Cognitive studies in simultaneous interpreting
Christoffels, I.K.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Comparing Retention following Simultaneous Interpreting and other Tasks involving Concurrent Articulation

7.1 Introduction

Text recall is consistently better when listening to a text than when simultaneously interpreting it (Darò & Fabbro, 1994; Gerver, 1974b; Isham, 1994; Lambert, 1988). Since simultaneous interpreting (SI) involves concurrent comprehension and production, a likely reason for poorer recall of interpreted material is that in SI speech production interferes with retention of the input (Christoffels & De Groot, in press, Chapter 2; Darò & Fabbro, 1994; Isham, 2000). SI is in some respects not unlike the situation created during articulatory suppression (AS) in memory research. In AS subvocal rehearsal is blocked by having participants utter irrelevant sounds continuously, such as ‘the, the, the, …’ during the presentation of words or digits with the effect of poorer recall in comparison to listening to these items without concurrent articulation (Baddeley, Lewis, & Vallar, 1984).

According to the working memory framework of Baddeley and colleagues (Baddeley, 1986, 2000, 2003; Baddeley & Logie, 1999; Gathercole & Baddeley, 1993) the effect of AS is located in the phonological loop component. This component is specialized in holding speech-based information. It is one of the two subsidiary systems that are controlled by the central executive, which is a limited capacity attentional system. The executive controls processes in working memory, including encoding and retrieval strategies, and the manipulation of material held in these systems. The phonological loop consists of a phonological store, which retains verbal traces for a few seconds, and an articulatory rehearsal system, which serves to refresh the verbal material held in the store. AS is assumed to block this rehearsal system, thus leading to poorer recall. A fourth component was recently introduced to the model: The episodic buffer. This is a limited-capacity store capable of integrating information from different sources, including long-term memory, in a multi-dimensional episodic structure (Baddeley, 2000, 2003). The buffer can maintain integrated information for periods and in amounts that exceed the phonological loop capacity, and is therefore important for the retention of prose.

By systematically comparing recall following SI to recall following other tasks that involve concurrent articulation we may gain some insight in the amount of central
processing involved in SI. An example of such a task is shadowing, in which spoken material is immediately literally repeated. Like SI it involves simultaneous comprehension and production (e.g., Marslen-Wilson, 1973) but in contrast to SI only one language is involved. Recall following interpreting has been compared to recall following other tasks previously, but these comparisons have produced equivocal results. Gerver (1974b) compared recall following interpreting, shadowing and listening of French texts in interpreter trainees. Recall was best following listening, intermediate following interpreting, and worst following shadowing. Similarly, Lambert (1988) tested a group consisting of trainees and professional interpreters. She found that free recall of the texts was similar following listening, simultaneous interpreting, and consecutive interpreting; again recall was poorest following shadowing. Also semantic recognition recall was poorer following shadowing, and for recognition recall also a difference between listening and simultaneous interpreting was obtained, with better recall following listening. Both authors suggested that the difference in recall between shadowing and interpreting can be explained by the fact that the latter task involves compulsory ‘deep’ or complex processing of the input (see also Christoffels & De Groot, 2003, Chapter 3). Interpreting requires a transformation from one language to another, whereas the input-output transformation in shadowing is far less complex.

Dará and Fabbro (1994) used a different approach and obtained different results. They measured digit span in the dominant (L1) and the second language (L2) in four conditions: Advanced interpreting students silently listened to the digits, performed AS while listening to the digits, shadowed, and interpreted the digits. Noteworthy is that they did not obtain any effect of what language the digits were presented in. Their main finding was that digit span was smaller in the interpreting condition than in any of the remaining conditions. Furthermore, digit span was smaller in the AS condition than in the silently listening conditions. Surprisingly, in the shadowing condition the digit span was not smaller than in the listening condition or any other condition. However, since recall of separate digits was measured rather than recall of ongoing text, and since shadowing digits seems to involve the verbal repetition of digits that are presented one second apart, recall under these circumstances may actually not have suffered much from concurrent articulation.

In a study conducted recently we compared sentence recall following SI and shadowing. The results were not consistent with any of the above mentioned studies, in the sense that we did not obtain any differences in recall across tasks (Christoffels & De Groot, 2003, Chapter 3). In contrast, in another study (Christoffels, 2003b, Chapter 6) we did not compare shadowing and interpreting directly but showed that type of articulatory suppression, i.e., uttering meaningless sounds or repeating three words, affected the amount of recall of short stories. This suggested that the type of articulation that is required from the participant influences recall.

One reason for the mixed results of the above-mentioned studies may be that different types of recall were measured. As a consequence, the relative importance of short- and long-term memory may have differed, implicating differences in the roles of the phonological
Comparing retention following simultaneous interpreting

loop and the episodic store. For example, Darò and Fabbro (1994) measured relatively immediate and verbatim recall, whereas Gerver (1974b) measured recall by questions following the presentation of a complete discourse.

It is clear that there is some discrepancy in the results described above. The first goal of the present study was to compare recall following listening, shadowing, and interpreting in a study that is specifically targeted at measuring recall. The question was whether any differences between shadowing and interpreting would emerge, or whether we would replicate the null-effect observed in Chapter 3 (Christoffels & De Groot, 2003). The second goal was to obtain an indication of the impact of the type of recall that is measured. We used two different ways of measuring story recall, either focusing on the semantic content (semantic recall) or on the exact wording of the stories (form recall).

In addition, we examined recall in two additional conditions that involve concurrent articulation. We included a traditional articulatory suppression condition and a random letter generation condition. Random letter generation is considered to be a task that involves high demands on the limited capacity of central executive processing, and has been highly successful in disrupting executive behavior across a range of different tasks (Baddeley, Emslie, Kolodny, & Duncan, 1998). Recall in the random letter generation condition is therefore expected to be lower than in all other conditions. In contrast, AS may be regarded as a condition that specifically disrupts the articulatory rehearsal process but imposes minimal demands on executive processing (Baddeley, Gathercole et al., 1998). AS and random letter generation may be regarded as benchmarks against which to compare recall in the shadowing and interpreting conditions. In shadowing and interpreting, the demands on working memory are likely to be by higher than in AS because of the complexity of the processing involved. Note, however, that what is being uttered in shadowing and interpreting is not irrelevant but directly related to the input instead.

Finally, we assessed the role of language. Previous studies looking at bilingual digit span, word span, or complex spans often found that bilinguals' performance is better when participants are tested in the first language (L1) rather than in the second (L2) (e.g., Chincotta & Underwood, 1998a; Christoffels et al., 2003, Chapter 5; Thorn & Gathercole, 2001), although this may depend on type of memory test (Harrington & Sawyer, 1992), and language proficiency (Service et al., 2002). The question was whether the effect of language extends to retaining short stories as well. Since the participants in this study are fluent in the L2 but certainly not balanced bilinguals we may expect recall to be better when they are tested in the native language than in the L2. Especially in form recall differences may be obtained between the L1 and L2 since recall of the exact wording is likely to depend more on the language that the story was delivered in than semantic recall does.

Summarizing, we compared immediate recall following listening to short stories in two languages under five different conditions: Participants just listened to the stories, uttered irrelevant syllables while listening, shadowed the stories, interpreted them simultaneously, or generated random letters while listening.
Chapter 7

7.2 Method

7.2.1 Participants

Thirty native Dutch (L1) university students who spoke English as a second language (L2) participated in this study. The age of the participants varied between 18 and 33 and was on average 21.2 years. The participants rated their passive and active knowledge of English on a scale from 1 to 10, on which 1 indicated no knowledge of the language and 10 indicated native knowledge. The average rating for active knowledge of English was 7.3 and the average rating for passive knowledge of English was 7.5. Participants were either paid or received course credit for participation.

7.2.2 Stimuli

Stories

Fifteen stories were constructed especially for this experiment. All stories had a simple structure and described a single event that happened to one or two main characters at a certain location. The stories were first written in Dutch, and then all translated into English. A native English speaker checked the English translations. The stories consisted of 15 to 16 sentences each (mean 15.8). The number of words in each story varied between 130 and 155 words, and averaged 140 words. The stories were recorded by a fluent Dutch-English bilingual. The duration of the recordings was about 50 seconds. Ten stories were used for the experimental conditions; the remaining five stories were used for practice purposes. In Appendix 8 an example of the stories is presented in Dutch and English.

7.2.3 Recall

Immediate recall of the stories was measured in two ways. The semantic recall measure was designed to reflect how well the gist of the story was retained, whereas the form recall measure was designed to give an indication of how well the exact wording was retained. All recall was measured from the middle 12 sentences of each story. The first and last sentences were excluded from scoring because in both the shadowing and interpreting condition there is a short delay before participants start to articulate their output, so no concurrent articulation takes place during these moments and effects of recency and primacy on recall were avoided.

In semantic recall each of the experimental stories consisted on average of 24 ‘idea units’ (between 22 and 29). Idea units consisted of propositions, determinants of place and time (e.g., ‘always’ and ‘on the market’), and causal relationships (e.g., ‘because’). For each idea unit it was determined whether it was present in the written free recall data of the
participants and the percentage of correct recall was calculated. The semantic content of the unit was considered important; there was no exact written reproduction necessary for a unit to be considered correct.

In form recall 50 items, consisting of single words, were selected from each story. Selection took place by systematically excluding words from each story along the following guidelines, until 50 items were left. Articles and very common verbs were completely excluded from the list, since they are likely to be produced by chance. Most items that were repeated in the story were only counted once. For each of the selected items it was determined whether it was correctly recalled and the percentage of correct recall was calculated. Items were only considered correct if they were reproduced literally.

7.2.4 Design and procedure

Two factors were manipulated in this study: listening condition (5 levels), and language (2 levels). In the control condition participants were instructed to just listen to the stories. In the articulatory suppression condition (AS) participants were required to utter 'de, de, de' continuously. In the shadowing condition participants were asked to literally repeat what they heard. In the simultaneous interpreting condition (SI) participants were instructed to immediately translate the meaning of the story they heard into the other language. Finally, in the random letter generation condition (RLG) the participants were required to generate a random sequence of the letters A, B, C, D, E, and F.

Listening conditions were presented blocked by language, in Dutch (L1) and English (L2). Half of the participants started with the Dutch condition, the other half with the English condition. The English and the Dutch version of each story were used equally often across the different listening conditions. Furthermore, each listening condition was presented first, second, third, fourth, and fifth equally often.

Before the start of each condition, the participants received instructions and practiced the condition with one of the practice stories. Stories were presented by computer over headphones at a volume that was comfortable to the participants. Each condition started with an alert sound. Three seconds later the story was presented. In the AS and RLG conditions this alert signaled the participants to start articulating. Immediately after the story was finished, an alert sounded to indicate the end of the story. Five seconds later the last alert sounded to indicate that the participants could stop articulating in the AS and RLG conditions and start recalling the story in all conditions. They were instructed to write down as much as they could remember from the stories and to try to use exact wording if possible. Time for recall was no longer than four minutes per condition, but this provided usually enough time for the participants to finish written recall. After a short break the next condition was presented. At the end of the session the participants filled in a short questionnaire concerning their language knowledge. Each session took about 90 minutes.
7.3 Results

7.3.1 Semantic recall

A repeated-measures ANOVA with listening condition (control, AS, shadowing, SI, RLG) and language (Dutch, English) as within-subject factors, revealed a main effect of listening condition, $F(4, 26) = 47.21, p < .001, \eta^2 = .88$. The main effect of language was not significant, nor was the interaction between listening condition and language, $p$’s > .10. The percentages semantic recall are presented in Table 1 and graphically in Figure 1.

Table 1. Mean semantic recall (% recall) and standard deviations (SD) per listening condition per language.

<table>
<thead>
<tr>
<th>Listening condition</th>
<th>CL % recall</th>
<th>AS % recall</th>
<th>SH % recall</th>
<th>SI % recall</th>
<th>RLG % recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>60</td>
<td>52</td>
<td>51</td>
<td>51</td>
<td>29</td>
</tr>
<tr>
<td>SD</td>
<td>12</td>
<td>18</td>
<td>19</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>English</td>
<td>60</td>
<td>44</td>
<td>50</td>
<td>56</td>
<td>26</td>
</tr>
<tr>
<td>% recall</td>
<td>17</td>
<td>19</td>
<td>16</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Mean</td>
<td>60</td>
<td>48</td>
<td>51</td>
<td>54</td>
<td>28</td>
</tr>
</tbody>
</table>

Note. CL = control; AS = articulatory suppression; SH = shadowing; SI = simultaneous interpreting; RLG = random letter generation.

For subsequent analyses, the data were collapsed across the two levels of the factor language. Simple effects analyses confirmed, as expected, that recall in the control condition was better than in the concurrent articulation conditions, all $p$’s < 0.01. We therefore replicated the finding of previous studies of poorer recall following interpreting a text than following listening to it.

We obtained no significant difference between SI and shadowing. Unlike Gerver (1974b), who found worse performance in shadowing, and Darò and Fabbro (1994), who found better performance, but in accordance with our own previous results (Christoffels & De Groot, 2003, Chapter 3), we could not differentiate SI from shadowing on the basis of recall performance, $F(1, 29) = 1.54, p > .10$.

Interestingly, recall in the AS condition was not better than in the shadowing condition, $F(1, 29) < 1, p > .10$, and poorer rather than better following SI, $F(1, 29) = 5.34, p = .028, \eta^2 = .16$, even though the shadowing and interpreting conditions may be considered to involve more complex processing.

Poorest recall was obtained in the RLG condition. In this condition, recall was significantly poorer than in all other conditions, all $p$’s < .001.
7.3.2 Form recall

A repeated-measures ANOVA with listening condition (control, AS, shadowing, SI, RLG) and language (Dutch, English) as within-subject factors was performed. This analysis showed a main effect of listening condition, $F(4, 26) = 44.71$, $p < .001$, $\eta^2 = .87$. The main effect of language was not significant, $F(1, 29) < 1$, $p > .10$, but the interaction between listening condition and language was, $F(4, 26) = 4.02$, $p = .011$, $\eta^2 = .38$. The percentages form recall are presented in Table 2 and graphically in Figure 2.

Interestingly, when comparing the simple effects of language per listening condition only in SI the language effect reached significance, with better performance for SI in the English condition (i.e., translation into Dutch, the L1), $F(1, 29) = 9.31$, $p = .005$, $\eta^2 = .24$.

However, simple main effects analyses revealed that some of the differences between listening conditions were language-dependent. As expected, recall was generally better in the control condition than in the other conditions, but this pattern did not hold for both languages in all conditions. Recall in the control condition was significantly better than in the AS condition for both Dutch and English, $F(1, 29) = 9.45$, $p = .005$, $\eta^2 = .25$, and $F(1, 29) = 21.08$, $p < .001$, $\eta^2 = .42$, respectively. Likewise, recall was better in the control condition than in the RLG condition, for Dutch $F(1, 29) = 86.18$, $p < .001$, $\eta^2 = .75$, and English, $F(1, 29) = 112.79$, $p < .001$, $\eta^2 = .80$. Although recall in the control condition was also significantly better than recall in the shadowing condition for English, $F(1, 29) = 4.37$, $p$
= .045, \( \eta^2 = .13 \), this was not the case for Dutch, \( F(1, 29) < 1, p > .010 \). SI showed the opposite pattern from shadowing: Recall in the control condition was better than in the SI condition for Dutch, \( F(1, 29) = 7.48, p = .011, \eta^2 = .21 \), but not for English, \( F(1, 29) < 1, p > .10 \). In other words, recall performance for the shadowing condition was more like the control condition when stimulus and recall were in Dutch, whereas for SI this was the case when stimulus and recall were in English (i.e., when SI took place from English into Dutch).

Also the difference between shadowing and interpreting depended on language. For English no significant difference was obtained, \( F(1, 29) = 2.33, p > .10, \eta^2 = .07 \), but for Dutch it was significant: Recall was better in shadowing than in SI, \( F(1, 29) = 4.82, p = .036, \eta^2 = .14 \). In other words, recall following SI was never better than following shadowing, and poorer than shadowing when the stimuli were presented in Dutch.

Table 2. Mean form recall (% recall) and standard deviations (SD) per listening condition per language.

<table>
<thead>
<tr>
<th>Listening condition</th>
<th>CL</th>
<th>AS</th>
<th>SH</th>
<th>SI</th>
<th>RLG</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dutch</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% recall</td>
<td>39</td>
<td>32</td>
<td>38</td>
<td>31</td>
<td>17</td>
</tr>
<tr>
<td>SD</td>
<td>14</td>
<td>14</td>
<td>17</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td><strong>English</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% recall</td>
<td>39</td>
<td>29</td>
<td>33</td>
<td>38</td>
<td>16</td>
</tr>
<tr>
<td>SD</td>
<td>13</td>
<td>14</td>
<td>11</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>39</td>
<td>31</td>
<td>36</td>
<td>34</td>
<td>17</td>
</tr>
</tbody>
</table>

*Note.* CL = control; AS = articulatory suppression; SH = shadowing; SI = simultaneous interpreting; RLG = random letter generation.

In comparison to AS, once more, the differences were language-dependent. In Dutch, recall in the shadowing condition was better than recall in the AS condition, \( F(1, 29) = 26.48, p < .001, \eta^2 = .48 \), and for English this difference was marginally significant in the same direction, \( F(1, 29) = 3.37, p = .077, \eta^2 = .10 \). Recall in the SI condition was also better than in the AS condition for English, \( F(1, 29) = 10.83, p = .003, \eta^2 = .27 \), but there was no significant difference for Dutch, \( F(1, 29) < 1, p > .10 \).

Finally, again poorest recall was obtained in the RLG condition for both Dutch and English, \( p < .001 \), in comparison to all other conditions.

### 7.4 Discussion

Across five listening conditions, as expected, recall was in general best in the control condition. Notable exceptions of this pattern were obtained for form recall in the English interpreting and Dutch shadowing conditions, which did not differ significantly from recall
in the control condition. In general, we replicated previous findings of poorer recall following SI and shadowing in comparison to ordinary listening, but this difference depended on the type of recall and language.

Of specific interest for the present study is the comparison between SI and shadowing. Replicating previous results (Christoffels & De Groot, 2003, Chapter 3), for semantic recall no differences between these two conditions were obtained. However, for form recall we found better recall in the shadowing condition than in the SI condition, but only when the stories were presented in Dutch, not when they were presented in English. The latter result contrasts with the results of Gerver (1974b), and Lambert (1988), who found poorest retention in shadowing.

The comparison of shadowing and SI to articulatory suppression showed that, against expectations, results were not better in the AS conditions. In semantic recall no difference between articulatory suppression and shadowing was obtained and recall following interpreting was even better than following articulatory suppression. In form recall, recall in the shadowing condition was (marginally) better than in the articulatory suppression condition in both languages, and recall in the SI condition was better for English.

The expectations concerning the random letter generation condition were confirmed. Poorest recall was obtained in this condition in comparison to all other conditions for both types of recall. The impact that this concurrent task has on memory was especially clear in form recall, for which the percentages of recall were extremely small.

7.4.1 Semantic versus form recall

Inspection of Figures 1 and 2 indicates that for semantic and form recall similar patterns of results were obtained, especially in English. Nevertheless, there are also some clear differences in results between the two types of recall. So, even when the recall scores were obtained using the same immediate free recall data, on the same stimuli, and at the same moment in time, results differ when the focus of the measurement is recall of gist or recall of exact wording. This complicates the comparability of recall across different studies, but can also explain why different patterns of results have been obtained in the past (Christoffels & De Groot, 2003, Chapter 3; Darò & Fabbro, 1994; Gerver, 1974b; Lambert, 1988).

On the other hand, it is important not to overemphasize the differences between the semantic and form recall results. They are mainly due to the fact that in form recall language interacts with listening condition. In fact, the results of semantic recall and form recall tend to be the same in at least one of the two languages. The similarity between the two measures is also indicated by the correlation between semantic recall and form recall in each condition (collapsed on the two languages). These correlations are significant and in general quite high ($r$ varied between .59 and .90). The impact of the different recall measures may become larger when the time lapse between stimulus and recall is larger, because in general when time passes less of the exact form of the input tends to be recalled.
7.4.2 Language

We obtained no effect of language in semantic recall, suggesting that retention of the gist of the story is relatively independent of language. Because in form recall exact wording is required, it was expected that in particular for this measure an effect of language would be obtained. Nevertheless, form recall in the control condition was the same in the two languages, suggesting that ordinary listening and retention of stories is not affected by language dominance in our group of participants. However, form recall tended to be better for L1 in the concurrent articulation conditions. As mentioned before, the only exception was retention in the SI condition, which was significantly better for L2 (i.e., when stories where presented in the L2 and SI took place from L2 into L1). When stories were presented in the L2, in the SI condition the actual articulation involved production the translation of the input in the L1, and vice versa when the stories are presented in the L1. Possibly, the language of output was more important in determining the amount of recall than the language of input. If production in the L2 consumes a relatively large amount of resources (in comparison to comprehension in the L2), this explains why in interpreting recall was better for English stories.

In conclusion, language did not show to be a very important factor in story recall, especially when recall of gist was concerned. In form recall it was a factor that mediates the effects of concurrent articulation on recall, i.e., when resources are stretched already.

To understand the differences between the articulatory suppression, shadowing, SI, and random letter generation conditions, we may take the working memory model into account (e.g., Baddeley & Logie, 1999). The components of interest are the phonological loop, since it maintains verbal information, the episodic buffer, since it maintains episodic structures that are important for the retention of stories, and finally, the central executive, because it is involved in the actual processing of information and in the transfer of information from the loop to the episodic buffer (Baddeley, 2000, 2003).

Articulatory suppression supposedly only prevents articulatory rehearsal in the phonological loop and should not consume many central resources (Baddeley, Gathercole et al., 1998). In the articulatory suppression condition less information is retained in the phonological loop since it cannot be rehearsed. Although the model is not very detailed on the transfer of information from the phonological loop to the episodic buffer as yet, we may assume that as a consequence of articulatory suppression the episodic structure must be built on the basis of less information than when no concurrent articulation takes place. Furthermore, we may assume that in this respect shadowing, SI, and random letter generation are the same.

Random letter generation not only prevents articulatory rehearsal because of concurrent articulation, but it also taxes the central executive. Therefore, few resources are available for creating an episodic structure of the content of the story (the episodic buffer heavily depends on central executive processing; Baddeley, 2000, 2003). This must result in
Comparing retention following simultaneous interpreting

a relatively poor representation of the stories in the episodic buffer. The effects of the demands posed by these conditions on recall were very clear in our data: recall following articulatory suppression was worse than following listening, but after random letter generation it was extremely poor.

The fact that we did not obtain better recall following articulatory suppression than following shadowing and interpreting may be considered surprising. After all, unlike articulatory suppression, that supposedly does not demand executive processing, interpreting and shadowing do involve the executive since the produced output in these two conditions is far more complex than in articulatory suppression. So why is recall in these conditions not more like recall in the random letter generation condition? There is a crucial difference between articulatory suppression and random letter generation on the one hand, and SI and shadowing on the other: Articulation in the former two conditions is irrelevant, whereas articulation in the latter two concerns the processing of information that is to be recalled. Possibly, the type of input processing involved in shadowing and interpreting facilitates episodic structure building, even though retention must be hampered by concurrent articulation and even though these conditions are likely to tax the executive. It seems that in interpreting and shadowing there is a trade-off of factors that either support or frustrate recall. The effects of concurrent articulation and resource consuming executive processing may obstruct recall, but the complex, or ‘deep’, level of processing that is involved in shadowing and SI probably facilitates recall. In fact, the input processing necessary for performing shadowing and SI may be precisely the type of processing that is necessary for building an episodic structure in the first place. This may result in a null effect (semantic recall) or even an advantage (form recall) for shadowing and interpreting as compared to articulatory suppression.

In line with the argument above we may expect that the amount of experience participants have in performing the task in a particular condition is likely to influence recall as well. It would be interesting to learn whether the pattern of recall would be different for professional interpreters since it is likely that for professionals SI is less demanding and therefore consumes less central resources. Then, more executive processing could be dedicated to the retention of information, leading to an advantage of SI over shadowing (as indeed Gerver, 1974b, found).

In conclusion, this study reveals some of the factors that may have caused the mixed results concerning recall following interpreting versus shadowing reported before (Christoffels & De Groot, 2003; Darò & Fabbro, 1994; Gerver, 1974b; Lambert, 1988, Chapter 3). The pattern of recall is influenced by the way recall is measured. For form recall but not for semantic recall also language influences the amount of input that is retained. Because recall following SI and shadowing tends to be similar or even better than recall following articulatory suppression, we may conclude that, rather than only interfering with recall, the processing necessary in shadowing and interpreting also assist recall.