revealed a main effect of Language in the OAS condition ($F(1,24) = 7.42$, $MSE = 108.5$, $p < .05$), but no effect of Language in the Control condition ($F(1,24) < 1$) or in the MAS condition ($F(1,24) < 1$).

Correlational analyses

Correlations were computed between the quality of the interpreters narratives in the Dutch-to-SLN and SLN-to-Dutch interpreting directions (Experiment 1) and the NSC* scores in the three conditions of the memory span tasks in each language (see Table 4). As in Experiment 1, the scores on both measures obtained by van Dijk et al. (2011) were combined into a third measure.

As shown in Table 4, the memory span scores for spoken words in the OAS condition were significantly correlated with the quality of interpreted narratives in both interpreting directions. Furthermore, the memory span scores for sign in the OAS condition were significantly correlated in the Dutch-to-SLN interpreting direction, and showed a trend towards significance in the SLN-to-Dutch direction ($p = .081$). Furthermore, no significant correlations (or trends towards significance) were found between the quality of interpreted narratives and the memory span scores for signs and words in the Control and MAS conditions.

Discussion

Experiment 2 revealed an intriguing pattern of results. First, oral articulatory suppression effects were found for

Table 4. *The correlations between a combined measure of quality of the interpreted stories (van Dijk et al., 2011) and the serial order memory NSC* scores in the control, oral articulatory suppression (OAS) and manual articulatory suppression (MAS) conditions in the STM tasks with signs and words (* = p < .1, ** = p < .05). NSC* refers to the Number of Series Correctly recalled * Series Length.*

STM task		Interpreting condition	
Language	Condition	SLN to Dutch	Dutch to SLN
SLN	Control	.10	.24
	OAS	.36 *	.49 **
	MAS	.15	.21
Dutch	Control	.15	.13
	OAS	.40 **	.49 **
	MAS	.10	.26

both SLN signs and Dutch words, whereas no manual articulatory suppression effects were found for SLN signs and Dutch words. The finding that even in the memory span task with SLN signs an oral articulatory suppression effect was observed, suggests that participants used a speech-based code to store the signs and spoken words in memory in the span tasks we administered. This finding is consistent with Hall and Bavelier's (2010) suggestion that bimodal bilinguals may use more than one code to store and retain information in short-term memory.

One possibility is that participants activate the translation equivalents of the SLN signs in spoken Dutch, and use the phonological representation of the translation equivalents to store and retain information in memory. The use of translation equivalents as stimulus materials in the memory span tasks for signs and words may have encouraged participants to recode signs into words.

Alternatively, and perhaps more likely, participants may have used a speech-based component of the signs to store and retain the signs in memory. In Sign Language of the Netherlands, a considerable proportion of signs consist of a mouthing, a component which is intimately related to the translation equivalent in spoken Dutch, which can be produced vocally and subvocally during the production of signs in a reduced or complete form (Schermer, 1990, 2001). Schermer (2001) conducted one of the few studies on the role of mouthings and mouth gestures (oral components not related to spoken Dutch) in SLN with Deaf native signers. She found that Deaf native signers spontaneously produced mouthings in over 50% of the signs. Although some of the mouthings clearly had a function in disambiguating the manual part of the sign or complemented the meaning of the manual part of the sign, the vast majority of the mouthings that were produced had no specific function.

Furthermore, an analysis of the production errors of (deaf and hearing) British Sign Language (BSL) has revealed that there is dissociation between the mouthing errors and manual errors of BSL signers (Vinson, Thompson, Skinner, Fox & Vigliocco, 2010). This finding suggests that the mouthing component and the manual components of signs are represented independently in the mental lexicon, and are independently accessed during sign language production. All of the signs in Experiment 2 (e.g., common nouns) were presented to the interpreters with a mouthing component. Sign language interpreters may have relied upon this speech-based phonological component of the sign in this type of experimental setting simply because they are more skilled (have more experience) in using a speech code to store and retain information in short-term memory.

Second, no main effect of language was observed, although additional analyses revealed that the span scores for words under oral articulatory suppression were larger than the span scores for signs under oral articulatory suppression.[5] To illustrate, in the control condition the average scores for signs and words were, respectively 38.1 and 37.9 for the NSC* scores (Number of Series Correctly recalled * Series Length) and 6.24 and 6.21 for the MLC (Maximum List length Correctly recalled) scores. In most of the studies on memory spans for words and signs (see Bavelier, Newport, Hall, Suppalla & Boutla, 2006; Boutla et al., 2004), however, the span scores for spoken words (average MLC scores 6−7) were found to be significantly higher than the span scores for signs (average MLC scores 4−5). To account for the higher span scores for spoken words in comparison to signs, Wilson and Emmorey (1997a) have argued that differences in articulation time underlie the difference in the short-term memory spans for signs and words. As signs generally take longer to articulate than spoken words (Bellugi & Fischer, 1972), the storage and retention of signs in short-term memory will be more difficult than the storage and retention of spoken words. Alternatively, working memory for speech may be more efficient in retaining serial order information in comparison to working memory for signs (e.g., Boutla, et al., 2004; Hall & Bavelier, 2010). We will return to this issue in more detail in the general discussion.

[5] The retention of words and signs under oral articulatory suppression will strongly reduce the involvement of active rehearsal (Unsworth & Engle, 2007) and will increase the impact of long-term language specific knowledge (Padilla et al., 2005). The observation of larger span scores for spoken words under oral articulatory suppression in comparison to the span scores for signs under oral articulatory suppression can be accounted for if it is assumed that spoken Dutch is the more dominant language of the interpreters (and that the subjective frequencies of Dutch spoken words are higher than the subjective frequencies of their SLN translation equivalents). In this scenario, Dutch words are more easily available from long-term memory than SLN signs, which will facilitate the constructing of episodes in the episodic buffer.

Third, and most importantly, significant positive correlations (or a positive correlation showing a trend towards significance) between the scores in the span tasks for signs and words under oral articulatory suppression and the quality of the interpreted narratives were observed in both interpreting directions. Thus, in the present study the memory span scores for words and signs under oral articulatory suppression were positively correlated with their interpreting performance regardless of the direction of interpretation and regardless of the language of the memory span task: higher span scores for words and signs under oral articulatory suppression were associated with better performance in both interpreting directions. The fact that we obtained significant correlations with the span scores not only for words, but also for signs under articulatory suppression is not so difficult to account for if we consider that participants were likely to have used a speech-based code (i.e., the mouthing component or the Dutch translation equivalent of the sign) to store and retain information in memory.

The correlation between the span scores for words and signs under oral articulatory suppression and the quality of interpretations in both interpreting directions is consistent with Padilla et al.'s (2005) and Christoffels' (2006) suggestion that the ability to bind information in episodic memory is important in producing high-quality interpretations. The simultaneity of language comprehension and production during interpreting will make it more difficult to retain information from the source language in short-term memory (Christoffels, 2006). An enhanced ability to bind information in episodic memory will reduce interpreters' reliance on short-term memory for the retention of information in the source language during interpreting, and will positively affect the quality of their interpretations.

The relative insensitivity for oral articulatory suppression in word and sign span tasks reflects the interpreters' ability to bind information in episodic memory. Participants will construct episodes in the episodic buffer during word span tasks (under articulatory suppression) and use these episodes during the recall task. The use of a relatively large set of words and signs ($n = 45$) in the present memory span experiments may also have increased participants' reliance on combining information from multiple sources in episodic memory (Cowan, 2001). Efficiency in constructing episodes in the episodic buffer will reduce reliance on short-term memory, and will result in a relative insensitivity to oral articulatory suppression in memory span tasks. In other words, the ability to bind information from different sources in episodic memory is likely to underlie the positive relationship between the quality of interpretations and the memory span scores for signs and words under oral articulatory suppression.

General discussion

In two experiments, the relation between the working memory skills of experienced sign language interpreters and the quality of their interpretations in two interpreting directions (SLN to Dutch and Dutch to SLN) was explored. In Experiment 1, one 3-back task with SLN signs and one 3-back task with spoken Dutch words were administered to a group of twenty-five experienced sign language interpreters. The results showed that the sign language interpreters' overall performance in SLN was as good as in Dutch, but analyses of the accuracy scores on the lure trials revealed that performance in Dutch was significantly better than in SLN. Furthermore, differences in the 3-back accuracy scores were not correlated with the quality of their interpreted narratives. We speculate that *n*-back tasks may not be the most ideal tasks to measure working memory skills (Conway et al., 2005; Kane et al., 2007).

In Experiment 2, experienced sign language interpreters conducted a series of memory span tasks with signs and spoken words under three different conditions: (1) oral articulatory suppression (saying "da-da-da"), (2) manual articulatory suppression (making an alternating two-handed movement forwards combined with a hand-internal movement) and (3) control condition (just watching / listening). Interestingly, the results revealed that the memory span scores in the control conditions in SLN and Dutch were equally large. This finding is inconsistent with studies that have consistently shown that the memory span scores for signs (for hearing and deaf signers) are lower in comparison to the memory span scores for spoken words (for a review, see Hall & Bavelier, 2010). The difference between the memory span scores for signs and words has often been interpreted in terms of a difference in the articulation time of signs and words (Wilson & Emmorey, 1997a) or in terms of working memory for speech to be more efficient for serial order recall than working memory for signs (Bavelier, et al., 2006; Boutla et al., 2004).

One possible explanation for the lack of this language effect is that the interpreters in our study used the mouthing component of the signs or the Dutch translation equivalents of the signs to store and retain (and update) the signs in short-term memory. This explanation was supported by our finding that the memory span scores for signs and words were affected by oral articulatory suppression, and not by manual articulatory suppression. If the use of a speech-based code indeed leads to better serial recall of signs, the next question is why hearing ASL−English bilinguals in a previous study on the memory spans for signs and words (e.g., Boutla et al., 2004) did not use a speech-based code to store and retain the ASL stimuli. There are several possible explanations. First, Boutla et al. (2004) used a small set of phonological simple stimuli (finger-spelling, digits) which did not differ in articulation time, whereas we selected a larger set of phonologically more complicated set of Dutch−SLN translation equivalents (nouns) which presumably were longer to articulate in SLN than in Dutch. The use of a large set of phonologically more complicated stimuli may have stimulated participants to use a speech-based code to store and retain the signs in short-term memory. Further research is obviously needed to clarify this issue. Nevertheless, the data obtained in the present study provide a new piece of the puzzle in the ongoing debate on short-term memory spans for signs and words in deaf and hearing people (Bavelier et al., 2006; Boutla et al., 2004; Wilson & Emmorey, 2006a, 2006b).

Finally, we found that the memory span scores of signs and spoken words under oral articulatory suppression were correlated with the quality of their interpretations in both interpreting directions. We argue that an enhanced ability to bind information from multiple sources in episodic memory underlies this relationship. The simultaneity of language comprehension and production during interpreting makes it difficult for interpreters to retain information from the source language, as it hinders the retention of information in short-term memory through phonological processes such as active rehearsal (Christoffels, 2006). An enhanced ability to bind information in episodic memory will reduce interpreters' reliance on short-term memory for the retention of information in the source language. As a consequence, this enhanced ability to effectively create episodes facilitates the production of high-quality interpretations. The sensitivity to articulatory suppression in memory span tasks reflects the ability to effectively create episodes in episodic memory.

References

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

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Neuroscience, 4*, 829−839.

Bajo, M. T., Padilla, F., & Padilla, M. (2000). Comprehension processes in simultaneous interpreting. In A. Chesterman, N. GallardoSan Salvador & Y. Gambier (eds.), *Translation in context*, pp. 127−142. Amsterdam/Philadelphia: John Benjamins.

Barik, H. C. (1975). Simultaneous interpretation: Qualitative and linguistic data. *Language and Speech, 18*, 272−297.

Bavelier, D., Newport, E. L., Hall, M. L., Supalla, T., & Boutla, M. (2006). Persistent difference in short-term memory span between sign and speech: Implications for cross-linguistic comparisons. *Psychological Science, 17*, 1090–1092.

Bellugi, U., & Fischer, S. (1972). A comparison of sign language and spoken language. *Cognition, 1*, 173−200.

Billies, P., Buchkoski, D., Kolvitz, M., Sanderson, G., & Walter, J. (2003). *Postsecondary Education Programs Network: Needs assessment, 2002* (Report of OSERS grant H078A60004). Rochester, NY: National Technical Institute for the Deaf, Rochester Institute of Technology, Northeast Technical Assistance Center.

Boutla, M., Supalla, T., Newport, E. L., & Bavelier, D. (2004). Short-term memory span: Insights from sign language. *Nature Neuroscience, 7*, 997–1002.

Christoffels, I. K. (2006). Listening while talking: The retention of prose under oral articulatory suppression in relation to simultaneous interpreting. *European Journal of Cognitive Psychology, 18*, 206−220.

Christoffels, I. K., & de Groot, A. M. B. (2005). Simultaneous interpreting: A cognitive perspective. In J. F. Kroll & A. M. B. de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 454–479. New York: Oxford University Press.

Christoffels, I. K., 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.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769–786.

Conway, A. R. A., Kane, M. J. & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences, 7*, 547−552.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–114.

Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging, 4*, 500–503.

Gerver, D. (1976). Empirical studies of simultaneous interpretation: A review and a model. In R. W. Brislin (ed.), *Translation: Applications and research,* pp. 165−207. New York: Gardner Press.

Gile, D. (1997). Conference interpreting as a cognitive management problem. In J. Danks, G. M. Shreve, S. B. Fountain & M. K. McBeath (eds.), *Cognitive processes in translation and interpreting*, pp. 196−214. Thousand Oaks, CA: Sage.

Hall, M., & Bavelier, D. (2010). Working memory, deafness and sign language. In M. Marschark and P. E. Spencer (eds.), *The handbook of deaf studies, language and education*, pp. 458−472. New York: Oxford University Press.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the *N*-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 615–622.

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

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1118–1133.

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

Miller, K. M., Price, C. C., Okun, M. S., Montijo, H., & Bowers, D. (2009). Is the *N*-back task a valid neuropsychological measure for assessing working memory? *Archives of Clinical Neuropsychology, 24*, 711−717.

Moser, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. In D. Gerver & H. W. Sinaiko (eds.), *Language, communication and interpretation,* pp. 353−368. New York/London: Plenum Press.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*, 46–59.

Padilla, F., Bajo, M. T., & Macizo, P. (2005). Oral articulatory suppression in language interpretation: Working memory capacity, dual tasking and word knowledge. *Bilingualism: Language and Cognition, 8*, 207−219.

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: University of Turku Press.

Power, D., & Hyde, M. (2002). The characteristics and extent of participation of deaf and hard-of-hearing students in regular classes in Australian schools. *Journal of Deaf Studies and Deaf Education, 7*, 302−311.

Powers, S. (2002). From concepts to practice in deaf education. *Journal of Deaf Studies and Deaf Education, 7*, 230−243.

Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research, 31*, 573–598.

Schermer, G. M. (1991). *In search of a language. Influences from spoken Dutch on Sign Language of the Netherlands*. Amsterdam: Universiteit van Amsterdam.

Schermer, G. M. (2001). The role of mouthing in Sign Language of the Netherlands: Some implications for the production of sign language dictionaries. In P. Boyes Braem & R. Sutton-Spence (eds.), *The Hands are the head of the mouth: The mouth as articulator in sign languages*, pp. 273−284. Hamburg: Signum Press.

Schick, B., Williams, K., & Kupermintz, H. (2006). Look who's being left behind: Educational interpreters and access to education for deaf and hard-of-hearing students. *Journal of Deaf Studies and Deaf Education, 11*, 3–20.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1089−1096.

Timarova, S. (2008). Working memory and simultaneous interpreting. In P. Boulogne (ed.), *Translation and its others. Selected papers of the CETRA Research Seminar in Translation Studies 2007*, pp. 1−28. Leuven: KU Leuven.

Tzou, Y.-Z. (2008). The roles of working memory, language proficiency, and training in simultaneous interpretation performance: Evidence from Chinese−English bilinguals. Unpublished doctoral dissertation, Texas A & M University, United States.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*, 1038–1066.

van Dijk, R., Boers, E., Christoffels, I. K., & Hermans, D. (2011). Directionality effects in simultaneous language interpreting: The case of sign language interpreters in the Netherlands. *American Annals of the Deaf, 156 (1)*, 47−55.

Vinson, D. P., Thompson, R. L., Vigliocco, G., Skinner, R., Fox, N., & Vigliocco, G. (2010). The hands and mouth do not always slip together in British Sign Language: Dissociating articulatory channels in the lexicon. *Psychological Science, 21*, 1158−1167.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation.

Wilson, M., & Emmorey, K. (1997a). A visuospatial "phonological loop" in working memory: Evidence from American Sign Language. *Memory & Cognition, 25*, 313−320.

Wilson, M., & Emmorey, K. (1997b). Working memory for sign language: A window into the architecture of the working memory system. *Journal of Deaf Studies and Deaf Education, 2*, 121−130.

Wilson, M., & Emmorey, K. (2000) When does modality matter? Evidence from ASL on the nature of working memory. In K. Emmorey & H. Lane (eds.), *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima,* pp. 135−142. Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, M., & Emmorey, K. (2006a). Comparing sign language and speech reveals a universal limit on short-term memory capacity. *Psychological Science, 17*, 682–683.

Wilson, M., & Emmorey, K. (2006b). No difference in short-term memory span between sign and speech. *Psychological Science, 17*, 1093−1094.