Bilingualism: Language and Cognition 7 (3), 2004, 227–240 © 2004 Cambridge University Press DOI: 10.1017/S1366728904001609

Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing*

INGRID K. CHRISTOFFELS
ANNETTE M. B. DE GROOT
University of Amsterdam

Simultaneous interpreting is a complex task where the interpreter is routinely involved in comprehending, translating and producing language at the same time. This study assessed two components that are likely to be major sources of complexity in SI: The simultaneity of comprehension and production, and transformation of the input. Furthermore, within the transformation component, we tried to separate reformulation from language-switching. We compared repeating sentences (shadowing), reformulating sentences in the same language (paraphrasing), and translating sentences (interpreting) of auditorily presented sentences, in a simultaneous and a delayed condition. Output performance and ear–voice span suggest that both the simultaneity of comprehension and production and the transformation component affect performance but that especially the combination of these components results in a marked drop in performance. General lower recall following a simultaneous condition than after a delayed condition suggests that articulation of speech may interfere with memory in SI.

Introduction

In international political or corporate meetings simultaneous interpreting plays an important role in mediating communication. In daily life, we may have encountered simultaneous interpretations of live broadcasted statements or interviews on television news channels, such as CNN, and may have been intrigued by this capacity to verbally transform online a message from one language, the source language, into another language, the target language. In simultaneous interpreting (SI) it is required that interpreters both listen and speak at the same time. In this regard it contrasts with so called consecutive interpreting, where the interpreter alternates between listening and speaking and only starts to translate after the speaker has finished speaking.

SI is a cognitively demanding task because many processes need to take place at one and the same moment in time. The interpreter has to comprehend and store input segments in the source language, transform an earlier segment from source to target language, produce an even earlier segment in the target language, and cope with time pressure since SI is externally paced; the speaker, not the interpreter, determines the speaking rate (e.g. Gerver, 1976; Lambert, 1992; Padilla, Bajo, Cañas and Padilla, 1995). That SI is intrinsically demanding is illustrated by the fact that even experienced professional interpreters sometimes make up to several mistakes per minute, and that they usually take shifts of only 20 minutes maximum to prevent fatigue (Gile, 1997).

During SI, in contrast to normal language use, the interpreter must both comprehend another person's speech and produce their own speech at the same time. Although interpreters appear to take advantage of pauses in the input (Goldman-Eisler, 1972, 1980; Barik, 1973), a comparison of the input to their spoken output suggests that almost 70% of the time they are actually talking while processing the input (Chernov, 1994). The simultaneity of comprehension and production is perhaps the most salient characteristic of simultaneous interpreting, and it is likely to be one of the main reasons why it is such a cognitively demanding task.

Many people would agree that SI is one of the most complex language processing skills, and hence, characterizing this ability may be regarded as an important objective of psycholinguistic research. It is therefore surprising that experimental research on SI is sparse (see Christoffels and De Groot, in press, for a review). On the one hand, we may want to understand the processes involved in SI to extend our understanding of bilingual language performance in general. On the other hand, studying SI may refine our models on language

* The authors thank Bregje van Oel and Anniek Sturm-Faber for their invaluable research assistance, and Susanne Borgwaldt, Lourens Waldorp, Jos van Berkum and Diane Pecher for their helpful comments on earlier versions of this paper. This research was conducted while I. K. Christoffels was supported by a grant (575-21-011) from the Netherlands Organization for Scientific Research (NWO) foundation for Behavioral and Educational Sciences of this organization. Portions of this research were presented at the Twelfth Meeting of the European Society of Cognitive Psychology in Edinburgh, UK, in September 2001.

Address for correspondence
Ingrid Christoffels, Maastricht University, Faculty of Psychology, Department of Neurocognition, P.O. Box 616, 6200 MD Maastricht, The Netherlands
E-mail: *i.christoffels@psychology.unimaas.nl*

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

comprehension and language production, and on bilingualism in particular. SI may also provide insight into the relation between language comprehension and production.

In the present study, we tried to gain more insight into SI by investigating what components of SI are responsible for the demanding nature of the task. In the literature two task components are mentioned that could be the main sources of difficulty or complexity of SI (e.g. Anderson, 1994; De Groot, 1997; Frauenfelder and Schriefers, 1997; MacWhinney, 1997; Moser-Mercer, 1997). First, during SI one has to understand and produce speech simultaneously. Second, there is the act of the actual transformation of the message into another language. The first component, the simultaneity of comprehension and production, is especially salient in SI. It entails that two streams of speech have to be processed simultaneously and that attention has to be divided: One focus is on understanding new input, the other is on conceptualizing and producing an earlier part of the message (MacWhinney, 1997, in press). Regarding the second component, that is, transforming the message into another language, the act of reformulating the message may be distinguished from the fact that two different languages are simultaneously involved (Anderson, 1994; De Groot, 1997). Rather than translating each incoming word separately, interpreting presumably involves reformulation at a higher level (Goldman-Eisler, 1980; Schweda-Nicholson, 1987). Literal word-by-word translation alone would render an unintelligible interpretation, if only because languages often differ in word order. On top of reformulation, a switch has to be made from one language into another. The two language systems concerned have to be activated simultaneously, albeit presumably to different degrees. These two 'sub-components' of transformation, the requirement of reformulation and the necessity of a switch of language, may separately contribute to the processing and memory demands of SI.

Is one of the above task (sub-)components perhaps the 'bottleneck', the main source of complexity, in SI? Or is it the case that two or maybe all three components combined are responsible for the demanding nature of the task? We studied this question by comparing interpreting with tasks that differ from it on one or more of the three potential sources of difficulty in SI (De Groot, 1997, 2000). One task that may be especially suitable for the comparison with interpreting is the shadowing task. Shadowing involves the immediate verbatim repetition of what is heard. A less known task, the paraphrasing task, forms another interesting candidate for comparison with SI. In the context of the present study, this task requires that the basic meaning of a message is restated in the same language but into different words and/or in a different sentence construction (see Moser, 1978). Likewise, in interpreting, the message is restated, but in this task it is done into a language different from the source language.

What shadowing, paraphrasing and interpreting have in common is that they require the simultaneous comprehension and production of speech. However, for shadowing, there is no additional complexity of having to reformulate a message and only one language is involved. It could even be argued that shadowing is not comparable to interpreting at all because shadowing could conceivably be done by just repeating the phonetic form of the input. However, this concern seems to be unwarranted because it has been shown that shadowers do analyze the input up to the semantic level (Marslen-Wilson, 1973).

In contrast to shadowing, in both the paraphrasing and interpreting tasks reformulation is necessary, since in both tasks the meaning of the message has to be extracted and restated into different words, but only in interpreting do two languages have to be activated simultaneously. Hence, by comparing performance on these three tasks it may be possible to assess the role of the transformation component and to disentangle the two sub-components of transformation: reformulating a message and doing so in another language.

The paraphrasing task seems ideal to assess the possible differences in demands between the latter two sub-components since the task seems indeed very similar to interpreting. In both cases, one has to comprehend an input message and produce an output message with the same meaning but in a different wording. In the literature, paraphrasing is often assumed to be similar to interpreting and, indeed, it is referred to as 'unilingual interpreting' or 'intralanguage translating' (Malakoff and Hakuta, 1991; Anderson, 1994). The task is often used as an exercise or assessment in the training of interpreters (Moser-Mercer, 1994). Also, interpreters may find themselves accidentally 'translating' into the same language (Anderson, 1994; De Bot, 2000), suggesting that paraphrasing is not an unnatural task. Moreover, in a study into relative hemispheric lateralization Green, Sweda-Nicholson, Vaid, White and Steiner (1990) compared bilingual interpreting with monolingual paraphrasing on the assumption that both tasks are similar enough to warrant such a comparison.

However, when we subject the paraphrasing task to closer scrutiny, it becomes clear that there may also be differences between interpreting and paraphrasing other than output language. For example, the vocabulary demands in paraphrasing may be larger than in interpreting (Malakoff and Hakuta, 1991). Although the vocabulary demands ultimately depend on the type of source text, in general interpreting may only require a basic vocabulary in each language, whereas paraphrasing may require the use of synonyms and, therefore, a larger vocabulary. Furthermore, having to change the grammatical structure in paraphrasing may be more demanding than finding the most dominant grammatical equivalent in the output language in interpreting, even

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

though interpreting often involves changing the word order of a sentence as well. A final difference in demands is perhaps most critical. The paraphrasing task forbids the participant to repeat the original stimulus sentence verbatim. At the same time, the stimulus sentence expresses the message that must be restated adequately. Therefore, the paraphraser is at risk of delivering the message in exactly the same form as it was phrased in the input. Not only is the given sentence form a legitimate way of conveying the intended message, the participant may even be primed to this particular sentence form and the exact words used to express the message (cf., Hartsuiker and Westenberg, 2000, who found priming for word order in sentence production). It seems plausible that, in order to comply with the task requirements, the paraphraser is actively involved in suppressing the original stimulus form and must rigorously monitor her output to avoid literal repetition. These two aspects may be less important in interpreting, because due to the required switch of language in the output, there is less risk of literal repetition in that task. Possibly, because of these differences, the cognitive demands of paraphrasing may in some respects be higher than those of interpreting.

Even though the paraphrasing task may not be a completely satisfactory comparison to interpreting, we decided nevertheless to include this task in the present experiment, the reason being that the task has been described and used before as a monolingual counterpart of interpreting. It was an objective of this study to learn whether or not the paraphrasing is justifiably regarded as a monolingual version of simultaneous interpreting. Moreover, if the unique task characteristics of paraphrasing, discussed above, are not so important in determining task performance, the task makes an interesting comparison to interpreting and shadowing. As argued earlier, we can then distinguish the separate demands of reformulation and of the involvement of two languages.

Task comparisons

Some previous studies have compared interpreting and shadowing and found that shadowing performance was better than interpreting performance and that the ear–voice span (EVS), the time lag between input and corresponding output, was shorter in shadowing than in interpreting (Treisman, 1965; Gerver, 1974a, b). Furthermore, the deteriorating effect of increased information density of the input (Treisman, 1965) and the effect of white noise in the input (Gerver, 1974a) were larger in interpreting than in shadowing. Also pupil dilation, taken as a measure of processing load, was smaller in shadowing than interpreting (Hyönä, Tommola and Alaja, 1995). Finally, shadowing and interpreting were contrasted using positron emission tomography (PET). It was shown that some brain areas were selectively activated in interpreting. The brain areas concerned are associated with lexical retrieval, verbal working memory, and semantic processing (Rinne, Tommola, Laine, Krause, Schmidt, Kaasinen, Teräs, Sipilä and Sunnair, 2000). Taken together, these studies suggest that interpreting is a more demanding and more complex task than shadowing and that it is more sensitive to factors that increase task difficulty than shadowing.

To our knowledge Anderson (1994) reported the only study that compared all three tasks mentioned before (paraphrasing was referred to as English–English translation). According to two measures of output quality, shadowing performance was significantly better than both interpreting and paraphrasing performance. Performance in interpreting was poorer than in paraphrasing, but only according to one of the quality measures. Furthermore, the EVS was smaller in shadowing than in interpreting and paraphrasing, but there was no difference in EVS between the latter two tasks. In other words, although Anderson replicated the difference between shadowing and interpreting described before, the results are not conclusive with respect to the demands involved in using two language systems instead of just one.

A few studies assessed memory performance across various tasks and found that text recall was systematically better when participants listened to text than when they interpreted text (e.g. Gerver, 1974b; Darò and Fabbro, 1994; Isham, 1994). This suggests that task complexity interferes with memory. However, the patterns of results were less consistent when SI was compared with shadowing. Gerver (1974b) found that both comprehension and recall were best following listening and worst after shadowing, whereas interpreting performance was in between. Likewise, Lambert (1988) obtained best recall following listening and simultaneous or consecutive interpreting, and worst following shadowing.

Darò and Fabbro (1994) measured digit span while participants just listened to the digits, performed articulatory suppression (continuous articulation of irrelevant syllables) during listening, shadowed, and interpreted the digits. Their main finding was that digit span was smaller in the interpreting condition than in any of the remaining conditions. Furthermore, the digit span was not smaller when shadowing than when listening. The results of Darò and Fabbro (1994) versus those of Gerver (1974b) and Lambert (1988) therefore differ on whether interpreting or shadowing leads to better memory performance. These studies are, however, difficult to compare because Darò and Fabbro did not measure recall of interpreted text but of digits. Shadowing of digits involved verbal repetition of these digits, presented one second apart, a procedure that may actually support recall. Moreover, they measured immediate verbatim recall instead of recall after presenting the complete text.

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

The present study

In the present study, we compared all three tasks discussed before. Unlike any of the earlier studies, we manipulated the simultaneity of input and output by administering, in addition to the simultaneous condition, a delayed condition. When simply comparing simultaneous interpreting, paraphrasing, and interpreting it is not possible to assess the effect of simultaneity of comprehension and production. Yet, the simultaneity of comprehension and production is probably the most salient feature of simultaneous interpreting and is considered to be one of its main sources of difficulty. The effect of this feature has yet to be established experimentally. To simulate the simultaneous interpreting situation, participants were to translate sentences on-line, one sentence immediately following the other. The participants were also asked to paraphrase and shadow sentences. In a delayed condition the participants only responded – by translating, paraphrasing or shadowing – after each sentence was completely presented. Interpreting in this condition is similar to consecutive interpreting in actual professional practice.

Online task performance was measured in several ways, by a measure indicating quality of task performance, by the amount of output, and by response latency. In measuring the quality of task performance, we used a rating system that indicated how accurately participants performed the translating, paraphrasing, and shadowing tasks. Furthermore, in contrast to previous studies a complementary analysis of the amount of output was performed. The response latency in the simultaneous condition was measured by calculating the lag between input and output (EVS). Since it is not possible in the delayed condition to calculate an EVS, the time lag between the end of a stimulus sentence and the onset of the response was calculated to serve as a measure of response latency.

We expect performance – both output quality and latency – to be better in the delayed condition than in the simultaneous condition since the simultaneity of comprehension and production is likely to be a major source of difficulty in all tasks. Since we expect the transformation component also to be an important factor, shadowing performance is expected to be generally better than performance in the other two tasks. Differential results in the paraphrasing and interpreting tasks may not only depend on whether or not two languages are involved but also on the other differences between the two tasks mentioned earlier. If paraphrasing performance turns out to be better than interpreting performance, we may conclude that the distinction between reformulation (present in both paraphrasing and interpreting) and the involvement of two languages (only present in interpreting) is valid. In other words, it would indicate that the involvement of two languages is indeed an important additional source of difficulty in SI. However, if paraphrasing performance turns out to be similar to interpreting performance, this can be either because there are no added demands of the language switch on top of the demands for reformulation, or because the effect of the added demands of a language switch in interpreting and the effect of task demands that are specific to paraphrasing cancel out. Since these two possible reasons for similar paraphrasing and interpreting performance cannot be distinguished, we would then not be able to separate the reformulation and language switch sub-components of the transformation component. If paraphrasing performance turns out to be even worse than interpreting performance, we may suspect that paraphrasing differs fundamentally from interpreting in aspects other than the number of languages involved. In that case, the assumed similarity of paraphrasing and interpreting should be questioned.

In this study we sought to combine online and offline performance measures for, to our knowledge, the first time. In addition to the measures of online performance, we tested memory performance across the various conditions by administering a cued recall task. There were two reasons for measuring recall. First, as described earlier, previous studies provided mixed results on the question whether recall is better after interpreting than after shadowing. Second, in some of the earlier studies the dependent measures concerned measures of online performance, such as the ear–voice span or the amount of errors, whereas other studies used offline measures, such as comprehension questions.

Concerning recall, depending on the theoretical point of view one takes, either one of two contrary patterns of results may be expected. On the one hand one could argue that interpreting is a task that consumes a relatively large amount of limited available working memory resources, so that in comparison to less complex tasks such as shadowing, less resources are available for remembering the stimuli. This view predicts poorer recall in the interpreting condition than in the remaining two task conditions. In contrast, as would be argued from a levels-of-processing framework (Craik and Lockhart, 1972; Craik and Kester, 2000; Lockhart and Craik, 1990), due to the transformation of input in the more complex tasks this input is processed more elaborately. According to this view, interpreting and paraphrasing should lead to better recall than shadowing (see also Lambert, 1988). The simultaneity of speech production and comprehension may be an additional source of differences in recall performance between conditions, since phonological interference may cause reduced recall in all simultaneous conditions (see also Darò and Fabbro, 1994; Isham, 2000). It has been found that short-term memory for lists of words is disrupted when participants engage in

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

articulatory suppression (e.g. Baddeley, Lewis and Vallar, 1984). Consequently, any situation where the participant is involved simultaneously in listening to speech and in articulating speech may mimic an articulatory suppression condition. In the simultaneous condition, we would therefore expect decreased recall due to articulatory suppression of the subvocal rehearsal process.

To summarize, in this study shadowing, paraphrasing, and interpreting are compared in both a simultaneous and a delayed presentation condition. Different performance measures should give an indication of the relative importance of the simultaneity and translation components in SI performance and of the suitability of regarding paraphrasing as a monolingual analogue of SI. From previous studies and the different theoretical stances that may be taken, it does not seem warranted to predict how recall in the interpreting condition will compare to that in the remaining conditions. However, the alleged detrimental effects of phonological interference on recall make it likely that recall will be better in the delayed than in the simultaneous condition.

Method

Two factors were manipulated within-subjects, type of *task* (shadowing, paraphrasing, interpreting) and *simultaneity* (simultaneous, delayed). During shadowing the participants literally repeated the input sentence. In paraphrasing the participants were asked to reformulate the sentence in the language of input while retaining its meaning, by changing the word order and/or using synonyms. Both shadowing and paraphrasing were performed in Dutch. Finally, in the interpreting condition participants were to give a translation of an English sentence into Dutch, again without changing its meaning.[1] In the simultaneous condition, participants were instructed to start their response during the presentation of the stimulus sentences, whereas in the delayed condition participants responded after completion of the presentation of each sentence. The main reason for using sentences rather than larger units of discourse, was to control the relatively large memory load that would be involved in interpreting, shadowing, and paraphrasing larger text units in a delayed condition.

Participants

Twenty-four native speakers of Dutch (13 females and 11 males, age 18–27) participated in this study, either voluntarily or in return for payment. All participants were unbalanced bilinguals, with Dutch as their native language and English as their strongest foreign language. Despite being dominant in Dutch they were fluent in English. Since the age of twelve, they had been instructed in English for three to four hours a week. They were university students, who either studied English at University level, used English on a daily basis for their doctoral study in other fields, or who had spent a few months abroad speaking English. In a language questionnaire no participants reported any (former) language problems (e.g. stuttering). On a scale from 0 to 10, participants rated their active knowledge of English at 8.2 on average and their passive knowledge at 8.5. All participants signed an informed-consent form before participating in the experiment.

Materials

Sentences

Forty English and 80 Dutch sentences were used in this study. Their length varied between 11 and 17 syllables. Of the complete sentence set, six different sets of 20 sentences each were constructed. Two of these sets consisted of English sentences and four consisted of Dutch sentences. The sets were matched on sentence length. Average sentence length per set varied between 12.7 and 13.3 syllables. The sentences were unambiguous, and the English sentences were relatively easy to comprehend and translate to make sure that difficulties would not be due to misunderstanding of the English input. Twelve participants with a similar English proficiency level participated in a control study that showed that the sentences were indeed easy to comprehend and translate. All sentences were read out loud by a fluent Dutch–English female bilingual and recorded on computer. In the simultaneous condition, the sentences were presented in succession, with a two-second pause in between sentences. Including this pause, the presentation rate was on average 119 words per minute. In the delayed condition, the next sentence was only presented after the participants had finished their response to the current sentence. The

[1] We decided to keep language of production the same across all tasks. This opens up the small possibility that comprehension processes are not completely the same between tasks because they take place in different languages, which in turn may affect task performance. However, the impact of this possible difference is likely to be negligible. In a recent review, Kroll and Dussias (2004) noted that although indeed some differences between monolingual and bilingual parsing strategies have been found, the pattern is complex and sometimes bilinguals resemble monolinguals. Moreover, Dussias (2001) concluded that the main differences in parsing strategies are found during early stages of language acquisition, where L1 strategies are used on L2 sentences. Dutch–English participants have been studied extensively (e.g. Kroll and Stewart, 1994; Dijkstra, Van Jaarsveld and Ten Brinke, 1998; Van Hell and De Groot, 1998) and are usually considered to be very proficient in English. No difficulties in comprehension of these sentences were reported by the participants and a control group's translation of the sentences was excellent. Finally, in a recent study we did manipulate language, and at least for recall in shadowing and interpreting tasks, we obtained no evidence for an effect of input language (Christoffels, 2004).

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

length of the recorded sentences was 2.7 s. on average. Six sets of five practice sentences were recorded in the same way. The first and the last sentence of each set were regarded as fillers and were not included in subsequent analyses. For the first sentence, participants still had to start their response so comprehension and production were, for a large part, not simultaneous, and, similarly, for the last sentence the input stopped before production was finished. Examples of English sentences are: *While Chris is partying until late, his parents are asleep* and *Sean wanted to go swimming but Dan preferred to play hockey*.

Cued-recall test

Of each sentence set, 10 sentences were selected from evenly distributed positions across the set. The first half of each of these sentences served as a cue in a recall test where the participant had to complete the sentence fragents as accurately as possible. The sentence fragments were presented on paper and participants were instructed to write down the remainder of the sentence. On average 4.5 words were required to complete the fragments. The presentation order of the fragments differed from the original presentation order of the sentences.

Procedure and design

Participants were tested individually. The sentences were played to the participant over headphones. On an Apple Macintosh Power PC 4400, using the software package Deck II, the stimulus material and the response of the participant were recorded synchronously on different tracks. The experimenter monitored the experiment by listening to the material over another headphone. Task instructions were presented before the start of each task and included a few example sentences (all different from the experimental sentences) and suggestions of correct responses. For the paraphrasing condition these examples showed to the participants that both a simple word order change and the use of synonyms would be considered correct responses. Each condition started with the presentation of five practice sentences, allowing the participants to familiarize themselves with the condition. In the simultaneous condition the sentences in each set were presented in one go. In the delayed condition, the presentation of the next sentence was controlled by the experimenter. A new sentence was presented only after the participant had finished the previous sentence.

After administering a complete set of 20 sentences, in each condition the corresponding recall test was administered. The participants were notified beforehand that recall tests would be administered after each task. It was stressed, however, that they should concentrate on the primary task. All conditions were administered in a single session, which took about 60 minutes in all, including short breaks between tasks and a short informal exit interview. In each sentence set the sentences were always presented in the same order. The presentation order of the tasks, simultaneity conditions, and stimulus sets were counterbalanced. For example, each of the stimulus sets used for the paraphrasing condition was presented to half of the participants in the simultaneous condition and to the other half in the delayed condition. Furthermore, half of the participants always received the simultaneous version of each task first (either set A or set B); the other half started with the delayed version. Finally, the order of the shadowing, paraphrasing, and interpreting conditions was counterbalanced. This resulted in a unique presentation order of the stimulus sets and conditions for each participant.

Data analysis

Output performance

Two analyses were performed on the output performance by transcribing the verbal output of the participants. One analysis concerned the quality of the task performance and one concerned the amount of output. Quality of task performance was measured by scoring for each sentence how well the participant performed the task. The system emphasized how well the meaning of the input was preserved in the output. To warrant the objectivity of this output measure, two independent judges rated the performance quality.[2] Other ways of measuring specifically the quality of SI, such as counting the number of different types of errors (Barik, 1994), have been criticized for focusing too much on specific words or the exact form of the output rather than on the meaning of the input and the output, and for counting errors that professional interpreters may not regard as errors (see Gile, 1991).

Output was variable across sentences and across participants, for example, some sentences were not translated at all by some participants. The quality of each output sentence was rated on a scale between 0 and 5. With 18 critical sentences per condition, the maximum quality score that could be assigned per participant, per condition was therefore 90. Zero points were given when the participant failed to say anything; 5 were given for excellent performance, taking task instruction into account. For interpreting and paraphrasing this meant that the response sentence completely conveyed the meaning

[2] This method of rating translation performance may not be suitable for output of professional interpreters. Between our participants the difference in quality of output was very large. For professionals, individual differences in performance are likely to be far smaller. They are likely to perform at the high end of the scale only and consistent rating would in general be more difficult.

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

of the stimulus sentence.[3] For shadowing the sentence should be reproduced literally. Intermediate scores were obtained for intermediate performance. The value 1 was assigned when only a few words of the stimulus sentence were reproduced; 2 was assigned when a more or less complete sentence was produced of which the meaning deviated strongly from the meaning of the stimulus sentence; 3 was assigned when the meaning of the response was related to that of the stimulus sentence but incorrect; this value was typically assigned when the opposite causal relation was produced in the output, or when subject and object had changed places; 4 was assigned when the meaning of the response was similar to that of the stimulus sentence and also approximately correct. The two judges were highly consistent in their rating of quality of performance; the inter-rater reliability was $\rho = 0.95$. This indicates that this measure was objective, in the sense that the judges in general agreed in their ratings. The average of their ratings was used in the statistical analyses to be reported below.

The second output-performance analysis involved counting the number of words in the transcribed responses in each condition. Only complete words, irrespective of their content, were counted and speech disruptions and comments on the task were excluded. The percentage was calculated of the number of output words produced in a particular condition relative to the number of input words that were presented in that condition. This measure did not involve any assessment of the quality of the output. It was hypothesized that in difficult conditions participants would be less able than in easy conditions to produce output at all, so the pattern of results of this analysis should be similar to the pattern of results of the quality of performance analysis.

Response latency

The latencies between stimulus and response were analyzed separately for the simultaneous and delayed conditions. For the simultaneous condition, we calculated the EVS for half of the sentences (the second sentence, the fourth, the sixth etc.). For each of the selected nine input sentences per condition, three content words were selected that were distributed evenly across the sentence, that is, the selected words were from the beginning, the middle, and the end of the sentence. For the three words per sentence in the input signal the onset time was determined. Subsequently, the onset time for the corresponding word in the output signal was determined. In the interpreting condition the corresponding word was the translation equivalent of the critical input word; in the paraphrasing condition it could be a synonym of the critical input word. The word-EVS was the difference between the onset time in the input and the onset time in the output. This word-EVS was averaged across the three selected words per sentence, to control for differences in word order changes across conditions. In other words, when a target word changed position in the output sentence with respect to the input sentence, other parts of the sentence changed positions too, which averages out by calculating the mean of three words from different positions within the sentence. Our results indicated that no negative EVS occurred. This could have happened if the participants had engaged in anticipating the output sentence and had readily produced it. The EVS per condition was calculated by averaging the mean EVS per sentence.

It was not possible to calculate an EVS for the delayed condition. To obtain a measure of latency for this condition we calculated the time lag between the completion of the stimulus sentence and the onset of the response sentence for the same nine sentences as used for calculating the EVS in the simultaneous condition.

Recall

Recall was measured by assigning each of the ten sentences in the test a score between 0 and 3 on how similar the recalled sentence was to the original stimulus sentence (0: no recall or error; 1: half finished; 2: similar in meaning; 3: completely correct recall). The highest possible assigned score was 30. We decided to use a rating system to measure recall because we focused on recall of meaning. Verbatim recall measures, such as the amount of overlap between words in the input and the recall, would focus more on recall of the exact form of the input. Recall performance was rated by two judges; inter-rater reliability was $\rho = 0.99$. The average rating was entered in the statistical analyses.

Results and discussion

Twenty-three participants were included in the analysis. One participant was excluded as an outlier because of extremely poor performance in the simultaneous shadowing condition (more than 4 standard deviations below the average). There was no significant effect of presentation order of the simultaneous and delayed conditions, nor of presentation order of sentence set in any of the analyses reported below. Therefore, presentation order was no longer included as a separate factor in further analyses. Omnibus repeated-measures analyses of variance (ANOVA) are reported for all dependent variables. Because our main interest was to assess the effect of simultaneity on each of the tasks and to compare performance across different tasks, we also conducted simple effect analyses.

[3] In the paraphrasing task participants could unintentionally have reproduced the exact input sentence and receive, according to our rating system, the maximum score. We had planned to exclude such responses before analysis but it turned out that they did not occur in our data.

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

Table 1. *Average rated output quality (standard deviation) per task, per condition (maximum score 90) and average percentage (standard deviation) of the number of output words relative to the number of input words.*

	Quality of performance		Amount of output*	
Task	Simultaneous	Delayed	Simultaneous	Delayed
Shadowing	88.63	89.65	100.24	100.08
	(1.80)	(.65)	(2.57)	(.41)
Interpreting	69.35	84.44	83.42	103.60
	(10.67)	(3.05)	(10.67)	(3.05)
Paraphrasing	65.26	81.72	89.71	102.25
	(16.94)	(7.79)	(9.90)	(3.47)

*When more words are produced than were present in the input the amount of output may be more than 100 percent.

Output performance

A two-way ANOVA was conducted with task (shadowing, interpreting, and paraphrasing) and simultaneity condition (simultaneous and delayed) as within-subject factors on QUALITY OF PERFORMANCE. The means and standard deviations of performance quality are presented in Table 1. The main effect of task was significant, $F(2,21) = 39.37$, $MSE = 79.2$, $p < .001$, as well as the main effect of simultaneity condition, $F(1,22) = 36.91$, $MSE = 110.2$, $p < .001$. These main effects were qualified by an interaction between task and simultaneity condition, $F(2,21) = 19.77$, $MSE = 42.5$, $p < .001$. As expected, the quality of performance was lower in the simultaneous condition than in the delayed condition in all tasks. Simple effects showed that these differences were significant for shadowing, $F(1,22) = 6.71$, $MSE = 1.8$, $p = .017$, interpreting, $F(1,22) = 40.00$, $MSE = 65.5$, $p < .001$, and paraphrasing, $F(1,22) = 24.37$, $MSE = 127.9$, $p < .001$. In the simultaneous condition, shadowing performance was significantly better than interpreting performance, $F(1,22) = 76.64$, $MSE = 55.8$, $p < .001$, as well as paraphrasing performance, $F(1,22) = 44.88$, $MSE = 140.0$, $p < .001$. Interpreting did not differ significantly from paraphrasing, $F(1,22) = 1.93$, $MSE = 99.9$, $p = .179$. In the delayed condition, the results show a similar pattern as in the simultaneous condition. Shadowing performance was significantly better than interpreting, $F(1,22) = 84.15$, $MSE = 3.7$, $p < .001$, and paraphrasing, $F(1,22) = 23.53$, $p < .001$, $MSE = 30.8$, whereas the comparison of interpreting and paraphrasing did not yield a significant effect, $F(1,22) = 2.45$, $MSE = 34.8$, $p = .132$. In other words, for both the simultaneous and the delayed condition, interpreting and paraphrasing performance appear equally good, whereas performance for shadowing is better than for both these tasks.

An ANOVA on the AMOUNT OF OUTPUT showed that the main effects of task and simultaneity condition were significant, $F(2,21) = 3.42$, $MSE = 151.9$, $p = .042$, and $F(2,21) = 32.36$, $MSE = 125.5$, $p < .001$, respectively. Also the interaction between these two factors was significant, $F(2,21) = 20.82$, $MSE = 58.3$, $p < .001$. Table 1 shows the mean percentage of output relative to the input for each condition and the corresponding standard deviations. Simple effects showed that the difference between the delayed and the simultaneous conditions was significant for interpreting, $F(1,22) = 29.91$, $MSE = 60.4$, $p < .001$, and paraphrasing, $F(1,22) = 26.25$, $MSE = 178.4$, $p < .001$, but not for shadowing, $F(1,22) < 1$, $MSE = 3.3$, $p = .77$. In the simultaneous condition, the amount of output for shadowing was larger than for both interpreting, $F(1,22) = 25.49$, $MSE = 50.0$, $p < .001$, and paraphrasing, $F(1,22) = 13.49$, $MSE = 241.2$, $p = .001$. Output was larger in paraphrasing than in interpreting but this difference was not statistically significant, $F(1,22) = 3.87$, $MSE = 117.7$, $p = .062$. In the delayed condition, only the difference between interpreting and shadowing reached significance, $F(1,22) = 9.14$, $MSE = 5.9$, $p = .006$; paraphrasing output did not differ significantly from shadowing output, $F(1,22) = 1.42$, $MSE = 100.8$, $p = .247$, nor from interpreting output, $F(1,22) < 1$, $MSE = 114.8$, $p = .672$.

Table 1 shows that for the quality of performance the absolute difference between the simultaneous and delayed conditions for shadowing is, albeit significant, small in comparison with this difference for interpreting and paraphrasing. This may explain the significant interaction between the factors task and simultaneity condition in the omnibus ANOVA. Moreover, for the amount of output, there was no difference between the simultaneous and the delayed condition for the shadowing task. These results suggest that coping with simultaneity does affect performance negatively to some extent but the effect of simultaneity is rather small. It appears that especially the combination of simultaneity and transformation had a detrimental effect on performance. Furthermore, as expected, output performance was better for shadowing than for the other two tasks, suggesting that the transformation component is a source of difficulty in SI, but this difference is especially large in the simultaneous condition, again indicating that it is the combination of these two components that is particularly problematic.

Finally, performance in the interpreting condition was not poorer than in the paraphrasing condition. This finding actually converges with the participants' perception of the relative difficulty of the tasks. Of the 21 participants who answered the question which task they found the most difficult to perform, 10 reported that they regarded simultaneous interpreting to be the most difficult task,

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

Table 2. *Average ear–voice span (standard deviation) in milliseconds per task in the simultaneous condition, and average latency (standard deviation) in milliseconds per task in the delayed condition.*

Task	EVS (simultaneous condition)	Latency (delayed condition)
Shadowing	1010 (399)	543 (160)
Interpreting	2085 (557)	1070 (316)
Paraphrasing	2626 (748)	2070 (1347)

whereas 11 felt that simultaneous paraphrasing was more difficult. The response latency analyses will shed some further light on this matter.

Response latency

Separate analyses were performed for the simultaneous and the delayed conditions. The mean EVS for the simultaneous condition and the mean response latency for the delayed condition as well as the corresponding standard deviations are reported in Table 2. In the simultaneous condition, a one-way repeated-measures ANOVA showed a significant effect of task, $F(2,21) = 67.25$, $MSE = 231323.6$, $p < .001$. Simple main-effect analysis confirmed what the means in Table 2 suggest: The EVS was significantly smaller for shadowing than for both interpreting and paraphrasing, $F(1,22) = 105.41$, $MSE = 125946.9$, $p < .001$, and $F(1,22) = 94.43$, $MSE = 317936.5$, $p < .001$, respectively, and, interestingly, the EVS was significantly larger for paraphrasing than for interpreting, $F(1,22) = 13.47$, $MSE = 250087.3$, $p = .001$. In the simultaneous condition some words were not reproduced; there was about 13.8 % of missing values.[4]

In the delayed condition a similar pattern emerged as in the simultaneous condition, even though a different measure for latency was used. Again, a main effect of task was found, $F(2,21) = 25.21$, $MSE = 548759.4$, $p < .001$, and simple effect analyses showed that the latency was shorter for shadowing than for interpreting, $F(1,22) = 77.85$, $MSE = 41122.4$, $p < .001$, and paraphrasing, $F(1,22) = 31.35$, $MSE = 855145.4$, $p < .001$. Again the latency for paraphrasing was larger than for interpreting, $F(1,22) = 15.34$, $MSE = 750010.3$, $p = .001$. In sum, as expected, for both the simultaneous and the delayed condition the response latency was shortest for shadowing. The latency was intermediate for interpreting and largest for paraphrasing. This result, in combination with output performance, suggests that the paraphrasing task is harder to perform than the interpreting task. We can therefore conclude, contrary to what has been suggested in the literature, that the paraphrasing task should not be regarded as the monolingual equivalent of SI. Unfortunately, in this study it is therefore not possible to disentangle the reformulation and language-switch aspects in the transformation component. Apparently, the specific task demands of paraphrasing, in so far as they differ from interpreting, play a more important role in determining the relative performance level in paraphrasing and interpreting than do the added demands of a language switch, that is only present in interpreting.

[4] Analysis of errors in the simultaneous conditions showed exactly the same pattern as the RT data, including a significant main effect for Task ($F(2,21) = 30.14$, $MSE = 9.5$, $p < .001$).

Table 3. *Average recall (standard deviation) per task in the simultaneous and delayed conditions.*

Task	Simultaneous condition	Delayed condition
Shadowing	15.35 (4.62)	17.41 (4.65)
Interpreting	16.35 (4.65)	22.54 (3.24)
Paraphrasing	15.26 (5.47)	19.70 (3.30)

Recall

Overall recall was better in the delayed condition than in the simultaneous condition. A repeated-measures ANOVA showed that this main effect of simultaneity condition was significant, $F(1,22) = 40.61$, $MSE = 15.22$, $p < .001$. Also, the effect of task was significant, $F(2,21) = 10.21$, $MSE = 10.87$, $p < .001$, and the interaction between task and simultaneity condition was marginally significant, $F(2,21) = 2.72$, $MSE = 3.1$, $p = .089$. The average recall scores and standard deviations are presented in Table 3.

For the recall data, simple effects analysis showed significantly lower recall in the simultaneous than in the delayed conditions for two of the three tasks: interpreting, $F(1,22) = 20.13$, $MSE = 21.9$, $p < .001$, and paraphrasing, $F(1,22) = 16.35$, $MSE = 13.8$, $p = .001$. For shadowing, the simultaneous and delayed conditions differed marginally, $F(1,22) = 4.18$, $MSE = 11.7$, $p = .053$. In the simultaneous condition, none of the differences between tasks reached significance (all Fs < 1), whereas in the delayed condition all differences between tasks reached significance. Recall for interpreting was higher than for shadowing, $F(1,22) = 17.41$, $MSE = 17.4$, $p < .001$, and paraphrasing, $F(1,22) = 9.92$, $MSE = 9.4$, $p = .005$, and recall for paraphrasing was higher than for shadowing, $F(1,22) = 7.68$, $MSE = 7.8$, $p = .011$.

These results indicate that the main effect of task is mainly due to differences in the delayed condition. Interestingly, the recall results are different from those of

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

the online performance measures reported earlier. Recall performance was similarly lower in the simultaneous condition, but there were no differences between tasks. For the delayed condition, there were task differences in both the analyses of the online data and of the recall data, but for recall, interpreting, not shadowing, showed the best performance. It is therefore not to be recommended to base conclusions on the use of recall measures when the actual interest lies, for example, in the comprehension of the input.

General discussion

The present experiment showed that a supposedly important source of difficulty in SI, simultaneity of speech comprehension and production, indeed leads to a small decrement in performance. However, it may only be a major problem for the language system if it is combined with the requirement to reformulate a message. In other words, although it seems that coping with simultaneity is demanding, it only appears to exhaust limited mental resources when it is combined with yet another demanding task component. These conclusions are based on, first, our findings of an interaction between task and simultaneity condition, which we obtained in the analyses on the quality and the amount of output. Second, for shadowing the quality of performance in the simultaneous condition was lower than in the delayed condition but still close to excellent and there was no difference in the amount of output for these conditions, suggesting that simultaneity per se did not have a large impact on shadowing performance. Third, both interpreting performance and paraphrasing performance were poorer in the simultaneous condition than in the delayed condition. This result indicates that for the latter two tasks the concurrence of comprehension and production adds to task difficulty.

Having to transform a sentence per se – both within the same language and into a different language – seems to be a source of difficulty, but again not a major one because in the delayed condition the difference in quality of performance between the three tasks was, albeit significant, rather small. Apparently, our participants were very capable of translating and paraphrasing sentences, and were almost as good as they were at repeating them. Although transformation on its own is possibly more demanding than is the simultaneity of comprehension and production on its own, again, it is mainly the combination of both these task components that caused performance to drop substantially. The interaction between simultaneity and task condition may suggest that both task components deplete the same reservoir of attentional resources.

In line with the finding of Treisman (1965) and Gerver (1974b) we found that shadowing performance was better than interpreting performance. Also, as expected, the average EVS was longer for interpreting than for shadowing. Interestingly, the EVS reported in different studies, including this one, is quite similar, even though different languages, materials, methodologies, and participant populations were used across studies (Treisman, 1965; Goldman-Eisler, 1972; Barik, 1973; Gerver, 1976). We observed averages of about 2 seconds for interpreting, and of 1 second for shadowing, which we estimated to be equivalent to about 5 words for interpreting and between 2 to 3 words for shadowing. This consistency between studies is likely to be caused by an upper boundary to the EVS due to limits of memory capacity, and by a lower boundary, due to the fact that a minimal amount of information is needed before an input fragment is disambiguated and can safely be produced (Christoffels and De Groot, in press).

We found very similar output performance for interpreting and paraphrasing.[5] This finding is similar to the corresponding result by Anderson (1994), who found a difference on only one of two performance measures that she used. We may take this to indicate that the language switch component per se does not add substantially to task difficulty in SI, that it is not important whether one or two languages are involved, and that it is the requirement of reformulating the input that taxes the processing resources most. An important result is that the latency was longer for paraphrasing than for interpreting, because this finding indicates that in the present study paraphrasing was the more difficult task. From the joint results of the output and latency analyses we can conclude that, contrary to what has been assumed before (Anderson, 1994; Green, Vaid, Schweda-Nicholson, White and Steiner, 1994), paraphrasing should not be regarded as a monolingual version of SI. This conclusion has implications for teaching SI since paraphrasing may have been wrongly perceived as an exercise preparing for interlanguage processing (see also Fabbro and Gran, 1997).

An interesting question is what aspects of the paraphrasing task are particularly demanding. As mentioned before, having to change the grammatical structure in paraphrasing may be more demanding than finding the appropriate grammatical equivalent in the output language in interpreting, even though also interpreting often involves changing the word order of a sentence. Interestingly, this line of argument implies that manipulating the input in the same (native) language is at least as demanding as the requirement to switch

[5] Note that with certain manipulations of stimulus material, it may be possible to reverse the relative performance in the interpreting and paraphrasing tasks. Some types of material may be easier to interpret whereas others may be easier to paraphrase. For example, sentences with low frequency words that have high frequency synonyms may be relatively easy to paraphrase. These same sentences may be relatively difficult to translate because low frequency target words must be found.

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

to another language system in interpreting. Furthermore, in paraphrasing literal repetition of the input, that is, reproducing the original input as the output form, must be avoided. This may be demanding because the original form of the input is an adequate way to express the intended message and likely to be highly activated. This activation must be suppressed. However, avoiding production in the source language instead of in the target language in interpreting may involve processing resources as well. Indeed, a recent account of language control, the inhibitory control model of Green (1986; 1998), suggests that even single word translation may involve high levels of control to prevent naming the word.

The results of this study have interesting theoretical implications for models of interpreting. Several authors have described two possible interpreting routes (Fabbro, Gran, Basso and Bava, 1990; Anderson, 1994; Fabbro and Gran, 1994; Isham, 1994; Isham and Lane, 1994; Paradis, 1994; De Groot, 1997, 2000; Massaro and Shlesinger, 1997), the 'transcoding route' and 'meaning-based interpreting'. Transcoding involves the literal transposition of source language linguistic units by their corresponding target language units. For the present discussion meaning-based interpreting is most relevant. It appears that, either implicitly or explicitly, models of the interpreting process assume such a route (e.g. Gerver, 1974a; Moser, 1978; Paradis, 1994; De Bot, 2000). Meaning-based interpreting refers to the idea that interpreting is conceptually mediated: The input is fully comprehended in a way similar to ordinary speech comprehension. From the non-verbal, conceptual representation of the input, production of the intended message takes place in a way similar to ordinary language production. According to a purely meaning-based route, interpreting proceeds without a separate language transformation stage or component. In that case, the difficulty of SI would only lie in the simultaneity of the comprehension and production processes. However, given the small but significant effect of the transformation component, that we found separately from the effect of simultaneity, our data suggest that transformation constitutes a processing step in the interpreting process, and that models of interpreting should incorporate such a step.

The pattern of results revealed by the recall tests was quite different from that revealed by the direct performance measures. In the delayed condition recall was highest for interpreting, followed by paraphrasing, and, then finally, by shadowing. Recall in the simultaneous condition was about the same across tasks. Note that the simultaneous condition is most comparable to earlier research. Since we found no task differences, the results for this condition differ from earlier research. Gerver (1974a) and Lambert (1988) found better performance for interpreting than for shadowing, whereas Darò and Fabbro (1994) obtained the reverse pattern, that is, a higher recall of digits after shadowing than after interpreting. Perhaps the different results can be explained by the fact that the tasks' measures and procedures were different in the various studies and/or by the fact that the participants in the other studies were professional interpreters (Gerver, 1974a; Lambert, 1988) or advanced students of interpreting (Darò and Fabbro, 1994), whereas our participants were untrained in SI. Professionals may have learned to cope with the simultaneity of speech comprehension and production and may, therefore, show differences in recall between tasks in a simultaneous condition. This idea is supported by the finding that interpreters have a large memory span in comparison to other groups of participants (Padilla et al., 1995; Bajo, Padilla and Padilla, 2000; Christoffels and De Groot, in press).

In fact, this discussion touches upon a much-debated issue in interpreting. A problem that frustrates research on simultaneous interpreting is the difficulty of finding enough experienced professional interpreters to participate in experimental studies, especially if a specific language combination is required (Massaro and Shlesinger, 1997). At the same time, professionals in the field of interpreting are concerned about the ecological validity of laboratory research and doubt whether studies that test non-professional participants tell us anything about the process of interpreting in highly skilled professionals. Indeed, given that we only included participants with no prior interpreting experience, we do not know whether our results generalize to professional interpreters. However, in this particular case the finding, for example, that simultaneity is not as demanding as might have been expected is likely to hold true for participants with much experience in handling such simultaneity situations as well.

Moreover, there are a number of reasons why we believe the use of fluent bilinguals who are not trained in SI as participants in studies on interpreting is both legitimate and informative. One is the existence of some empirical data that suggests that professional interpreters and inexperienced bilinguals process language in the same way. In the present study the EVS in SI was similar to studies involving participants with SI experience. This suggests that the temporal constraints on performing this task is similar for different groups. Furthermore, Dillinger (1994) found only small quantitative differences and no qualitative differences between these two groups and argued that interpreting is not a special, acquired skill but an ability that accompanies bilingualism naturally. Moser-Mercer, Frauenfelder, Casado and Künzli (2000) did not report any differences between first year students of interpreting and professional interpreters in comprehension scores after shadowing in a pilot study. Green et al. (1990, 1994) reported no behavioural laterality differences in shadowing, paraphrasing, and

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

interpreting between professional and ordinary bilinguals, although they did find significant differences between interpreters and monolinguals.

Another reason is that the use of professional interpreters may be also problematic: Shlesinger (2000) points out that the use of idiosyncratic strategies by individual interpreters may contaminate the basic interpreting process that one might want to understand separately of those strategies. Finally, we are not only interested in trying to understand interpreting and how it is supported by our language system, but also in what the SI task can tell us about the language system proper. In this respect it makes sense to study interpreting in non-professionals.

It is striking that our participants produced interpretations at all and that they did not perform at floor level even though none of them had any previous experience in SI. Admittedly, their level of SI performance is unlikely to be close to the level obtained by professional interpreters. Still, it appears from the present data that interpreting at a less skilled level is within reach of every reasonably proficient bilingual, which suggests that it is supported by normal bilingual language processing. It also suggests that bilinguals, when engaged in a translation task, may suffer less interference from the non-target language than might be expected given that both languages have to be active simultaneously. The results of a recent study on Stroop-effects in word translation (Miller and Kroll, 2002) converge with this suggestion in showing that language selection occurs relatively early in the production of a translation. In other words, there may exist no competition from the non-target language in translation. The input language may give information about what language is not to be produced. Possibly, bilinguals can use this language cue somehow to prevent interference from the non-target language (Miller and Kroll, 2002).

As mentioned earlier, in the delayed condition recall after interpreting was better than after shadowing, even though quality of performance was far better in shadowing. In other words, in the delayed condition performance was better in the more complex task condition. The type of processing involved in interpreting apparently leads to better recall. Although not without criticism, the levels-of-processing framework (Craik and Lockhart, 1972; Lockhart and Craik, 1990) is helpful in explaining this finding because in this framework the type of processing of the to be remembered material is instrumental in recall performance (see also e.g. Craik, 2002; Lockhart, 2002; Watkins, 2002): Material that is analyzed semantically is recalled better than material that is processed at a more shallow, non-semantic, level. Even though in shadowing the input is analyzed up to a semantic level (Marslen-Wilson, 1973), it seems reasonable to assume that interpreting and paraphrasing require deeper semantic analysis of the input. Hence the higher recall scores in interpreting and paraphrasing than in shadowing. However, in a levels-of-processing account it is not clear why recall after paraphrasing is not equally good as recall after interpreting, because there is no obvious difference between these tasks in the depth of analysis of the input. One possibility is that in interpreting language could serve as a cue to recall, whereas in paraphrasing the participants' paraphrase may have interfered with recall of the stimulus sentence.

In the introduction we argued that the production of speech in SI may negatively affect recall as in any other task involving the production of speech while processing input. We found that across tasks recall in the simultaneous condition was clearly poorer than in the delayed condition. It thus seems that having to comprehend and produce speech at the same time may indeed have reduced recall in the simultaneous conditions. The fact that there were no significant differences between tasks in the simultaneous condition suggests that this disruptive effect on recall may have overshadowed effects caused by differences in task difficulty. Alternatively, the reduced recall in the simultaneous condition can be explained as an effect of divided attention: The simultaneous condition can be regarded as a dual task situation, where both comprehension and production consume attentional resources. Recently, we replicated the lack of difference between simultaneous interpreting and shadowing when comparing recall of prose (Christoffels, 2004).

In sum, one conclusion suggested by this study is that the paraphrasing task cannot be considered the monolingual equivalent of SI that some researchers have considered it to be. Probably due to a number of unique characteristics of the paraphrasing task, a comparison between the paraphrasing and the interpreting tasks does not provide a suitable way to answer the question whether, within the transformation component, reformulating or language switching is the more important sub-component. More importantly, we established that both the simultaneity of comprehension and production and the transformation component are sources of cognitive complexity in SI. However, our results also indicate that the role of each of these components separately is limited, and that of the two, transformation is the more demanding component. Especially the combined demands of simultaneity and transformation make simultaneous interpreting the complex task that it is.

References

Anderson, L. (1994). Simultaneous interpretation: Contextual and translational aspects. In Lambert & Moser-Mercer (eds.), pp. 101–120.

Baddeley, A. D., Lewis, V. & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, 36A, 233–252.

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

Bajo, M. T., Padilla, F. & Padilla, P. (2000). Comprehension processes in simultaneous interpreting. In A. Chesterman, N. Gallardo San Salvador & Y. Gambier (eds.), *Translation in context*, pp. 127–142. Amsterdam: John Benjamins Publishing.

Barik, H. C. (1973). Simultaneous interpretation: Temporal and quantitative data. *Language and Speech*, 16, 237–270.

Barik, H. C. (1994). A description of the various types of omissions, additions and errors of translation encountered in simultaneous interpretation. In Lambert & Moser-Mercer (eds.), pp. 121–137.

Chernov, G. V. (1994). Message redundancy and message anticipation in simultaneous interpretation. In Lambert & Moser-Mercer (eds.), pp. 139–153.

Christoffels, I. K. (2004). Cognitive studies in simultaneous interpreting. Unpublished doctoral dissertation. University of Amsterdam, Amsterdam.

Christoffels, I. K. & De Groot, A. M. B. (in press). Simultaneous interpreting: A cognitive perspective. In Kroll & De Groot (eds.).

Craik, F. I. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.

Craik, F. I. M. (2002). Levels of processing: Past, present . . . and future? *Memory*, 10, 305–318.

Craik, F. I. M. & Kester, J. D. (2000). Divided attention and memory: Impairment of processing or consolidation? In E. E. Tulving (ed.), *Memory, consciousness, and the brain: The Tallinn Conference*, pp. 38–51. Psychology Press/Taylor & Francis.

Danks, J. H., Shreve, G. M., Fountain, S. B. & McBeath, M. K. (1997) (eds.). *Cognitive processes in translation and interpreting*. Thousand Oaks: Sage Publications.

Darò, V. & Fabbro, F. (1994). Verbal memory during simultaneous interpretation: Effects of phonological interference. *Applied Linguistics*, 15, 365–381.

De Bot, K. (2000). Simultaneous interpreting as language production. In Englund Dimitrova & Hyltenstam (eds.), pp. 65–88.

De Groot, A. M. B. (1997). The cognitive study of translation and interpretation. In Danks et al. (eds.), pp. 25–56.

De Groot, A. M. B. (2000). A complex-skill approach to translation and interpreting. In Tirkkonen-Condit & Jääskeläinen (eds.), pp. 53–68.

Dijkstra, T., Van Jaarsveld, H. & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, 1, 51–66.

Dillinger, M. (1994). Comprehension during interpreting: What do interpreters know that bilinguals don't? In Lambert & Moser-Mercer (eds.), pp. 155–189.

Dussias, P. E. (2001). Sentence parsing in fluent Spanish–English bilinguals. In J. L. Nicol (ed.), *Two languages: Bilingual language processing*, pp. 159–176. Cambridge, MA: Blackwell Publishers.

Englund Dimitrova, B. & Hyltenstam, K. (2000) (eds.). *Language processing and simultaneous interpreting*. Amsterdam: John Benjamins Publishing.

Fabbro, F. & Gran, L. (1994). Neurological and neuropsychological aspects of polyglossia and simultaneous interpretation. In Lambert & Moser-Mercer (eds.), pp. 273–317. Amsterdam: John Benjamins Publishing.

Fabbro, F. & Gran, L. (1997). Neurolinguistic research in simultaneous interpretation. In Y. Gambier, D. Gile & C. Taylor (eds.), *Conference interpreting: Current trends in research*, pp. 9–27. Amsterdam: John Benjamins Publishing.

Fabbro, F., Gran, L., Basso, G. & Bava, A. (1990). Cerebral lateralization in simultaneous interpretation. *Brain and Language*, 39, 69–89.

Frauenfelder, U. H. & Schriefers, H. (1997). A psycholinguistic perspective on simultaneous interpretation. *Interpreting*, 2, 55–89.

Gerver, D. (1974a). The effect of noise on the performance of simultaneous interpreters: Accuracy of performance. *Acta Psychologica*, 38, 159–167.

Gerver, D. (1974b). Simultaneous listening and speaking and the retention of prose. *Quarterly Journal of Experimental Psychology*, 26, 337–341.

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

Gile, D. (1991). Methodological aspects of interpretation (and translation) research. *Target*, 3, 153–174.

Gile, D. (1997). Conference interpreting as a cognitive management problem. In Danks et al. (eds.), pp. 196–214.

Goldman-Eisler, F. (1972). Segmentation of input in simultaneous translation. *Journal of Psycholinguistic Research*, 1, 127–140.

Goldman-Eisler, F. (1980). Psychological mechanisms of speech production as studied through the analysis of simultaneous translations. In B. Butterworth (ed.), *Language production*, vol 1: *Speech and talk*, pp. 143–153. London: Academic Press.

Green, A., Sweda-Nicholson, N., Vaid, J., White, N. & Steiner, R. (1990). Hemispheric involvement in shadowing vs. interpretation: A time-sharing study of simultaneous interpreters with matched bilingual and monolingual controls. *Brain and Language*, 39, 107–133.

Green, A., Vaid, J., Schweda-Nicholson, N., White, N. & Steiner, R. (1994). Lateralization for shadowing vs. interpretation: A comparison of interpreters with bilingual and monolingual controls. In Lambert & Moser-Mercer (eds.), pp. 331–355.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language*, 27, 210–223.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1, 67–81.

Hartsuiker, R. J. & Westenberg, C. (2000). Persistence of word order in written and spoken sentence production. *Cognition*, 75, 27–39.

Hyönä, J., Tommola, J. & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology*, 48A, 598–612.

Isham, W. P. (1994). Memory for sentence form after simultaneous interpretation: Evidence both for and against

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press

deverbalization. In Lambert & Moser-Mercer (eds.), pp. 191–211.

Isham, W. P. (2000). Phonological interference in interpreters of spoken-languages: An issue of storage or process? In Englund Dimitrova & Hyltenstam (eds.), pp. 133–149.

Isham, W. P. & Lane, H. (1994). A common conceptual code in bilinguals: Evidence from simultaneous interpreting. *Sign Language Studies*, 85, 291–316.

Kroll, J. F. & De Groot, A. M. B. (eds.) (in press). *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press.

Kroll, J. F. & Dussias, P. E. (2004). The comprehension of words and sentences in two languages. In T. Bhatia & W. Ritchie (eds.), *Handbook of bilingualism*, pp. 169–200. Cambridge, MA: Blackwell Publishers.

Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.

Lambert, S. (1988). Information processing among conference interpreters: A test of the depth-of-processing hypothesis. *Méta*, 33, 377–387.

Lambert, S. (1992). Shadowing. *Méta*, 37, 263–273.

Lambert, S. & Moser-Mercer, B. (1994) (eds.). *Bridging the gap: Empirical research in simultaneous interpretation.* Amsterdam: John Benjamins Publishing.

Lockhart, R. S. (2002). Levels of processing, transfer-appropriate processing, and the concept of robust encoding. *Memory*, 10, 397–403.

Lockhart, R. S. & Craik, F. I. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology*, 44, 87–112.

MacWhinney, B. (1997). Simultaneous interpretation and the competition model. In Danks et al. (eds.), pp. 215–232.

MacWhinney, B. (in press). A unified model of language acquisition. In Kroll & De Groot (eds.).

Malakoff, M. & Hakuta, K. (1991). Translation skill and metalinguistic awareness in bilinguals. In E. Bialystok (ed.), *Language processing and language awareness*, pp. 141–166. New York: Oxford University Press.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522–523.

Massaro, D. W. & Shlesinger, M. (1997). Information processing and a computational approach to the study of simultaneous interpretation. *Interpreting*, 2, 13–53.

Miller, N. A. & Kroll, J. F. (2002). Stroop effects in bilingual translation. *Memory and Cognition*, 30, 614–628.

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

Moser-Mercer, B. (1994). Aptitude testing for conference interpreting: Why, when and how. In Lambert & Moser-Mercer (eds.), pp. 57–68.

Moser-Mercer, B. (1997). Beyond curiosity: Can interpreting research meet the challenge? In Danks et al. (eds.), pp. 176–195.

Moser-Mercer, B., Frauenfelder, U. H., Casado, B. & Künzli, A. (2000). Searching to define expertise in interpreting. In Englund Dimitrova & Hyltenstam (eds.), pp. 107–131.

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. 61–72. Turku: University of Turku.

Paradis, M. (1994). Toward a neurolinguistic theory of simultaneous translation: The framework. *International Journal of Psycholinguistics*, 10, 319–335.

Rinne, J. O., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., Sipilä, H. & Sunnari, H. (2000). The translating brain: Cerebral activation patterns during simultaneous interpreting. *Neuroscience Letters*, 294, 85–88.

Schweda-Nicholson, N. (1987). Linguistic and extralinguistic aspects of simultaneous interpretation. *Applied Linguistics*, 8, 194–205.

Shlesinger, M. (2000). Interpreting as a cognitive process: How can we know what really happens? In Tirkkonen-Condit & Jääskeläinen (eds.), pp. 4–15.

Tirkkonen-Condit, S. & Jääskeläinen, R. (eds.) (2000). *Tapping and mapping the processes of translation and interpreting*. Amsterdam: John Benjamins Publishing.

Treisman, A. M. (1965). The effects of redundancy and familiarity on translating and repeating back a foreign and a native language. *British Journal of Psychology*, 56, 369–379.

Van Hell, J. G. & De Groot, A. M. B. (1998). Disentangling context availability and concreteness in lexical decision and word translation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 51A, 41–63.

Watkins, M. J. (2002). Limits and province of levels of processing: Considerations of a construct. *Memory*, 10, 339–343.

Received June 24, 2003
Revision received January 16, 2004
Accepted January 21, 2004

https://doi.org/10.1017/S1366728904001609 Published online by Cambridge University Press