

Working memory and expertise in simultaneous interpreting*

Minhua Liu, Diane L. Schallert and Patrick J. Carroll
The University of Texas at Austin

This study describes an experiment that aimed to determine if performance differences exist in simultaneous interpreting by individuals with similar general cognitive abilities, but different skills specific to the task of simultaneous interpreting. Professional interpreters' performance in simultaneous interpreting from English into Mandarin was compared to that of two groups of student interpreters, beginners and advanced. The results showed that the professional interpreters who were not different from students in their general working memory capacity outperformed student interpreters. This difference was attributed, at least in part, to the development of specific skills in managing competing demands on limited cognitive resources. One important domain-specific skill observed in this study is the ability to select more important ideas from the speech input under conditions where stringent task demands jeopardize completeness and accuracy of the output. Professional interpreters' generally superior performance is discussed within the descriptive framework of working memory theory.

Keywords: domain-specific skills, expertise, resource allocation, simultaneous interpreting, working memory

Introduction

Simultaneous interpreting involves listening to a message in one language and immediately rendering that message verbally into another language, while at the same time continuing to listen to the incoming message. The moment-by-moment operations in the process of simultaneous interpreting involve expressing in the target language the meaning of segment A, just heard from the speech in the source language, attending to the incoming segment B and temporarily holding

segment B and/or its meaning in memory while continuing to translate segment A, and at the same time monitoring the target language output for accuracy and smoothness of delivery. Several studies using different language combinations found that interpreters engaged in concurrent speaking and listening for a substantial percentage of the speech input's total speaking time (e.g. Chernov 1979; Gerver 1975). By its very nature, simultaneous interpreting imposes large demands on an individual's cognitive resources. Simultaneous interpreters often seem to work near the saturation level of their cognitive resources, as can be seen in the deterioration in professional interpreters' performance even when no difficulty can be identified in the source speech (Gile 1999).

Psychologists have developed the construct of working memory to describe the ways that normal processing is influenced by the capabilities and limitations of various cognitive systems that work together. The working memory construct has proved useful for addressing issues related to the degradation and even breakdown of performance when task demands exceed the individual's ability to perform with complete success, as, for instance, in the study of dyslexia (e.g. Gathercole & Baddeley 1993). The construct of working memory can thus serve as a consistent framework for understanding the cognitive aspects of the complex and demanding skill of simultaneous interpreting.

Working memory is considered to be a limited-capacity mechanism that is involved in both the processing and storing of currently active information while tasks are being carried out (Baddeley & Hitch 1974). The functions of the storage and processing of information compete with each other for the limited resources in working memory and tend to trade off against each other when working memory capacity approaches its limit (Daneman & Carpenter 1980). In proposing a "capacity theory of comprehension," Just and Carpenter (1992) suggested that individual differences in working memory capacity account for performance differences in language comprehension. The *reading span test*, a test for measuring individual differences in working memory capacity, was devised by Daneman and Carpenter (1980). Reading span was found to correlate highly with both reading comprehension and listening comprehension, and was consequently interpreted as a measure of general working memory capacity available to be used in language processing of all sorts (Daneman & Carpenter 1980).

Just, Carpenter, and colleagues proposed an alternative to the traditional limited capacity approach for explaining working memory limitations. They suggested that individual differences are a function of differences in the efficiency with which resources are allocated to various concurrent tasks

(Carpenter et al. 1994; Just & Carpenter 1992). The two explanations are often considered compatible, because a less efficient working memory may require consumption of more working memory resources for the same processing, and thus result in a virtually smaller working memory capacity (Just & Carpenter 1992; Miyake et al. 1994). Just and Carpenter (1995) developed the concept of a “resource allocation policy” in discussing the efficiency of working memory in allocating resources. Resource allocation policy is a process that is invoked when resources in working memory are stretched and about to be exceeded. It regulates the trading-off that is needed as one attempts to coordinate all the processes required by a task. According to them, resource allocation policy includes sensitivity to the characteristics of the task being performed, and is subject to strategic control. An important consequence of emphasis on efficiency over fixed capacity is that working memory limitations can be altered by practice or training (Just & Carpenter 1992). Therefore, rather than thinking of working memory capacity as a fixed attribute, Just, Carpenter and colleagues suggested that working memory varies as a function of an individual’s efficiency with the specific task processes (Daneman 1991; Daneman & Carpenter 1983). If they are correct, there is no absolute measure of working memory; it can only be measured with respect to the specific operations in a given domain.

This domain-specific efficiency view of working memory capacity is in line with research on expert performance. Studies have shown that expertise is highly domain-specific (e.g. Ericsson & Charness 1994; Walker 1987). The primary difference between experts and novices in a domain was not their general aptitude or working memory capacity, but their domain knowledge or skills (e.g. Chase & Simon 1973; Chi et al. 1982). Experts were shown to have acquired more domain-specific declarative as well as procedural knowledge that allowed them to circumvent limits on their processing capacity and thus outperform novices in performing a domain task (Chase & Simon 1973; Chi et al. 1982). For example, in sports or the game of chess, acquired anticipatory skills circumvent general limits and allow a domain-specific expansion of working memory capacity (Ericsson & Kintsch 1995; Ericsson & Lehmann 1996). The expansion of the functional working memory capacity may also be a result of selective encoding of relevant information (Ericsson & Lehmann 1996). Studies in several domains have reported that experts are better able to select the more relevant information for processing and to reduce the amount of effort spent on processing less relevant information, whether in the field of computer programming (Adelson 1984), reading (McGee 1982), medicine (Groen & Patel 1988) or chess (Holding & Reynolds 1982).

By extension, then, one would expect that an examination of the working memory of expert simultaneous interpreters would also hint at strategies these experts are using to manage the challenges of the task. In fact, several studies comparing the performance of experts and novices in simultaneous interpreting showed that expert interpreters demonstrate quantitative as well as qualitative differences from novice interpreters in their interpretation performance. Professional interpreters not only correctly processed more propositions in a text than novices (Dillinger 1994), but the types of omissions observed in their interpretation output were also less serious than those of the novices (Barik 1975). In addition, more experienced interpreters seemed to process larger chunks of the input (e.g. Davidson 1992; McDonald & Carpenter 1981) and their translation was less literal than that of novice interpreters (e.g. Barik 1975; McDonald & Carpenter 1981). In an error detection task using the dichotic listening paradigm, it was found that while there was no significant difference between professional and student interpreters in detecting correctly translated sentences, students detected significantly more syntactic errors and professionals significantly more semantic errors (Fabbro et al. 1991; Ilic 1990). Although extremely useful in allowing us to understand what expert interpreters are doing as they interpret, these studies did not focus specifically on the role of working memory in explaining differences between more and less expert interpreters.

Some research has suggested that expert interpreters have superior general working memory. If this is so, it will be difficult to separate the effects of task-specific skills from the greater general cognitive capacity indicated by the superior working memory. Darò and Fabbro (1994) and Padilla et al. (1995) measured working memory capacity using traditional digit span or Daneman and Carpenter's (1980) reading span measures. They claimed that interpreters with more expertise had a larger working memory span. One implication drawn by these researchers, given the general nature of the digit span and reading span measures, was that expertise is intrinsically related to a general cognitive quality of the expert interpreter and that this general cognitive quality contributes to performance differences in the task of simultaneous interpreting.

The present work provides a somewhat different view of the role of working memory in explaining performance differences between more and less expert interpreters. We compare three groups that do not differ in their general working memory capacity, so general capacity cannot be used to explain performance differences. We then look for specific differences in the selection of input and the quality of output by interpreters engaged in simultaneous interpreting using specially constructed texts. Given the near-equivalence of the

groups' general working memory abilities, we can attribute differences to specific, skill-related abilities which, according to our conceptual framework, are present in working memory.

The study

In the study reported here, professional interpreters' performance in simultaneous interpreting from English into Mandarin was compared to that of two groups of student interpreters. The general working memory capacity of participants was measured by using the *listening span test*, a listening version of the reading span test which, unlike traditional digit and word span measures of working memory, has been shown to positively correlate with comprehension (Daneman & Carpenter 1980).

Two domain-specific skills of expertise in simultaneous interpreting were investigated in the present study. First, we examined whether interpreters with higher expertise levels were better at selecting the more important over the less important information to interpret when their working memory was being challenged during simultaneous interpreting. Challenge to working memory in this task was introduced by varying the difficulty level of the text to be interpreted as well as the speed of delivery, two aspects of the simultaneous interpreting task that have been demonstrated in the past to impede performance (e.g. Gerver 1969; Tommola & Helevä 1998). Under these challenges, the skill of selective encoding was investigated by way of differentiating performance on what were called *essential idea units*, which were critical components of a sentence's message, from *secondary idea units*, which had an elaborative function. We predicted that the more expert interpreters would be more effective than the less experienced interpreters in focusing on essential idea units, as demonstrated by better selective interpretation of these portions of the input. The second domain-specific skill investigated in this study was along the lines of what Just and Carpenter (1995) meant in describing the resource allocation policy. The skill of efficient resource allocation in simultaneous interpreting was examined by the construction of sentences with important information at the beginning of the sentence that would require the immediate attention of the interpreter. It was predicted in the present study that interpreters with higher expertise levels would demonstrate better efficiency in the allocation of working memory resources, a better ability to withstand the challenges put on them by other task demand variables such as difficulty of the previous sentence and speed of input, by managing to interpret all the information presented in these sentences. In response to the call by Just and Carpenter

(1995) for more research on the resource allocation policy of working memory elicited by tasks with different characteristics, the present study focused on using the task of simultaneous interpreting to better understand the construct of working memory and its allocation policy.

Method

Participants

Eleven professional interpreters and 22 interpretation students participated in this study. The professional interpreters had, up to the time of the experiment, received at least one year of full-time interpretation training and had had at least two years of professional experience, with at least 40 working days per year. The 22 students consisted of 11 students who were at the end of their second-year (final year) of training and 11 at the end of their first-year of training at the time of the experiment. All participants either spoke Mandarin as their first language and English as their second language, or had comparable first-language competence in speaking both of these languages. The three groups — *professionals*, *advanced students*, and *beginning students* — represented three different levels of mastery in simultaneous interpreting.

All professional interpreters were recruited in Taiwan, one of the major markets for Mandarin-English conference interpretation. The interpretation students were recruited from two graduate-level translation and interpretation training institutions in Taiwan and one in California, USA. Common to all participants was that they had been admitted into their respective institutions with a required 600 score on TOEFL and/or had passed an entrance examination before beginning interpretation training.

Materials

Three English texts were chosen and modified for this study. To minimize potential effects of individual differences in background knowledge, all three texts were chosen from different publications in the field of translation and interpretation and were all on the general subject of professional issues in the field. The first text discussed how professionals should behave on the job (the Manners text; Snyder 1994), the second discussed problems in translating for the media (the Media text; Paajannen-Manila 1990), and the third discussed the code

of ethics for translators and interpreters (the Ethics text; Weber 1984). A text of a similar nature was also chosen for the warm-up session in the experiment.

Critical sentence

In each of the three experimental texts, 20 sentences were identified as *critical sentences*, and distributed throughout the texts so that critical sentences were never adjacent to each other. In the next step, both difficult and easy versions of each critical sentence were constructed using the Flesch-Kincaid Readability Index to differentiate the difficulty levels. The difficult version of each critical sentence had a readability level of 14 or higher, and the easy version had a level of 12 or lower. The difficult and easy versions of a critical sentence were similar in meaning and differed only in form. The sentence preceding a critical sentence was always constructed to have a readability level of 11 or lower so as to reduce the possibility of a carry-over effect on the interpretation of the critical sentence from the preceding sentence. Table 1 lists samples of the difficult and easy versions of three critical sentences used in the experiment.

Either the difficult or the easy version was assigned to each critical sentence

Table 1. Examples of critical sentences from each of the three texts used in this study

Difficult	Easy
<p>Example 1 (from the Manners text)</p> <p>Now, this longing for independence and freedom, a longing to structure your own quality of life is a common desire, especially among those of us who came of age in the sixties. (14.59)</p>	<p>Now, most of us want independence and freedom, and to choose our own lifestyle. This is especially true for those of us who came of age in the sixties. (5.93)</p>
<p>Example 2 (from the Media text)</p> <p>The documentary reported an attempt that had been made to relieve symptoms of Parkinson disease where dopamine-producing cells were transplanted from the kidney into the brain. (14.52)</p>	<p>The documentary was about a way to help Parkinson disease by moving cells that made dopamine from the kidney into the brain. (11.76)</p>
<p>Example 3 (from the Ethics text)</p> <p>With the exception of work done for government agencies or the courts, translators and interpreters are generally not sworn to professional secrecy. (14.98)</p>	<p>Translators and interpreters are usually not made to keep their work a secret, except when working for the government and the courts. (11.76)</p>

Note. The number in parentheses indicates the readability level of the sentence.

in a random manner so that there was a mixture of 10 difficult and 10 easy critical sentences in a text. This combination of sentences became the first version of a text used in the experiment. The other version was constructed by using the easy version of every difficult sentence and the difficult version of every easy sentence. Sentences other than the critical sentences in a text were identical in the two versions. The six versions ranged in length from 1626 words (Version 2 of the Media text) to 2161 words (Version 2 of the Manners text), and the two versions of any one text did not differ in length by more than 22 words. Similarly, the six versions ranged in overall readability from 7.61 (Version 1 of the Manners text) to 9.5 (Version 2 of the Ethics text), with no more than a difference of .36 in overall readability between any two versions of the same text.

Idea units

In an effort to test the hypothesis that interpreters with more expertise are better at selecting the more important information to interpret than less expert individuals, each critical sentence was further divided into idea units, ranging in number from two to ten units per sentence. Two native speakers of English rated the importance of each idea unit and categorized the idea units as being either essential or secondary to the meaning of the critical sentence. The inter-rater reliability (Kappa) was 0.75. After discussing any discrepancy in judgment with the raters, the first author acted as the third and final judge on the importance of the idea units on which the two raters disagreed. The numbers of essential idea units were 34, 30 and 30 for each of the three texts respectively, and those of secondary idea units were 34, 34 and 30. Table 2 lists samples of critical sentences from each of the three texts used in the experiment with their essential and secondary idea units identified.

Continuation sentence

Following Gile's (1995) description of how difficulties in resource management may plague interpreters and affect how well they interpret subsequent text, the sentence immediately following a critical sentence was constructed to become the *continuation sentence*. A continuation sentence was written in such a way that the first two or three words of the sentence were essential for establishing a full and correct meaning of that sentence. These two or three words, if misheard or not heard at all, would hinder the comprehension of the entire sentence. The construction of a continuation sentence was intended to determine whether participants in this study would be able to coordinate their

Table 2. Essential and secondary idea units of each of the critical sentences in Table 1

Example 1

1. Most of us want independence and freedom.
2. Most of us want to choose our own lifestyle.
3. Those people who grew up in the 60's are especially so.

Example 2

1. The documentary was about a way to help Parkinson disease.
2. The way was to move cells from the kidney to the brain.
3. The cells produced dopamine.

Example 3

1. Translators and interpreters are generally not sworn to professional secrecy.
2. They must swear to professional secrecy when working for the government.
3. They must swear to professional secrecy when working for the court.
4. 2) and 3) are exceptions.

Note. The essential idea units are underlined.

Table 3. Continuation sentences for each of the critical sentences in Table 1

1. Life is not about nine to five.
 2. It was difficult to get the title changed.
 3. Nonetheless, sensitive material should not be talked about while one is working.
-

Note. Words underlined are the critical words at the beginning of a continuation sentence.

resources in such a way so as to allow them to pay attention to the following continuation sentence even when they were still in the midst of interpreting the critical sentence. Each continuation sentence was constructed to be short in length and simple in syntactic structure. All continuation sentences used in the experiment had a readability score of 10 or lower based on the Flesch-Kincaid Readability Index. A continuation sentence remained the same for both the easy and difficult versions of a preceding critical sentence. Table 3 shows a sample of continuation sentences used in the experiment.

Recording and speech digitization of the experimental texts

Two native speakers of English, one female and one male, read and recorded the warm-up text as well as the first version of each of the experimental texts in their entirety on separate audiocassettes. In a separate recording, they recorded just the critical sentences in the other version of each text. For this second round of recording, the readers were told to read the text between critical sentences silently in order to produce a more natural reading of these critical sentences in the absence of context. The readers were also asked to become

familiar with the texts to such a degree that they could record their reading of each text as if they were delivering a talk to an audience.

All recorded speeches and critical sentences were transformed into SoundEdit-16 files using a Macintosh computer to allow further modification of the recordings. To create the second version of a speech, the set of critical sentences separately recorded were spliced into the complete version of that speech to substitute the original critical sentences in the first version. In total, six versions were created, two for each of three speeches, in the form of SoundEdit files.

The speed of the six versions was modified through speech digitization using SoundEdit-16 so that each version would have two further versions at two different speech rates, 150 words per minute (wpm) and 130 wpm, respectively. The speech rate for the speech used in the warm-up session was set at 110 wpm. The durations of the pause before each critical sentence and that before each continuation sentence were equalized so that in the fast version there was a 1.8-second pause before each critical sentence and a one-second pause before each continuation sentence. The duration of pauses in the slow version was automatically adjusted when the speech rate was set at 130 wpm and thus, the pauses before the critical and continuation sentences were slightly longer in the slow version.

Thus, the final set of stimulus materials for this experiment consisted of 12 different recordings of speeches, a fast and a slow recording of each of the two versions for each of the three texts. The speed of delivery and the difficulty level of sentences were expected to act conjointly as challenges to processing. All stimulus materials lasted from 11 to 17 minutes.

As a way to help participants familiarize themselves with the topic of each speech, an abstract of approximately 50 words was prepared in English for the participants to read before they began interpreting a particular speech.

Listening span test

In this study, the participants' working memory capacity was measured by using a test similar to the listening span test developed by Daneman and Carpenter (1980). Participants were required to listen to sets of unrelated English sentences that ranged from 13 to 16 words in length. The number of sentences in each set increased from two to five as the task proceeded. There were five sets of sentences for each size. Seventy unrelated sentences of a non-technical nature were chosen from various texts and modified for this test. No specific domain

knowledge was required to comprehend these sentences. Out of the 70 sentences, 13 were made nonsensical but were kept grammatically correct. As participants listened to each sentence, they were required to indicate whether the sentence was *good*, defined to them as a sentence that made sense, or *bad*. After listening to all the sentences in a set, the participants were required to recall the last word of each sentence. Each sentence ended with a different word. For example, participants might be presented with the following set of two sentences: *Fearing more air strikes, people are stocking up on food, and men are planning escape routes; Rats cause enormous damage to buildings by gnawing wood, pipes, walls and foundations.* After listening to the two sentences, the participants were to recall the words “routes” and “foundations.” Participants were asked to recall as many of the last words as they could of a set of sentences.

In this task, the final words in each sentence constitute a concurrent load on working memory because the words from previous sentences must be stored and maintained even as a participant is making a judgment about the comprehensibility of the current sentence. The nonsensical sentences were used to ensure that participants were engaged in both storing the sentence-final words and processing a new sentence instead of merely trying to remember the last word of each sentence. The maximum number of sentences for which the participants successfully recalled all of the final words for at least three out of five sets was defined as their working memory span. Half credit was given if the participants were correct on two out of five sets. For example, if participants successfully recalled three out of five three-sentence sets, they were considered to have a working memory span of 3. If they were correct on two out of the five three-sentence sets, a working memory span of 2.5 was assigned.

A male native speaker of English recorded all sentences on an audiocassette at a normal speaking rate. At the beginning of the five sets of sentences of each size, the speaker announced the size of the sets. At the end of each set there was a tone indicating the end of a set. The sets were presented in order of increasing size.

Design

The experiment involved five factors, including one between-subject factor — *group*, and three within-subject factors — *speech*, *importance of idea unit*, and *difficulty*. The *speed* variable, even though incorporated in the design of the stimuli, was only considered as a between-subject factor in the analysis for each speech due to the fact that the design was not completely balanced.

There were two main dependent variables in this study. The first involved

calculating for each critical sentence the proportion of idea units correctly translated and then calculating a mean proportion for all the critical sentences in a speech. The proportion of idea units was computed separately to provide two within-subject measures — importance of idea unit (essential or secondary) and difficulty (easy or difficult). Therefore, each participant had a total of four scores for each speech, i.e., the mean proportion correct scores of the essential idea units in easy sentences, essential idea units in difficult sentences, secondary idea units in easy sentences, and secondary idea units in difficult sentences.

The second dependent variable consisted of the proportion of correctly translated continuation sentences in each speech. Each participant, for each speech, had a mean proportion correct score for all continuation sentences with a preceding easy critical sentence and one for all continuation sentences with a preceding difficult critical sentence.

Procedure

The order of administering the two versions of the stimulus texts was counterbalanced across participants for each expertise group. Passage order was counterbalanced and determined by a blocking strategy to distribute practice and/or fatigue effects.

All participants were run individually in a simultaneous interpretation room sitting with a microphone in front of them. They listened to the speeches through headphones. A blank audiocassette tape was used to record their interpretation output on one channel with the speech input recorded on the other channel. A practice session was administered in order to familiarize participants with the facilities and allow them to warm up for the experiment. After the warm-up trial, the summary for the speech in the first session was handed to the participants to read at their own pace. The summary was retrieved after the participants indicated that they were ready for the session. Upon finishing each session, the participants were given a chance to take a short break, after which the summary for the speech used in the next session was handed out. All participants interpreted the three experimental speeches. This part of the experiment lasted from one hour to one-and-half hours for each participant.

The listening span test was administered after the simultaneous interpreting sessions. The participants listened to recorded sentences from an audiocassette tape player put in front of them. As mentioned earlier, to encourage participants to listen to each sentence carefully rather than simply to pay attention to the last

word of each sentence, participants were asked to judge the meaningfulness of each sentence while they were listening to it. They were provided with a sheet of paper with sets of two underlined spaces marked with either *Good* or *Bad* for each sentence. Participants were instructed to check *Good* if a sentence was considered meaningful or *Bad* if it was nonsensical. They were also told that each sentence, whether meaningful or not, was grammatically correct and that grammar should not affect their judgment. Whether or not participants correctly verified the meaningfulness of the sentences was ignored in the analysis.

At the tone indicating the end of each set of sentences, the audiocassette player was paused to allow the participants to write down the last words of the sentences they had just heard. A sheet of paper with underlined spaces marked with the numbers for each set was provided to the participants to write down the last words. The participants worked at their own pace in writing down the words and made a gesture upon finishing a particular set of sentences. The listening span test lasted approximately 30 minutes for each participant.

Data scoring

Each participant's translation of all critical sentences and continuation sentences was transcribed. To prepare for scoring, the translated versions of a particular sentence produced by all 33 participants were listed on one sheet of paper and their order randomized so as to mask the original order of participants. A sheet itemizing the idea units for each critical sentence in each original speech was prepared for the scorers.

Two scorers judged the transcribed output against each idea unit in the critical sentences as either correctly translated (score 1) or incorrectly translated (score 0). Both scorers were native speakers of Mandarin and were highly proficient users of English. The two scorers first judged 25% of all translated sentences and achieved a 0.79 inter-rater reliability on Kappa. They then discussed any discrepancy in their judgment. One of the scorers then continued to score all the remaining translated sentences and her scores were used in all subsequent analyses of the data.

For the scoring of the continuation sentences, one of the scorers checked the transcribed output against the original sentence. Each continuation sentence was treated as one single unit and was scored as either correct (score 1) or incorrect (score 0).

Results

Listening span test

Before analyzing participants' performance in the simultaneous interpreting task, we looked at their responses to the listening span test. Means of the three groups, professional interpreters, advanced students and beginning students, were 3.64 ($SD=1.23$), 3.23 ($SD=1.19$), and 3.36 ($SD=1.33$) respectively, with a possible maximum score of 5. An ANOVA with listening span scores as the dependent variable indicated that there was not a significant difference among the groups, $F(2,30)=0.31$, NS. Due to the small number of participants in each group, it is possible for a significance test (e.g. ANOVA) to disguise a substantial effect; therefore, an effect size measure was also calculated. The effect size as measured by eta squared was very small, $\eta^2=.02$. Thus, only 2% of the variability in working memory capacity can be accounted for by the group differences.¹ We may safely assume that the three groups of participants did not differ in their general working memory capacity.

Critical sentence

An ANOVA using the mean proportion correct scores for each of the three speeches as dependent variables and group, speech, importance of idea unit and difficulty of critical sentence as independent variables showed a highly significant group effect, $F(2,30)=11.43$, $p<.001$. This result indicated that expertise influenced the participants' performance in simultaneous interpreting (means and standard deviations are provided in the last column of Table 4). Professional interpreters were better than advanced students, who in turn, performed better than beginning students. Post-hoc pair-wise comparisons indicated that the performance difference between professional interpreters and advanced students, $p<.01$, and that between professional interpreters and beginning students, $p<.001$, were significant. The difference between the two student groups was not significant, but approached conventional levels of significance, $p=.06$.

Importance of idea unit

A significant main effect of importance of idea unit was observed in the analysis, $F(1,30)=6.14$, $p<.05$. Participants correctly interpreted more essential (mean = .36, $SD=.14$) than secondary idea units (mean = .33, $SD=.12$). There was also a significant interaction effect between importance of idea unit and group, $F(2,30)=3.59$, $p<.05$. The mean scores of each group for essential versus

Table 4. Mean proportion correct scores and standard deviations across all speeches as a function of importance of idea unit

Group	Essential units	Secondary units	All units
	Mean (SD)	Mean (SD)	Mean (SD)
Beginning Students	.25 (.10)	.25 (.11)	.25 (.10)
Advanced Students	.34 (.09)	.32 (.10)	.33 (.10)
Professionals	.48 (.13)	.42 (.08)	.45 (.10)
Mean	.36 (.14)	.33 (.12)	

Note. $n = 11$ in each group.

secondary idea units (Table 4) suggest that professional interpreters not only performed significantly better at interpreting both essential and secondary idea units than the other two groups, but they were also much better at differentiating essential from secondary idea units. Individual t -tests were used to compare performance on essential versus secondary idea units for each group. There was a significant difference between types of idea units for the professionals, $t(11) = 2.60$, $p < .05$. The difference was not significant for either the beginning students, $t(10) = -.08$, NS, or for the advanced students, $t(10) = 1.12$, NS.

Difficulty of critical sentence

The difficulty levels of the critical sentences had a significant effect on performance, $F(1,30) = 5.24$, $p < .05$. Combining across levels of expertise, participants performed better in the easy condition (mean = .36, $SD = .13$) than in the difficult condition (mean = .33, $SD = .14$). The mean scores of each group for easy versus difficult versions (Table 5) suggest that the professionals dealt with the two versions of critical sentences in a different manner than the two student groups. Sentence difficulty did not seem to affect the professionals (mean = .45 for both conditions) as it did the student groups. However, there was a lack of a difficulty X group interaction effect, $F(2,30) = 1.14$, NS, implying that the three groups were not significantly different from each other in the way they reacted to sentences with different difficulty levels.

Continuation sentence

In the analysis of continuation sentences, an ANOVA using the mean proportion scores of correctly translated continuation sentences for each of the three speeches as dependent variables, and group, speech, and difficulty as independent

Table 5. Mean proportion correct scores and standard deviations across all speeches as a function of difficulty of the critical sentences

Group	Critical Sentence		Continuation Sentence	
	Easy	Difficult	Easy	Difficult
Beginning Students	.27 (.10)	.23 (.11)	.33 (.13)	.27 (.08)
Advanced Students	.35 (.11)	.32 (.09)	.39 (.15)	.35 (.14)
Professionals	.45 (.10)	.45 (.12)	.50 (.17)	.44 (.10)
Mean	.36 (.13)	.33 (.14)	.41 (.17)	.35 (.13)

Note. $n = 11$ in each group.

variables showed a highly significant group effect, $F(2,30) = 6.04$, $p < .01$. Post-hoc comparisons indicated that professional interpreters (mean = .47, $SD = .12$) performed significantly better than advanced students (mean = .37, $SD = .14$), $p < .05$, and beginning students (mean = .30, $SD = .09$), $p < .01$. Advanced students' performance, however, was not significantly different from that of the beginning students.

Difficulty of critical sentence

The main effect of difficulty also turned out to be significant, $F(1,30) = 6.69$, $p < .05$, indicating that the difficulty level of the preceding critical sentence had a strong impact on participants' performance in interpreting the following continuation sentence, even though the continuation sentence remained the same for the two versions of a critical sentence. The continuation sentences that were preceded by a difficult critical sentence were less correctly interpreted (mean = .35, $SD = .13$) than the ones following an easy critical sentence (mean = .41, $SD = .17$), indicating a bigger carry-over effect from a difficult critical sentence. However, the interaction between difficulty and group was not significant, $F(2,30) = .17$, NS. As can be seen in Table 5, the difference between the proportion of correct scores of continuation sentences under the two difficulty levels of the critical sentences was quite consistent across the three expertise groups, suggesting that the three groups were not performing differently from each other in interpreting the continuation sentences under the two difficulty levels of the critical sentences.

Speech effect

In discussing the analyses of these two dependent variables, we have so far omitted

any mention of the speech effect. In the analysis involving the scores for the critical sentences, there was not a main effect for speech, but the interaction effects for both speech X importance of idea unit, $F(2,60) = 6.03$, $p < .01$, and speech X difficulty, $F(2,60) = 4.17$, $p < .05$, were both significant. For the analysis of scores for the continuation sentences, a significant speech effect emerged, $F(2,60) = 24.05$, $p < .001$, but there was no speech X difficulty interaction. Thus, participants' performance seemed affected by the speech they were interpreting. However, the fact that there was no speech X group interaction for either dependent measure suggests that the three groups of participants were responding to the three speeches similarly.

Speed effect

Finally, the speed of delivery was analyzed by conducting analyses separately for each speech, using speed as a between-subject factor. Results indicated that only for the Ethics speech was there an effect for speed, $F(1,28) = 9.68$, $p < .01$, with performance significantly lower for the faster delivery (mean = .25, $SD = .12$) than for the slower delivery (mean = .37, $SD = .16$). However, the group X speed interaction effect was not significant for this speech, or the other two, indicating that, on the whole, the speeds chosen for the delivery of these speeches did not differentially affect these participants.

Discussion

Simultaneous interpreting is a particularly challenging task because it involves many language processes competing with each other as the individual strives to produce a reasonably high quality output. The study included texts that were designed specifically to allow us to see the influence of text characteristics such as difficulty and importance of idea unit on the performance of simultaneous interpreting. Professional interpreters were able to interpret more of the source language input accurately than student interpreters. More importantly, a qualitative difference was also observed in that professional interpreters were better at selecting the more important idea units over the less important ones when interpreting.

In choosing our participants, we drew from the ranks of those who were already professional interpreters and, for our comparison groups, from those fully committed to training as interpreters. We hoped to find three groups that

were comparable in their general cognitive abilities while distinctly different in their experience with simultaneous interpreting. Insofar as our measure of the general cognitive ability (the listening span test) was concerned, we succeeded in recruiting three very similar groups. Entrance requirements for these graduate-level training programs further guarantee similarity in general language proficiency and academic ability among the three groups compared in this study. Therefore, we cannot attribute the consistently better performance by professional interpreters in this study to a larger general working memory capacity or better general cognitive ability. It is interesting to note that the differences in general working memory capacity that have been noted as a distinction between experts and novices by some other researchers (Darò & Fabbro 1994; Padilla et al. 1995) are apparently not reliable characteristics of expert and novice interpreters.

If general intellectual ability and general working memory capacity cannot account for performance differences among groups in this study, differences observed in the results of the listening span test and the actual simultaneous interpreting task may be understood if we accept the claim that professional interpreters draw on strategies and knowledge specific to the simultaneous interpretation process, and perhaps even specific to the interpretation of the two particular languages involved. We may cautiously infer that, when performing outside of that skilled domain, as in the listening span test and in countless daily activities, professional interpreters revert to more general memory management strategies and have no advantage over other fluent speakers.

As has been observed in other domains, in which experts often demonstrate a better ability at selecting the more relevant information for further processing (e.g. Holding & Reynolds 1982; McGee 1982), professional interpreters in this study were better at selecting the more important idea units in simultaneous interpreting than the student interpreters. The results of this study are consistent with the suggestion by Just, Carpenter, and colleagues that working memory varies as a function of an individual's efficiency with the specific task processes (Daneman 1991; Daneman & Carpenter 1983).

A more efficient working memory in performing the task of simultaneous interpreting should manifest itself even more clearly when the task is made more challenging by way of introducing more difficult sentences. It was hypothesized in this study that the gap in performance difference between professional interpreters and student interpreters would expand under greater difficulty. Unfortunately, this result was not supported statistically, as shown by the lack of a significant group X difficulty interaction in the analysis of the critical sentences.

Nevertheless, the means of the three groups for easy and difficult critical sentences (Table 5) showed the expected pattern, with professional interpreters' performance not affected by greater difficulty, while performance of both advanced and beginning students suffering in the difficult condition. Although the statistical analysis must determine our current interpretation of these results, the pattern of means certainly motivates further investigation in future research.

Just and Carpenter's (1995) view is that the resource allocation policy of working memory may be sensitive to the characteristics of a task. The task of simultaneous interpreting particularly demands efficient resource allocation of working memory as the task involves overlapping cognitive processes. It is the act of temporarily switching attention from an unfinished utterance to an incoming message that makes the task of simultaneous interpreting so different from what we are used to in our usual mode of verbal communication. This demand was especially obvious in interpreting the continuation sentences in this study because of the way these sentences were designed. The two or three critical words at the beginning of a continuation sentence, together with a very short pause (approximately one second) before the sentence began, made guessing about meaning and the use of the pause time for finishing uttering the previous sentences unlikely. The fact that continuation sentences were more accurately interpreted by professional interpreters indicated that the professionals were better at understanding the first two or three words that were critical in comprehending the meaning of the whole continuation sentence.

However, the hypothesis that expertise was related to how efficiently an interpreter moves from interpreting one sentence to taking in the next one was not fully supported by the results of the continuation sentences, where no group X difficulty effect was found. Interpreters with more experience seemed to be just as likely to be affected by the difficulty of the preceding critical sentences as less experienced interpreters. The mean scores of the three groups for continuation sentences under two conditions (Table 5) show that all three groups were affected to a similar degree.

Although the interaction of group and difficulty was not significant for either the critical sentence or for the continuation sentence, it may be instructive to observe the pattern of means shown in Table 5. This exercise is prompted by the fact that the two types of sentences were designed to have mutual influence on each other, although they have been analyzed separately in the ANOVAs reported above. By looking at the means, we can see if there are different working memory resource allocation patterns between the professional interpreters and the two groups of students.

In interpreting the results from Table 5, it should be kept in mind that *easy* and *difficult* refer only to the difficulty level of the critical sentence; the continuation sentence is the same for both the easy and difficult versions. Professional interpreters showed equivalent performance on the easy and difficult critical sentences, suggesting that their working memory resources were managed in a way to allow them to adjust to immediate stimulus difficulty in order to maintain the quality of their output. This is unlike the two groups of student interpreters, who showed degraded performance with the difficult critical sentences, suggesting their lack of ability to make immediate adjustments to the difficulty of the critical sentence. When we view performance on the continuation sentences, we find that all three groups were affected by the difficulty of the preceding critical sentence. This inter-sentential influence is an important motivation for the use of a working memory model. The fact that the degree of influence was about the same for all three groups (4 to 6 percentage points) indicates that not even the professional interpreters can escape this lagging effect of input demands on performance. However, the professional interpreters differed from the students in that they were able to maintain their high level of performance, adjusting to the difficulty of the critical sentence. The difficulty of the critical sentences did influence processing of the continuation sentences, but only to the same degree that it influenced the students. All the while, the professional interpreters were maintaining a significantly higher general level of performance than either student group.

It is interesting that advanced students, compared with professional interpreters who consistently outperformed either group of students, did not distinguish themselves clearly from beginning students. Advanced students' performance, though generally better than beginning students', as can be seen from the mean scores in both Table 4 and 5, only turned out to be marginally better than beginning students statistically. It is quite possible that for gaining true expertise in simultaneous interpreting, training and practice in the classroom are not sufficient. To become expert interpreters, individuals may simply have to test their learned skills in the real world — to interpret at real meetings for people who are truly dependent on their interpretation.

A positive message about training for simultaneous interpreting, and perhaps for most specialized training, that we can extract from the results of the present study is that expertise may rely much more on acquisition of identifiable domain-specific skills than on general qualities such as a large working memory capacity. One thing suggested by the present study is that expert interpreters, as a way to circumvent limits of working memory resources, are

selective in terms of what to interpret and what not to interpret. This is consistent with what every experienced interpreter knows, i.e. that the skill involves constant judgment and selection, not a mere transposition from one language to the other. The literature on expertise development is inconclusive about the degree to which rules and strategies can be directly taught; but whether learned from direct or indirect instruction, selection and judgment will be critical components of the expert's skill.

Another implication from the results of this study is that professional interpreters, through extended practice and experience in the task of simultaneous interpreting, have developed a working memory that processes information in a different, and often times more efficient pattern. However, how expert interpreters allocate their working memory resources in a moment-by-moment manner cannot be inferred from the results of the present study. Therefore, this study does not claim that making working memory more efficient is the key to successful training in simultaneous interpreting. Indeed, we have used the construct of working memory as a shorthand for the process of integrating multiple cognitive systems to serve skilled performance. Unnecessary reification of working memory only obscures the more valuable task of specifying the nature of the underlying cognitive systems and their means of integration. Our attempt to focus on the processes of selection of more important sentence elements and adjusting to overall sentence difficulty are a small contribution to this complex goal.

Due to the small number of participants in this study, we do not claim that our findings can be generalized across different degrees of expertise or across different experimental procedures. Nor do we intend to claim that the skills investigated in this study explain all the effects or lack thereof in the present study. Such a claim would overly simplify the view of the process and influential factors involved in the extremely complex cognitive and communication act that we call simultaneous interpreting.

Notes

* This paper is based on part of Minhua Liu's unpublished doctoral dissertation submitted to the University of Texas at Austin. We thank Daniel Gile for his valuable comments during the preparation of the manuscript and an anonymous reviewer for constructive suggestions on an earlier version of this paper.

1. Eta squared is interpreted as the proportion of systematic variability measured by the dependent variable that may be accounted for by the independent variable.

References

- Adelson, B. (1984). When novices surpass experts: The difficulty of the task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory and Cognition* 10, 484–495.
- Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8). New York: Academic Press, 47–89.
- Barik, H. C. (1975). Simultaneous interpretation: Qualitative and linguistic data. *Language and Speech* 18, 272–297.
- Carpenter, P. A., Miyake, A. & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego: Academic Press, 1075–1122.
- Chase, W. G. & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 215–281.
- Chernov, G. V. (1979). Semantic aspects of psycholinguistic research in simultaneous interpretation. *Language and Speech* 22, 277–295.
- Chi, M. T. H., Glaser, R. & Rees, E. (1982). Expertise in problem solving. In R. S. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates, 1–75.
- Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research* 20, 445–464.
- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19, 450–466.
- Daneman, M. & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition* 9, 561–584.
- Darò, V. & Fabbro, F. (1994). Verbal memory during simultaneous interpretation: Effects of phonological interference. *Applied Linguistics* 15, 365–381.
- Davidson, P. M. (1992). Segmentation of Japanese source language discourse in simultaneous interpretation. *The Interpreters' Newsletter, Special Issue* 1, 2–11.
- Dillinger, M. (1994). Comprehension during interpreting: What do interpreters know that bilinguals don't? In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins, 155–190.
- Ericsson, K. A. & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist* 49, 725–747.
- Ericsson, K. A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review* 102, 211–245.
- Ericsson, K. A. & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology* 47, 273–305.
- Fabbro, F., Gran, B. & Gran, L. (1991). Hemispheric specialization for semantic and syntactic components of languages in simultaneous interpreters. *Brain and Language* 41, 1–42.
- Gathercole, S. E. & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gerver, D. (1969). The effects of source language presentation rate on the performance of simultaneous conference interpreters. In E. Foulke (Ed.), *Proceedings of the Second*

- Louisville Conference on Rate and/or Frequency-Controlled Speech. Louisville, KY: Center for Rate-Controlled Recordings, University of Louisville, 162–184.
- Gerver, D. (1975). A psychological approach to simultaneous interpretation. *Meta* 20, 119–128.
- Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. Amsterdam: John Benjamins.
- Gile, D. (1999). Testing the Effort Models' tightrope hypothesis in simultaneous interpreting — A contribution. *Hermes. Journal of Linguistics* 23, 153–172.
- Groen, G. J. & Patel, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates, 287–310.
- Holding, D. H. & Reynolds, R. I. (1982). Recall or evaluation of chess positions as determinants of chess skill. *Memory and Cognition* 10, 237–242.
- Ilic, I. (1990). Cerebral lateralization for linguistic functions in professional interpreters. In L. Gran & C. Taylor (Eds.), *Aspects of applied and experimental research on conference interpretation*. Udine: Campanotto, 101–110.
- Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99, 122–149.
- Just, M. A. & Carpenter, P. A. (1995). The intensity dimension of thought: Pupillometric indices of sentence processing. In J. M. Henderson, M. Singer & F. Ferreira (Eds.), *Reading and language processing*. Mahwah, NJ: Lawrence Erlbaum Associates, 182–211.
- Liu, M. (2001). *Expertise in simultaneous interpreting: A working memory analysis*. Unpublished doctoral dissertation, The University of Texas at Austin.
- McDonald, J. Carpenter, P. A. (1981). Simultaneous translation: Idiom interpretation and parsing heuristics. *Journal of Verbal Learning and Verbal Behavior* 20, 231–247.
- McGee, L. M. (1982). Awareness of text structure: Effects on children's recall of expository text. *Reading Research Quarterly* 17, 581–590.
- Miyake, A., Carpenter, P. A. & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology* 11, 671–717.
- Paajannen-Manila, A. (1990). Translating medical texts for the media. In M. Jovanović (Ed.), *Proceedings of the 12th World Congress of FIT*. Belgrade: Prevodilac, 418–421.
- Padilla, P., Bajo, M. T., Cañas, J. J. & Padilla, F. (1995). Cognitive processes of memory in simultaneous interpretation. In J. Tammola (Ed.), *Topics in interpreting research*. Turku: Centre for Translation and Interpreting, University of Turku, 61–71.
- Snyder, N. (1994). The paradox of professionalism. In D. L. Hammond (Ed.), *Professional issues for translators and interpreters*. Amsterdam: John Benjamins, 13–21.
- Tammola, J. & Helevä, M. (1998). Language direction and source text complexity: Effects on trainee performance in simultaneous interpreting. In L. Bowker, M. Cronin, D. Kenny & J. Pearson (Eds.), *Unity in diversity? Current trends in translation studies*. Manchester: St. Jerome, 177–186.
- Walker, C. H. (1987). Relative importance of domain knowledge and overall aptitude on acquisition of domain-related information. *Cognition and Instruction* 4, 25–42.
- Weber, W. K. (1984). Professional ethics and code of conduct. In W. K. Weber, *Training translators and conference interpreters*. Orlando, FL: Harcourt Brace Jovanovich, 59–63.

Authors' address

Minhua Liu
Laboratory for Cognitive Neuropsychology
Yang-Ming University
Taipei, Taiwan.
E-mail: flcg1068@mails.fju.edu.tw

About the authors

Minhua Liu received her MA in translation and interpretation from the Monterey Institute of International Studies and her PhD in Foreign Language Education from the University of Texas at Austin. She has been a conference interpreter and an interpretation instructor for 15 years. Her main research interest lies in the cognitive aspects of simultaneous interpreting. She is now a postdoctoral fellow with the Laboratory for Cognitive Neuropsychology of Yang-Ming University in Taiwan.

Diane L. Schallert is a professor of Educational Psychology and Foreign Language Education at the University of Texas at Austin where she teaches undergraduate courses in the teacher preparation program and graduate courses on the psychology of learning and language processes. Her research is focused on the nature and experience of classroom talk, both oral and written, and on the emotional, motivational and cognitive factors associated with learning.

Patrick J. Carroll received his PhD from the University of Massachusetts in Amherst in experimental psychology. He was a Sloan Postdoctoral Fellow at the Center for Cognitive Science and is now senior lecturer of Psychology at the University of Texas at Austin. He has published research in both written and spoken language processing. His current research interests concern the cognitive modulation of the experience of pain and the nature of communication of pain and other health symptoms.