Note-taking in consecutive interpreting: An empirical study drawing on eye-tracking, pen-recording, and voice-recording data

KUANG, HUOLINGXIAO

Note-taking in Consecutive interpreting:

An Empirical Study Drawing on Eye-tracking, Pen-recording, and Voice-recording Data

Huolingxiao Kuang

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

School of Modern Languages and Cultures
University of Durham

September 2022

Declaration

The candidate confirms that the work submitted is her own and that it has not been submitted, in whole or in part, in any previous application for a degree. Appropriate credit has been given where reference has been made to the work of others. Part of the discussions in Chapter 3, Chapter 5, Chapter 6 and Chapter 7 are developed from the author's preliminary work of this project which has been published in the form of refereed papers:

- Kuang, H., & Zheng, B. (2022). Note-taking effort in video remote interpreting: Effects of source speech difficulty and interpreter work experience. *Perspectives*. https://doi.org/10.1080/0907676X.2022.2053730
- Kuang, H. & Zheng, B. (2022). How does interpreting performance correlate with note-taking process, note-taking product and note-reading process? An eye-tracking and pen-recording study. *Across Languages and Cultures,* 23(2), 167-186. doi: https://doi.org/10.1556/084.2022.00281
- Zheng, B. & Kuang, H. Working memory and interpreting studies, in J. W. Schwieter, & E. Weng (Eds.). *The Cambridge handbook of working memory and language* (pp. 698-721). Cambridge University Press.

Copyright permissions have been granted for all the figures reproduced from others' work in this thesis.

Huolingxiao Kuang

(Signature)

Contents

Abstract

This empirical study provides a systematic investigation into the process of note-taking, the product of note-taking, and the process of note-reading, in remotely-conducted and video-mediated consecutive interpreting. With an eye-tracking and pen-recording approach, this investigation collects data from 29 student interpreters and 20 professional interpreters, in an experiment containing four English-to-Chinese interpreting segments with two levels of task difficulty.

The project, firstly, examines the effects of interpreter work experience and source speech difficulty on the process and product of note-taking. Results show that the experience and difficulty effects are mainly found in the overt visual attention that the participants pay to the notes and the physical effort of note-writing. These effects are less detected in the cognitive effort, temporal management or note choices during note-taking. In addition, the participants' note-taking effort is affected by their distorted perception of interpreting difficulty, which is caused by the different sequences of interpreting segments. Secondly, by combining visualization tools of eye-tracking such as heat maps with fixation-related measures, this study identifies a group-based processing pattern and a high level of cognitive load during the process of note-reading. Thirdly, it discovers a positive relationship between the average effort of taking a note and that of reading a note. Meanwhile, it finds a trade-off between various note forms (full words vs. abbreviations) and note languages (Chinese vs. English) in the cognitive effort of note-taking and that of note-reading. Lastly, by looking into the complex associations between the participants' note-taking behaviour and interpreting performance, it finds that, compared with the note-taking process, the note-reading process is more closely related to interpretation quality. In addition, note quantity is positively correlated with the interpreters' interpreting scores in the easy segments of interpreting, but not in the difficult ones. These findings provide pragmatic implications for interpreting practice and interpreter training.

Keywords: Note-taking process and product, note-reading process, eye-tracking, pen-recording, interpreter work experience, source speech difficulty

Acknowledgments

I would like to extend my sincerest gratitude to my supervisors Prof. Binghan Zheng and Prof. Sergey Tyulenev for their generous support throughout my Ph.D. study. I still remember that when I was applying to this doctoral program, Prof. Zheng squeezed his time for a short interview with me before his urgent flight back to the U.K. from China. I had no choice but frankly told him that I had very limited experience in research. Prof. Zheng was also very frank by replying that "then, it's gonna be really hard if you are going to do this". Words cannot express my gratitude Prof. Zheng for trusting this bold young girl and giving her an opportunity to start a new journey in her life. Additionally, this Ph.D. research project would not have been possible without the enormous supports from Prof. Zheng who generously provided knowledge and expertise. I will always be grateful to Prof. Zheng for what he has taught me about the right ways to conduct scientific research and what he has shown me, and also the rest of the TPR team, the true spirit of researchers, which is to discover the truth and make an utmost contribution to the advancement of human knowledge.

My thanks also go to China Scholarship Council for funding my Ph.D. study and the faculty of the School of Modern Languages and Cultures at Durham University for providing all the support, especially during the difficult time of COVID-19 pandemic. Without their generous support, I would not be able to conduct any experiments during the pandemic or even start this doctoral program.

I would like to thank Prof. Sijia Chen, Prof. Chao Han, Prof. Xia Xiang, Prof. Yanmei Liu, and Dr. Yanfang Jia for their insightful comments on my research and great support during my experiments. When I first decided on this research topic, I went online and seared for literature where Prof. Chen's work came out immediately. I did not realize that I would read and cite this excellent researcher's work dozens of times in the following four years. No words can express my gratefulness for Prof. Chen and all the researchers whose excellent research outputs have inspired me in my Ph.D. study. My special thanks go to Dr. Yanfang Jia who helped me to set up all my experiments in Changsha city without any hesitation. It is such a great help that Dr. Jia managed to arrange 15 experiments in two days and found me a perfect venue for free. I sincerely hope that one day I could also give my hand to her, whenever it is necessary.

I am also grateful to Prof. Michael Carl and colleagues who participated in the 2019 boot camp held at Kent State University. There is much gratefulness at the bottom of my heart which was overwhelmed by the pressure of research at that time and was not delivered to the participants during the boot camp. I feel obliged to say it here that after all, I am grateful for everything.

Moreover, COVID-19 accompanies my Ph.D. study from Year 2 till the end. At the beginning of the pandemic, my mom as a doctor in Wuhan stayed at the front line for half a year and I could not find my way back to my hometown in China. My dad who was not allowed to take one step out of the house managed to live on himself for six months. The great resilience my parents showed me during the pandemic and the unconditional support they gave me during this special time made me realize how powerful we can be as we face surmounting difficulties in life.

Thank you to all my amazing friends and colleagues, Dr. Rachel Yu Weng, Dr. Scarlett Yixiao Cui, Mr. Faustino Dardi, Miss Mingqing Xie, Miss Dariga Baktygereyeva, Miss Renwen Xu, Miss Xueni Zhang, Miss Jenney Mao and Mr. Haoshen He who have shared this experience along the way. I always call Rachel “学姐姐” and Scarlett “小学姐” because they are indeed sisters to me. Thank you Rachel for tolerating and answering all my basic questions during data analysis and thesis writing, and thank you Scarlett for keeping me away from Ph.D. depression with your beautiful smile and warm words. Thank you Fausto for telling me again and again that I deserve the most wonderful things in this complex world. Thank you Dariga for your wisdom and kindness. You always knew instantly what I was confused about in research and life, and you always have the most touching words for me to move forward.

My special thanks go to Miss Natalia Zhou and Miss Stephanie Tu, two excellent professional interpreters in the Chinese interpreting market. It is through their wide professional networks that I was able to recruit around 30 professional interpreters to participate in my experiments. It is also because of them that I was encouraged not to give up my career as a professional interpreter. You are friends, colleagues and tutors to me. And I am sure that we will meet each other in the booth soon after the pandemic.

I'd also like to thank my best friend, Miss Mélanie Xu, for calling me every night in the first two years of my Ph.D. to ease my anxiety while she was herself pursuing a MA degree in France. We have been supporting each other for 27 years and we will continue to do so for the rest of our lives. Thank you Miss Doris Fan and Mr. Adrian Wang for allowing me to go back to my early 20s every time when we met. Meeting you guys is healing to me. Last but not least, I want to thank you, Mr. Bill Wang, my fiancé, for showing up before I finish my Ph.D. and allowing me to be vulnerable when things went wrong.

The list of thanks will never end. I would like to extend my most sincere gratitude to all the people who have helped me during the last four years. Their unwavering support is much appreciated.

List of Tables

List of Figures

List of Abbreviations

AIIC	International Association of Conference Interpreters
AOI	Area of interest
ARI	Automatic Readability Index
C	Coordination
CC	Click count
CI	Consecutive interpreting
CIRIN Bulletin	Conference Interpreting Research Information Network Bulletin
CSSCI	Chinese Social Sciences Citation Index
EPS	Ear-pen span
FFD	First fixation duration
FluDel	Fluency of delivery
FPD	First pass duration
GFP	Gaze sample to fixation percentage
GTS	Gaze time on the screen
H	Hypothesis
IBM	International Business Machines Corporation
InfoCom	Information completeness
L	Listening and analysis
LTM	Long-term memory
M	Memory
MFD	Mean fixation duration
MMI	Man-machine interaction
MTI	Master of Translation and Interpreting
ms	millisecond
NASA TLX	NASA Task Load Index
NP	Note production
NR	Note-reading

P	Production
RI	Remote interpreting
RQ	Research question
RVC	Revisit count
S	Second
SD	Standard deviation
SI	Simultaneous interpreting
SL	Source language
SPD	Second pass duration
SR	Speech reconstruction from memory
TFC	Total fixation count
TFD	Total fixation duration
TL	Target language
TLQual	Target language quality
TSB	Translation Studies Bibliography
VC	Visit count
VMI	Video mediated interpreting
VRI	Video remote interpreting
VWM	Verbal working memory
WM	Working memory
wpm	Words per minute

Chapter 1: Introduction

1.1 Research background

The past two decades have witnessed a notable growth of interest in research on the interpreting process (e.g., Andres, 2002; S. Chen et al., 2021; Seeber, 2013; Stachowiak-Szymczak & Korpal, 2019). Nonetheless, the majority of the research focuses on simultaneous interpreting (SI) rather than on consecutive interpreting (CI) (e.g., S. Chen, 2020b; Chmiel, 2021). One important reason for this imbalance of research on the two interpreting modes is the general notion that SI is more cognitively demanding than CI. For the same reason, in China, the National MTI (Master of Translation and Interpreting) Commission advises MTI programs to arrange CI courses before SI courses (Wang & Lei, 2009). However, recent corpus-based studies have shown that "the cognitive load of CI, if not higher, may be as high as that of SI" (Lv & Liang, 2019, p. 91). Eye-tracking studies on video remote interpreting (VRI) have also proven that it is in the consecutive mode that interpreters present longer mean fixation duration and more frequent shifts of attention, implying a higher cognitive load of interpreting in CI than in SI (Doherty et al., 2022). At the same time, researchers have started to emphasize the special cognitive demands of CI because of its "storage-plus-processing" (Dong & Cai, 2015, p. 74) operation of working memory (WM). As long as interpreters cannot successfully store the source information in their short-term memory during the input stage of CI, they will not be able to process the stored information and deliver it in the target language (TL) during the output stage. All of these findings and discussions suggest that the complexity of cognitive processing involved in CI could be underestimated, but insofar as CI on its own is concerned, only very limited interpreting process studies are available. It remains rarely researched what happens inside the 'black box' of interpreters' minds as they perform interpreting consecutively.

One important cause of the high cognitive load of CI lies in note-taking, a subtask that is not shared by any other mode of interpreting. In CI, interpreters deliver their rendition only after the speaker finishes a segment of speech, which usually lasts from between 45 seconds to 3 minutes long (Setton & Dawrant, 2016). Therefore, interpreters generally resort to note-taking to release the pressure on their short-term memory. To be more specific, the integrated concept of "note-taking" can be divided into three parts chronologically: the process of note-taking, the product of note-taking, and the process of note-reading (e.g., Albl-Mikasa, 2008; S. Chen et al., 2021). During note-taking, interpreters encode selected information in the source speech into written notes; while during note-reading, interpreters decode their notes to recall the source speech and deliver the interpretation. As long as interpreters can retrieve their memory of the source speech from the product of note-taking, there are no absolute rules governing the appearance of notes. The inclusion of note-taking in CI is a double-edged sword: on the one hand, it facilitates memory retrieval during the output stage of CI by serving as a "bridge connecting the gap between the interpreter's memory and the production" (Chuang, 2008, p. 95) of the target speech; on the other hand, most errors and omissions in CI can be traced back to the comprehension phase because of the "strong cognitive pressure" caused by the "cognitive and mechanical aspects of note-taking during comprehension" (Gile, 2020, p. 13). These substantial impacts that note-taking can exert on interpreters' cognitive load of interpreting and interpretation quality demonstrate the importance of researching effective approaches to note-taking (e.g., Abuín González, 2012; Gile, 1995/2009; Gillies, 2017).

Discussions concerning effective note-taking can be traced back to Jean-François Rozan (1956), who firstly proposed the seven principles of note-taking. However, empirical studies on the process of note-taking in CI were not available until 2002 when Dörte Andres (2002) for the first time video-taped interpreters' note-taking scenes. In recent years, S. Chen (2017a) made a breakthrough in note-taking studies by examining the processes of note production and note-reception with a digital pen and a pair of eye-tracking glasses. Her findings support a series of conventional note-taking guidelines such as upholding the principle of economical note-taking (e.g., Szabó, 2006), but also throw doubt on some traditional note-taking guidelines. For example, in an L2-to-L1 interpreting task where the cognitive demand for listening comprehension is considered heavier than that in the other interpreting direction, interpreters rush to take notes and result in poorer interpreting accuracy (S. Chen, 2020a). This approach to note-taking is not in accordance with conventional note-taking guidelines (e.g., Rozan, 1956; Seleskovitch, 1975) where listening comprehension is prioritized over note-taking to ensure a meaning-based rendering. The observed gaps between what has been proposed in theory and what has been followed in practice suggest that probing into the process of note production and note reception in interpreting practice is important in exploring the bases of successful note-taking.

Note-taking can be even more demanding when it is involved in VRI or Video Mediated Interpreting (VMI) (Braun & Taylor, 2012; Napier et al., 2018), where interpreters have to distribute their limited visual attention and cognitive resources to watching the computer screen and processing the notes simultaneously. Although the interpreting community has shown “considerable skepticism” (Roziner & Shlesinger, 2010, p. 216) toward such remote modes of interpreting because they can induce “a number of physiological (sore eyes, back and neck pain, headaches, nausea) and psychological complaints (loss of concentration and motivation, feeling of alienation)” (Mouzourakis, 2006, p. 52) to interpreters, they have become the new norm for the interpreting industry and interpreter training programs in the post-pandemic era. According to the 2019-2020 European Language Survey, remote interpreting (RI) has become one of the language services that create the most revenue for language service companies, and it is an irreversible trend in the market (see European Commission, 2020). The International Association of Conference Interpreters (AIIC) has already published an interpreter checklist to help interpreters perform interpreting at distance (AIIC, 2020). At the same time, online interpreter training programs have also been introduced in schools around the world (e.g., Ko & Chen, 2011; Zhang & Ding, 2021). In light of the increasing popularity of RI in the interpreting market and the education field, professional interpreters and student interpreters have to adapt to such a remote way of working and training accordingly.

However, insofar as RI is concerned, only a small fraction of existing studies have been conducted with empirical methods (Ko, 2006), and most of these have focused on SI (e.g., Moser-Mercer, 2005; Roziner & Shlesinger, 2010; Korpal & Jasielska, 2019). By comparison, very little attention has been paid to CI, let alone note-taking. Nevertheless, recent research has demonstrated that it is the extra cognitive demand brought by note-taking that makes CI more cognitively demanding than SI in the remote mode of interpreting (Doherty et al., 2022). Therefore, finding out “how to reduce processing capacity and time requirements of note-taking while maintaining the efficiency of notes” (Gile, 1995/2009, p. 178) in the remote mode of CI is an essential step for professional interpreters and student interpreters to adapt to this increasingly popular mode of interpreting.

In CI, note-taking starts from the reception of the source speech and features high individuality among interpreters (e.g., Albl-Mikasa, 2008). Therefore, in the exploration of note-taking in CI, two important factors should be considered: the source material of note-taking, and the characteristics of the note-taker. For instance, when the source speech is difficult to interpret, interpreters can decrease their note-taking effort to make more processing capacity available to comprehend the source speech; or they can increase the effort to maintain note quantity and note quality. When the source speech is easy to interpret, interpreters can take a large number of notes to ensure an accurate and detailed interpretation, or they might note nothing and simply rely on their short-term memory. On the other hand, interpreters with different levels of interpreting expertise and note-taking skills can adopt different approaches to note-taking in the same interpreting task (e.g., Abuín González, 2012; S. Chen, 2022; Szabó, 2006). In S. Chen (2022), it was found that the professional interpreters can follow the source speech more tightly in note production to generate more notes than the student interpreters, thus leading to better interpreting performance in the former group than in the latter one. Hence, both the source material of note-taking and the producer of the notes are important variables that have to be considered while investigating the issues of note-taking in CI. In respect to note-taking, only very limited empirical studies (Cardoen, 2018; Hu, 2008) have considered the features of the source material and those of the note-taker. Furthermore, none of these studies have been conducted in a remote setting of interpreting, leaving the issue of how to take notes effectively in complex VRI scenarios underexplored.

Noting these issues, the present research attempts to revisit note-taking in CI and address some observed limitations by: 1) including both experienced professional interpreters and inexperienced student interpreters as participants; 2) using both easy and difficult speech segments as the source material of interpreting; 3) examining interpreters' note-taking behaviour in a stimulated VRI environment; and 4) combining process-oriented methods such as eye-tracking and pen-recording with product-oriented approaches such as the analysis of note features and interpretation quality, to provide a more comprehensive account of note-taking in interpreting.

1.2 Methodology

The present study was conducted with a between-subject and within-subject design. The effects of interpreter work experience on interpreters' note-taking behaviour were evaluated based on the data collected from two groups of participants, 20 professional interpreters and 29 student interpreters, and the effects of source speech difficulty on interpreters' note-taking behaviour were investigated by using both easy and difficult interpreting materials. Process-oriented methods including eye-tracking and pen-recording were adopted to measure the interpreters' note-taking effort and note-reading effort from visual, cognitive, physical and temporal perspectives. Product-oriented methods such as a descriptive analysis of the notes and a summative assessment of interpretation quality were used to analyse the interpreters' note choices and interpreting performance. In addition to that, a subjective scale called the NASA Task Load Index (NASA-TLX) was used to measure the subjective difficulty that the interpreters perceived during the interpreting task, and a questionnaire concerning their attitude towards note-taking in CI and basic demographic information was included for the sake of providing potential clues for the observed research findings.

The statistical analysis of the collected data was conducted mainly through SPSS Statistics, a software developed by the International Business Machines Corporation (IBM). For between-subject comparisons, independent *t*-tests were conducted for normally distributed data, and Mann-Whitney tests were adopted for non-normally distributed data. For within-subject comparisons, paired *t*-tests were used for normally distributed data, and the Wilcoxon signed rank test was adopted for non-normally distributed data. In addition, Pearson's and Spearman's correlation tests were conducted to examine the relationships between the note-taking and note-reading effort as well as the associations between interpreters' note-taking behaviour and interpretation quality.

1.3 Aims, research questions and hypotheses

In light of the significance and lack of investigation on interpreters' note-taking behaviour, four specific aims are addressed in the present study to fill the research gaps:

> *Aim 1*: to explore the effects of interpreter work experience and source speech difficulty on the process and product of note-taking, which includes the visual, cognitive, physical and temporal aspects of the note-taking effort and the distribution of note quantity, note form, note language and note-taking strategy in the note-taking product.

> *Aim 2*: to examine the process of note-reading by visualizing the interpreters' fixation patterns on the notes with the help of heat maps, gaze plots and event logs, and by quantifying the interpreters' note-reading effort with a range of fixation measures that point to both the early and late stages of note processing.

> *Aim 3*: to investigate the relationship between the effort of note-taking and the effort of note-reading, with the former being measured from four aspects (visual,

cognitive, physical and temporal), and the latter being quantified with eye-tracking measures that reflect the cognitive effort of different stages of note processing.

Aim 4: to explore the associations between the interpreters' note-taking behaviour on the one hand, namely their note-taking effort, note-taking product and note-reading effort, and the interpretation quality on the other hand, which includes information completeness, fluency of delivery and TL quality.

To pursue these four aims, five research questions (RQ) and eleven corresponding hypotheses (H) that cover each stage of note activities in CI are formulated. The questions start from the note-taking process and extend to the product of note-taking as well as the process of note-reading. After each of these note activities is examined individually, this study further explores the complex relationships among interpreters' note-taking behaviour, note-reading patterns, and interpretation quality, for the purpose of providing a full account of note-taking in CI from source speech comprehension to target speech delivery.

The first research question focuses on the process of note-taking, and it was raised under the background that interpreter work experience and source speech difficulty had been rarely examined in note-taking studies. Taking the interpreters' work experience and the materials' difficulty level as two independent variables, the present study explores this issue with a series of eye-tracking and pen-recording measures. The specific research question and associated hypothesis are formulated as follows:

RQ 1: During the input phase of CI, how would interpreter work experience and source speech difficulty affect interpreters' note-taking process?

H1: Between easy and difficult speech segments, as the cognitive demand for listening comprehension increases in the difficult segments, this could limit the availability of cognitive resources for note-taking. Thus, the interpreters are expected to expend less effort on note-taking in the difficult speech segments than in the easy ones. At the same time, compared with the students, the professionals who are assumed to be less reliant on the notes during interpreting would expend less effort on note-taking, regardless of source speech difficulty.

It is worth mentioning that "source speech difficulty" is investigated from two perspectives in this study: one suggesting the difficulty level of the speech segments (easy vs. difficult); and the other referring to the presentation sequence of the speech segments at different levels of interpreting difficulty (from-easy-to-difficult vs. from-difficult-to-easy). Therefore, for the sequence effects of source speech difficulty on the interpreters' note-taking behaviour, there is another hypothesis for RQ1:

> H2: For the speech segments that are presented in an easy-to-difficult order, the interpreters are expected to reduce their note-taking effort in the latter segment; while for the two segments arranged in a difficult-to-easy order, they would maintain their note-taking effort at a similar level because of the potential distorted perception of task difficulty. As student interpreters could be more sensitive to the change in source speech difficulty, such sequence effects of task difficulty on the interpreters' note-taking effort are expected to be more obvious in the student group than in the professional one.

Since the present study is framed under a VRI scenario where interpreters deliver interpretations through a computer screen, the hypotheses for RQ1 are examined through two methods. The first method is to analyse the proportions of the overt visual attention and cognitive resources that the participants allocate to the different areas on the computer screen. Results obtained from this method illustrate how the participants distribute their limited processing capacity to watching the video and producing the notes during the input stage of CI. The second method is to examine the hypotheses by using the original eye-tracking and pen-recording values obtained from the Areas of Interest (AOIs) drawn on each individual note. To put it simply, the first method focuses on the proportioned effort that the interpreters devote to completing the note-taking task, whereas the second method investigates the average effort that the interpreters expend on taking one note during the input stage of CI.

Secondly, the effects of interpreter work experience and source speech difficulty on the process of note-taking can very possibly affect the interpreters' product of note-taking. The second RQ examines the experience and difficulty effects on note-taking from a product perspective, with both within- and between-group comparisons being conducted to investigate the interpreters' note quantity, note form, note language and note-taking strategy in different task conditions. One research question with two associated hypotheses are formulated on this facet of the investigation:

> RQ2: With the experience and difficulty effects observed on the process of note-taking, how would the two independent variables, namely interpreter work experience and source speech difficulty, affect interpreters' note-taking product?
>
> > H3: As source speech difficulty increases and interpreter work experience decreases, the participants would deprioritize note-taking to make more processing capacity available for listening comprehension. As a result, note quantity would decrease and the interpreters would opt for note forms, note languages and note-taking strategies that require less note-taking effort.
> >
> > H4: In the easy-to-difficult direction of interpreting segment sequence, the interpreters would opt for more effort-saving note choices in the difficult segment than in the easy one. In the reverse interpreting direction, the interpreters could maintain their note preferences. The students are expected to show more obvious changes in their choices of notes in the easy-to-difficult direction of interpreting difficulty level as compared with the professionals.

Thirdly, the interpreters' choices of notes in the first stage of CI can directly decide their cognitive load of note-reading in the second stage of CI (e.g., S. Chen, 2020a). Meanwhile, as note-reading competes with target speech production for interpreters' limited cognitive resources during the reformulation phase of CI, inspecting how interpreters read back their notes is very important for researchers to understand the reasons behind the observed interpretation quality. Therefore, a research question and associated hypotheses are drawn up as follows:

> RQ3: During the output phase of CI, how do interpreters read back their notes? Is note-reading a cognitively demanding task?

H5: As interpreters usually take notes in groups based on the meaning units in the source speech (e.g., Albl-Mikasa, 2008; S. Chen, 2022), they are expected to read notes in groups during target speech production.

H6: Note-reading could be very cognitively demanding considering that the notes are highly condensed and varied in forms and languages.

Fourthly, exploring the relationship between the two stages of the note activities can offer practical implications for interpreting practice and interpreter training. Based on the Levels of Processing Hypothesis (Craik & Lockhart, 1972), a deeper level of language processing during the process of note-taking would lead to easier retrieval of memory during the process of note-reading. Hence, one research question and associated hypotheses are proposed for this aspect of investigation:

RQ4: Is there a trade-off between the note-taking effort at the input phase of CI and the note-reading effort at the output phase of CI?

H7: A trade-off is expected to be found between the overall effort of note-taking and that of note-reading.

H8: This trade-off between the two types of noting effort is also expected to be observed in the comparisons of different note forms (language notes vs. symbols; full words vs. abbreviations) and note languages (Chinese vs. English).

Last but not least, although note-taking has long been regarded as an essential factor of CI quality, in existing note-taking studies, there are no widely accepted conclusions about how note-taking could enhance or decrease interpretation quality. Since inspecting the role of note-taking from a "performance" perspective in CI can help interpreters to conduct note-taking more effectively, a related question and associated hypotheses are proposed as follows:

RQ5: What is the relationship between interpreters' note-taking behaviour and interpretation quality?

H9: Increased effort in note-taking would be associated with better interpreting performance because of a potentially deeper level of SL processing.

H10: A larger quantity of notes, a larger proportion of notes in symbols (compared with language notes), and a higher percentage of notes in TL (compared with SL) could lead to better interpretation quality.

H11: Increased note-reading effort would be correlated with poorer interpretation quality, as the processing capacity left for target speech production could be limited.

Figure 1-1 displays the research questions and the variables that are investigated in this study.

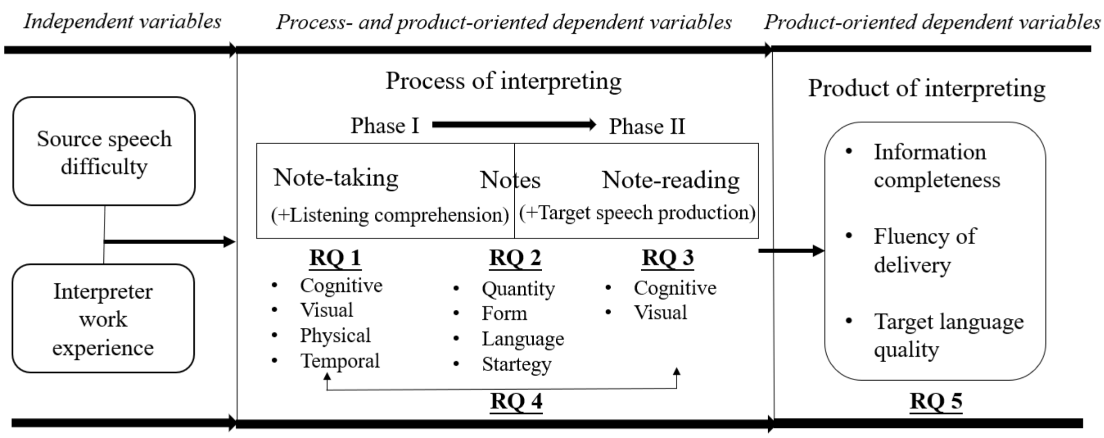


Figure 1-1. Scheme of the independent and dependent variables of the study

1.4 Structure of the thesis

This thesis is comprised of eight chapters:

Chapter 1 gives an overview of the study by introducing the research background, methodology, research aims, questions and hypotheses.

Chapter 2 presents a comprehensive review of note-taking studies in the field of CI. Specifically, theoretical discussions about the nature of note-taking in interpreting are introduced in Section 2.1, and empirical studies on the process of note-taking, the product of note-taking and the process of note-reading are examined in Section 2.2.

Special attention has been paid to previous investigations of the two independent variables in the present study: interpreter work experience, and source speech difficulty.

Chapter 3 continues to dig into the literature in both interpreting and neighbouring disciplines, which can provide insights about the cognitive processing involved from the processes of note-taking to the product of interpreting. It firstly introduces factors that can contribute to the cognitive load of note-taking (Section 3.1), then explains the types of resources that are demanded in notes activities (Section 3.2), and finally investigates the potential associations among note production, note reception and interpretation quality (Section 3.3). Altogether, the reviewed literature depicts how interpreters' note-taking behaviour can be shaped by task-, environment- and interpreter-related characteristics and how such behaviour affects the cognitive load of interpreting and interpretation quality.

Chapter 4 introduces the research methodology, including the selection and manipulation of the source speech (Section 4.1), the applications of eye-tracking and pen-recording methods in this study (Sections 4.2 and 4.3), and the analysis of note patterns and interpretation quality (Section 4.4).

Chapter 5 provides detailed information about the design, setting and procedures of the experiments (Sections 5.1-5.3), as well as the methods adopted in data quality assessment (Section 5.4) and statistical analysis (Section 5.5).

Chapter 6 firstly reports on the participants' perceived difficulty during interpreting (Section 6.1) and then provides all the results of the present study's five research questions. Sections 6.2, 6.3 and 6.4, respectively, report on the data concerning the process of note-taking, the product of note-taking and the process of note-reading. The relationship between the effort devoted to the production phase and the deciphering phase of note-taking is illustrated in Section 6.5, and the associations between the participants' note-taking behaviour and interpreting performance are reported in Section 6.6.

Chapter 7 discusses the major results of the present study by comparing them with findings in previous note-taking studies and applying them to relevant theories and models. In accordance with the four research aims, discussions include the observed effects of interpreter work experience and source speech difficulty on the process and product of note-taking (Section 7.1), the special reading patterns and cognitive demands of note-reading (Section 7.2), the relationship between the note-taking effort and the note-reading effort (Section 7.3), and the role of note-taking in interpreting from the perspective of interpreting performance (Section 7.4).

Chapter 8 concludes the thesis with a summary of the findings (Section 8.1), a discussion of the academic and pragmatic significance of the present study (Section 8.2), an introduction to the strengths and limitations of the present investigation (Section 8.3), and some suggestions for possible directions for future research (Section 8.4).

Chapter 2: Note-taking in Consecutive Interpreting

Note-taking guidelines have been well developed for a variety of language pairs (e.g., Becker, 1972; Gillies, 2017; Gran, 1982; Ilg, 1988; Matyssek, 1989; Wu, 2008). However, interpreting practitioners are always stuck in the application of these guidelines (e.g., Alexieva, 1994; Arumí Ribas, 2012; Her, 2001). This notable gap between what has been proposed in theory and what has been followed in practice highlights the importance of digging into the cognitive mechanisms underlying interpreters' note-taking behaviour. This chapter aims to provide a comprehensive review of previous theoretical discussions and empirical studies on the complex cognitive and physical attributes of note activities in CI. Specifically, Section 2.1 focuses on the theories that explore the nature of note-taking in CI, and Section 2.2 reviews the empirical studies that probe into the process of note-taking, the product of note-taking and the process of note-reading in interpreting practice. Section 2.3 provides a summary of the findings and methodologies of previous studies, during which interpreter work experience and source speech difficulty are identified as important potential factors of interpreters' note-taking behaviour in CI.

2.1 Note-taking in nature

This section is comprised of two parts, with the first part introducing how note-taking is perceived as storage of "sense" that is extracted from the source speech, and the second touching on how note-taking is perceived as storage of "language" which is closely related to the linguistic expressions in the source speech. Taken together, this section reviews previous theoretical discussions about the nature of note-taking from cognitive and linguistic perspectives.

2.1.1 Noting the idea

The theory of sense, which is widely discussed in note-taking studies, was proposed by Danica Seleskovitch based on her findings about 12 French-native professional interpreters' notes in an L2-to-L1 CI task:

> 1) few of the words in the source speech appear in the notes (on the sample passage, only one word, family, is found in all three sets);
> 2) the notes include many words not found in the source speech (structure, system, children);
> 3) the renditions express much more than is in the notes;

> 4) some items appear in three different forms, e.g. 'Similarly...' in the source speech is noted as '2nd', and rendered as 'par exemple...' (as cited in Setton, 2002, p. 119, Seleskovitch, 1975).

In summary, the expressions in the source speech, the composition of the notes, and the expressions in the target speech were detached from each other on the linguistic surface. At the same time, the interpreters showed obvious individual differences in the choice of note forms and note languages. Without knowing the corresponding mental activities, it would not be possible to understand the interpreters' notes. Based on these findings, Seleskovitch (1975) proposed the theory of sense, which is also referred to as the deverbalization theory of note-taking, claiming that note-taking was only linked with the source and target speeches in CI at a semantic level but not at a linguistic one. Specifically, during the input phase of CI, interpreters decode the linguistic expressions in the source speech to extract the "sense". Then, they record this mental representation of the source speech through notes without relying on the SL or the TL of the interpreting task. In the output phase of CI, they retrieve the "sense" in note-reading and re-deliver this to the audience with the linguistic tool of the TL. During this process, note-taking situates at a deverbalized stage where only the abstract sense, rather than concrete words, is stored in the notes. Seleskovitch (1975, p. 5) observed that "notes are not a third, interlingual text but syncretic reminders of the experience of (source speech) comprehension; and that interpretation is 'not transcoding, but the result of two people's thoughts".

Seleskovitch's theory of sense provided a solid theoretical foundation for the most widely-advocated principle of note-taking, "(n)oting the idea and not the word" (Rozan, 1956, p. 15), which regards source speech comprehension as a prerequisite of note-taking in interpreting. According to Albl-Mikasa (2008, p. 208), such "calls for dissociating sense from language, for concentrating on the conceptual content or essence and for taking notes on the macropropositional level" can be attributed to "the ideal of meaning-based interpreting" which is assumed to be the standard strategy of interpreting (e.g., Dam, 2001). It appears that researchers naturally extend this meaning-based approach from interpreting to note-taking (Albl-Mikasa, 2008). Notes are therefore assumed to be storage of sense rather than storage of words. If meaning-based interpreting and note-taking are regarded as the standard practice of interpreting, then the question remains: how to extract the pure "sense" from the "language" in the source speech? Seleskovitch (1975) does not elucidate this issue, whereas Mackintosh (1985) fills this gap by referring to studies on text comprehension and production from the perspective of propositions.

A proposition is a set of word concepts that are grouped together based on certain rules, and serves as a basic element of the semantic structures in a text (Kintsch, 1974). Mackintosh (1985) explains the process of interpreting by referring to two forms of propositions, micropropositions and macropropositions, which are generated through two types of language processing: micropropositional processing (or local level and propositional textbase processing); and macropropositional processing (or global level and mental representation processing). Albl-Mikasa (2008) provides a clear explanation of how these concepts from studies on text comprehension and production are applied in note-taking contexts:

> …text comprehension involves building multi-level representations of the text: on a lower level the representation of the (lexical and syntactical) surface structures and a propositional text base (explicit text propositions plus local-level inferences) and on a superordinate, more global level a situation or mental model (see van Dijk & Kintsch 1983; Johnson-Laird 1983; Schnotz 1994)…On the subordinate level of surface representation, language structures are maintained; on the level of the propositional textbase, information is represented in a conceptual way but closely reflecting the text; on the superordinate level the mental representation models the situation described by the text rather than the text itself and is therefore much less text-specific. (p. 203)

In Mackintosh's (1985) view, during the input stage of CI, interpreters should first identify the micropropositions in the source speech and then apply appropriate macro-rules to transform the micropropositions into macro-ones. Only when all the macropropositions of the source speech are connected in a coherent way can interpreters form a holistic understanding of the source speech and produce corresponding notes. Such a transformation from micropropositions into macropropositions is realized through three macro-rules: 1) *deletion*: removing a proposition if it is not helpful in decoding the next proposition; 2) *construction*: creating a new proposition to indicate the global fact of several propositions; and 3) *generalization*: using a superordinate proposition to cover a group of subordinate propositions. Interpreters should apply these three macro-rules repeatedly to create the macro-level propositions of the source speech, in order to achieve a global-level understanding of the source text. When it comes to the output phase of CI, interpreters should apply the reverse rules, i.e., *addition*, *specification* and *particularization*, to decipher the macropropositions in notes and transform them into micropropositions that can be re-delivered through the TL to produce the target speech.

In accordance with Seleskovitch (1975), Mackintosh (1985) upholds the view that notes should be created based on a thorough understanding of the source speech, and that they only function as reminders of the logical relations among the formed macropropositions. However, it is important to bear in mind that the retrieval and application of these macro-rules which are stored in interpreters' long-term memory (LTM) can cause extra cognitive load on their WM. As suggested by Gile (1999), interpreters can often work close to saturation. If they apply these macro-rules for several rounds in order to transform all the micropropositions into a macroproposition, note-taking could be delayed significantly and incoming information could be missed. Furthermore, as explained by Mackintosh (1985), the fundamental purpose of these macro-rules is to help interpreters to integrate the last, the present and the incoming propositions so as to comprehend the source speech at a global level. This means that, while interpreters are processing the current proposition, they will still have to keep the preceding and proceeding propositions in mind. Under this situation, their WM can be fully occupied in a short time. In addition to the concerns mentioned above, criticisms of Mackintosh's (1985) model of interpreting also extend to the purpose of interpreting. Albl-Mikasa (2008) argues that generating a thorough understanding of the source speech is not the ultimate goal of interpreters. Instead, interpreters should focus on transmitting the source information to the audience in an accurate and detailed way. Here, interpreting accuracy not only concerns the word choices in the source speech but also the style of the speaker. From this perspective, compared with recording the abstract sense of the source speech with notes, following closely the microstructures in the source speech is a more effective note-taking strategy. This proposal of a micropropositional approach toward note-taking in interpreting, which can be traced back to Kirchhoff (1979), forms another major argument about the nature of note-taking.

2.1.2 Noting the word

In contrast to Danica Seleskovitch who regards notes as the externalization of sense, Hella Kirchhoff perceives note-taking as "a primarily linguistic process, based on the microstructures of the source text" (as cited in S. Chen, 2016, p. 158, Kirchhoff, 1979). In Kirchhoff (1979), she finds that interpreters followed very closely the surface structures in the source speech to conduct note-taking. Therefore, she claims that "(n)oting the idea and not the word" (Rozan, 1956, p. 15) might be the ideal approach to note-taking but is difficult to follow in practice. Such a linguistic view of note-taking in CI is not in accordance with the conventional view that notes should be created based on a thorough understanding of the source speech. Neither is it highly promoted in the majority of the note-taking guidelines and textbooks. Despite this, Albl-Mikasa (2008) supports and extends Kirchhoff's (1979) claim about the importance of micropropositional processing in note-taking by establishing a relevant theoretical framework based on theories borrowed from text and language processing.

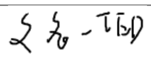
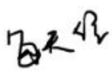
Albl-Mikasa's (2008) theoretical framework of note-taking is based on the notion that notes can be regarded as an independent text in its own right. Compared with the source and target texts in CI, the notation text has two distinctive features. Firstly, the source and target texts are delivered and received by different parties, namely the speaker, the interpreter and the audience. However, the notation text is produced and received by the same person, i.e., the interpreter herself. From this perspective, note-taking is an intrapersonal communication process between the interpreter at the input stage of CI and the same interpreter at the output one. Secondly, the source and target texts in CI are comprised of complete sentences and full words, whereas the notation text is comprised of a special notation language that has "highly reduced or even fragmentary" (p. 211) surfaces and "pictographic and iconic signs" (p. 211) in non-linear structures. This means that, during the process of note-reading, interpreters first have to recognize the notes that are varied in forms and languages, and then recover the complete source information based on the limited number of notes. Based on these two distinctive features of the notation text, the key to taking good notes in CI lies in achieving effective intrapersonal communication across the two stages of note activities. According to Wilson and Sperber (2004), the basic principle of effective communication is that all interlocutors in a conversation should abide by the Relevance Theory. Specifically, by inferring the abundant implicature behind the very limited explicature provided by the last interlocutor, the next interlocutor could make correspondent responses to the previous utterance without much processing effort. Applying this theory to note-taking, this means that interpreters should be able to maximize the referential effects of the limited notes with a minimum amount of effort.

In this way, they retrieve a great amount of implicature (i.e., source information) from the limited explicature (i.e., the notes). In summary, interpreters should "find the optimal balance between noted (i.e., explicit) and memorised (i.e., implicit) information" (p. 264) in order to facilitate memory retrieval without much note-taking and note-reading effort.

To find out how interpreters realize such "solipsistic" (p. 211) communication across note-taking and note-reading stages, Albl-Mikasa (2006) analyses five student interpreters' note-taking strategies in a CI task. Among the three identified strategies, the least effortful one is the ellipsis strategy, which refers to "an omission of source text units and a transfer of selected, often central content words from the source text into the notation text" (p. 3). Notes created in this way look like a shorthand for the linguistic expressions in the source speech. The second strategy is restructuring, which involves "substituting non-source text structures for source text structures" (p. 3) during note-taking. Compared with the ellipsis strategy, the restructuring strategy requires more cognitive effort from interpreters because of the involvement of syntactical processing, and often leads to simplified source structures in the notation text and in the target text. The third strategy is high condensation, which is a combinational use of the previous two strategies. This refers to the situation where interpreters use only one or several notes to represent "source text clauses, sentences or even whole passages" (p. 4). When adopting this strategy, interpreters have to greatly condense the contents in the source speech and greatly reduce syntactic structures in the source text. This requires them to fully comprehend the source speech before note-taking and rely more on their own memory rather than on the notes to recall the source information during target speech production. Table 2-1 provides specific examples for the three note-taking strategies with notes collected in the present study. Overall, following the order of the strategies introduced above (ellipsis, restructuring and high condensation), there is a decrease in note quantity, an increase of detachment from the source surface structures in notes, and an increase of cognitive effort involved in strategy adoption (Albl-Mikasa, 2006).

Table 2-1. Examples of the ellipsis, restructuring and high-condensation strategies of note-taking

Strategy	Source speech transcript	Notes	Explanations

Ellipsis	…because my father's name is Ted.		• 父: abbreviation of "father" (父亲) in Chinese • 名: abbreviation of "name" (名字) in Chinese • Ted: full name for "Ted"
Restructuring	…what happens every day.		• 每天："every day" in Chinese • 经: abbreviation for "happens" ("经过") in Chinese
High condensation	How was that passed on?		

After comparing the source text, the students' notation texts and the target texts, Albl-Mikasa (2006) found that the three texts stayed very close to each other at their linguistic surface; and the students' notes were dominated by the ellipsis strategy, a strategy that only involves a shallow level of language processing. This finding indicates that the students followed a micropropositional approach while producing and receiving the notation text. The same phenomenon has been observed among professional interpreters. In S. Chen (2020b), she adopted an indicator called ear-pen span (EPS), the time lag between the offset of source speech delivery and the onset of note-taking acts, to measure the temporal demand of note-taking. Since EPS can only be calculated for "note units that have a one-to-one correspondence with the source speech" (p. 126), the report that around 90% of the notes collected in S. Chen (2020b) had EPS data indicates that, for most of the time, the professionals followed the word choices in the source speech to produce their notes. All of these findings corroborate Kirchhoff's (1979) claim that, in interpreting practice, interpreters usually note in *word* rather than in *idea*. In the series of Albl-Mikasa's work (Albl-Mikasa, 2006, 2008; Kohn & Albl-Mikasa, 2002), she summarizes four reasons for interpreters' prevalence of this micropropositional processing approach during note-taking in CI. Firstly, this approach to note-taking is in accordance with the Relevance Theory in that it only requires a shallow level of language processing and a minimum amount of effort in note production. Therefore, interpreters can adopt this approach easily during note-taking. Secondly, notes created in this way usually are featured in large quantities, thus providing a great number of memory cues for interpreters to retrieve the source information during the output phase of CI. Thirdly, micropropositional processing of the source text is an integral procedure of CI that cannot be avoided. On the one hand, no matter how thoroughly interpreters comprehend the source text, such a global level understanding of the source text is built on the interpreters' processing of the SL at a local level. On the other hand, even though the interpreter has successfully extracted the pure "sense" from the "language" in the source text, she still needs to "revert (it) to a propositional representation for re-textualisation" (Albl-Mikasa, 2008, p. 225). Otherwise, the abstract sense cannot delivered to the audience through concrete words. Fourthly, Albl-Mikasa (2008) emphasizes that the interpreter's duty is to transmit the most detailed and accurate source information to the audience. Therefore, a notation text that follows the micropropositions in the source text should be beneficial in ensuring the accuracy and completeness of the target text. By comparison, working on the macropropositions in the source text, which requires a more thorough understanding of the text, is more suitable for those who intend to learn from the text and store the learning outcomes in their LTM. Based on these four reasons, Albl-Mikasa (2008, p. 225) supports that interpreters should adopt "(micro-)propositional processing and hence a somewhat form-based attitude to note-taking" to alleviate their pressure on the short-term memory and maintain a detailed interpretation of the source speech.

In summary, there are two major arguments about the nature of note-taking in CI: notes as the storage of sense which is extracted from the source speech; and notes as the storage of linguistic expressions in the source text. The former requires interpreters to generate a semantic whole of the source speech based on the formation of macropropositions, whereas the latter only involves micropropositional processing of language during note-taking. To understand how these cognitive and linguistic aspects of note-taking are reflected in interpreting practice, Section 2.2 systematically reviews 23 empirical studies that have explored interpreters' note-taking behaviour in three stages: the process of note-taking, the product of note-taking and the process of note-reading.

2.2 Note-taking in practice

With "note", "notation" and "note-taking" as the search keywords, 33 empirical studies are identified in the present study from the Conference Interpreting Research Information Network Bulletin (CIRIN Bulletin 2010-2022), Translation Studies Bibliography (TSB), and the Chinese Social Sciences Citation Index (CSSCI), and 13 more articles are found based on the references of the retrieved items. Among these 46 pieces of work, 16 will not be discussed in detail because the topics are not closely related to that of the present study. These latter items mainly focus on the pedagogy of note-taking (Chuang, 2008; Duman, 2014; X. Han, 2013; Hradilová, 2019; Hui, 2019; Lee & Choi, 2012; Lung, 1999, 2003; Orlando, 2014, 2015; Yamada, 2018), the necessity of note-taking in CI (Gile, 1991; Campos et al., 2017), the application of note-taking in written-translation training (Sakamoto, 2011), the use of note-taking in English listening tests (Kim, 2009), and the association between the interpreters' personality and note-taking style (Volpe, 2015). Another seven studies (Błaszczyk & Hanusiak, 2010; Chang, 2015; S. Chen, 2018; Kohn & Albl-Mikasa, 2002; Someya, 2017; Winkler, 2015; Y. Zhao, 2022) are excluded from the present review because they provide little product data, i.e., notes, or process data such as the interpreter's pen movements during note-taking, all of which are essential for the present study to generate hypotheses. Ultimately, there are 23 selected items remaining which are introduced in this section in a thematic and chronological order as follows.

2.2.1 Note-taking process

Among the 23 selected studies, only 6 probed into the process of note-taking (presented chronologically in Table 2-2). Two studies investigated the difficulties that interpreters experienced during note-taking through retrospective interviews and questionnaires, while four presented quantitative data that were obtained through video-recording, dual-task paradigm and pen-recording methods. In this section, special attention is paid to these latter four quantitative studies, as the methodologies and findings reported in them are essential references for the selection of research methods in relation to the present study. These studies will be introduced in this section according to their adopted methodologies: video-recording (Section 2.2.1.1), dual-task paradigm (Section 2.2.1.2), and pen-recording (Section 2.2.1.3).

Table 2-2. Studies on the process of note-taking[1]

Literature	Research focus	Language pair	Interpreting direction	Research method	Participant type
Andres (2002)	Cognitive effort	French -German	L2-L1	Video-recording	14 professionals and 14 students
Hu (2008)	Cognitive effort	Chinese -English	Both	Dual-task paradigm	10 professionals and 10 students
Xu & Chai (2008)	Note-taking difficulty	Chinese -English	L1-L2	Stimulated recall	6 professionally-trained students and 6 non-professionally-trained students
Abuín González (2012)	Note-taking difficulty	English -Spanish	L3-L1	Questionnaire	10 beginner students, 10 advanced

[1] Papers based on the same research project, such as S. Chen (2020a), S. Chen (2020b), and S. Chen (2022), are counted only once.

					students, and 10 professionals
S. Chen (2017b)	Cognitive, physical and temporal demands	Chinese -English	Both	Pen-recording	5 professionals
S. Chen (2020a, 2020b, 2022)	As above	Chinese -English	Both	Pen-recording	18 professionals for 2020a, 4 more professionals for 2020b, 22 more beginner students for 2022

2.2.1.1 Video-recording

Dörte Andres made the first attempt to examine the process of 14 professionals' and 14 students' note-taking in an L2-L1 CI task. Through recording the note-taking scenes, Andres (2002) manually identified the delivery onset of each word in the source speech and the onset of each note-taking act in the video (as shown in Figure 2-1) to calculate interpreters' EPS. She found that the professionals' EPS was four to six seconds, while that of the students sometimes extended to ten seconds. Since Andres detected that the interpreters generally had listening comprehension problems when their EPS was longer than seven seconds, she suggested that the prolonged EPS among the students indicated a cognitive overload of note-taking. In addition, she observed that, for the student group, the part of the source speech that made them pause during note-taking at the input stage of CI also impeded them from note-reading and target speech production during the output. By contrast, this phenomenon was not observed in the professional group. These observed professional-novice differences demonstrate that the two groups of interpreters varied in the cognitive load they experienced during the process of note-taking.

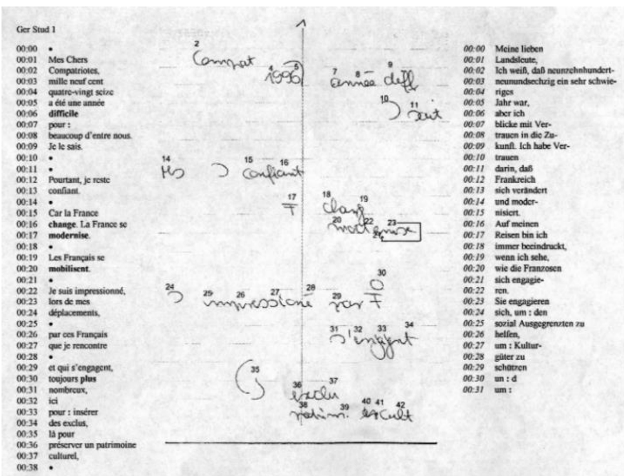


Figure 2-1. The identification of the onsets of source speech delivery and note-taking

acts in Andres (2002)

With innovative methodologies and inspiring findings, Andres (2002) greatly expanded the research scope in note-taking studies. This was the first time that the process of note-taking was visualized and analysed with quantitative methods. It was also the first piece of empirical evidence that proves the important role of interpreter work experience in affecting the process of note-taking. Nevertheless, concerns have been raised about Andres' calculation and interpretation of the EPS data. In terms of calculation, S. Chen (2017b, 2020a, 2020b, 2022), who also adopted EPS as an indicator of the note-taking effort, pointed out that Andres' (2002) result could be imprecise for three reasons. Firstly, Andres did the calculations in seconds rather than in milliseconds, which means that interpreters' differences in EPS at a millisecond level would be obscured. Since note-taking is conducted under extreme time limits in CI, such subtle differences in milliseconds can contain abundant implications about the process of note-taking. Secondly, all the timepoints in Andres (2002) were identified manually. Although they must have been checked carefully, there still could be human error in the reported data. Thirdly, Andres calculated the interpreters' EPS from the onset of each word in the source speech rather the offset of each word. As the duration of EPS could be heavily affected by the length of the source speech unit, Andres' method of measurement might miss crucial aspects of the data. For instance, the EPS of the note for the word "opportunity" is very likely to be longer than that for "ops" because it takes a longer time for the speaker to say the first word than the second.

Except for these three issues raised by S. Chen, another problem with using EPS to indicate the cognitive effort of note-taking is that it can only be calculated for notes that have corresponding units in the source speech. Notes that are improvised with no reference to the surface structures in the source speech, which usually involve more cognitive processing than those trackable notes, are not included in the EPS data. Hence, EPS cannot provide a full picture of interpreters' note-taking behaviour in CI. Finally, the length of EPS cannot be directly mapped onto the cognitive effort of note-taking. In Andres (2002), EPS was selected to be the indicator of the cognitive effort of note-taking because she found that the participants had comprehension problems when their EPS exceeded seven seconds. However, in essence, EPS only reflects how interpreters distribute their time between note-planning and note-writing. A shorter EPS can entail more notes, while a longer EPS leaves more time for listening comprehension.

Therefore, solely relying on EPS to measure interpreters' cognitive effort of note-taking is not a reliable approach. The cognitive implications behind the EPS data should only be inferred when other reliable measures of cognitive effort are available.

2.2.1.2 Dual-task paradigm

After Andres (2002), Hu (2008) continued to examine the effects of interpreter work experience, source speech difficulty and interpreter training experience on the cognitive effort of note-taking, through a methodology called the dual-task paradigm, borrowed from psychology. The dual-task paradigm measures the cognitive effort of performing a primary task by calculating the task performer's reaction time to a secondary task while he or she is performing the primary one. In Hu (2008), the participants were asked to respond to the random auditory probes sent by a computer program by pressing the left button of the mouse while they were taking notes for a source speech. In other words, the primary task was to conduct CI with note-taking and the secondary task was sound detection. Two groups of interpreters participated in the experiment: 10 professional interpreters with more than 3 years' interpreting-related work experience (at least 70 working days per year) and 10 student interpreters with 2 months' CI training. There were two types of interpreting materials: political talks that contained clichés the participants were familiar with; and speeches on topics such as steel production and the aging population that the participants were not familiar with. The latter situation was considered to be more difficult to interpret than the former one. The overall hypothesis was that the participants' cognitive effort of note-taking would decrease as their interpreter work experience increased and/or the difficulty level of the source speech decreased.

The first experiment simply focused on the effect of interpreter work experience on the two groups of participants' cognitive effort of note-taking in bidirectional CI tasks. Hu (2008) found that, in both directions of interpreting, the professionals' reaction time in the dual-task paradigm was shorter than that of the students, corroborating Andres' (2002) finding that the professionals bore less cognitive load during the process of note-taking than the students. In the second experiment, Hu (2008) added another independent variable to the exploration: source speech difficulty. The participants were asked to interpret talks with topics they were familiar with and not familiar with. The results show that, regardless of interpreting directionality, both groups of participants' reaction time increased in the difficult task condition.

Nonetheless, it is worth mentioning that the time difference was only significant in the professional group and not in the student group. The same group of students then participated in the third experiment which aimed at exploring the effect of interpreter training experience on the cognitive effort of note-taking. They completed another dual-task paradigm after receiving 48-hour training (in two months) on the same text type as the source speech that was used in the previous experiment. The results show that the students speeded up significantly in the second experiment, indicating a significant decrease in the cognitive effort of note-taking. Overall, the three experiments in Hu (2008) confirm the effects of interpreter work experience, source speech difficulty and interpreter training experience on the cognitive effort of note-taking in CI.

Despite the implicative findings reported in Hu (2008), there could be a series of concerns about the adoption of a dual-task paradigm in measuring the interpreters' cognitive effort of note-taking. First of all, the dual-task paradigm seriously broke the continuity of the note-taking process as it required interpreters to constantly switch between the primary and secondary tasks. Such interruptions could directly impede the interpreters' listening comprehension and note production, leading to small note quantities and poor interpreting performance. In addition, as the participants become familiar with the dual-task paradigm and figure out the purpose of the experiment, they might focus more on the completion of the secondary task rather than that of the primary one. Hu (2008) also pointed out these issues in her thesis, and explained that the dual-task paradigm had been shown to have "no deterioration of the quality of the primary task's final products" (p. 47). Although conducting a secondary task might slow down the participants' completion of the primary task, "the management of the task as a whole is not affected by the interruptions" (p. 48). However, none of the three pieces of literature that Hu (2008) referred to, Levy and Ransdell (1995), Piolat et al. (1999) and Piolat et al. (2001), put their focus on note-taking for interpreting purposes. Compared with general writing and note-taking tasks in the three referred studies, note-taking in interpreting can be much more cognitively demanding because it requires the interpreters to constantly transfer between the SL and TL of the CI task, and demands immediate recall after note production. It has been proven that noting in one's L2 is more effortful than in one's L1 (Barbier et al., 2006). Frequently switching between one's L1 and L2 during note-taking in interpreting can be even more demanding (Dong & Li, 2020). In addition, research on note-taking for general purposes has demonstrated that reviewing notes can help note-takers to achieve a deep level of text comprehension and a good performance in memory recall (Slotte & Lonka, 1999). However, consecutive interpreters do not have the time to review their notes and are required to deliver interpretations right after note-taking. These differences, among general writing tasks, note-taking for general issues and note-taking for interpreting purposes, in the involved cognitive processing imply that the methodologies adopted in these studies cannot be directly used interchangeably. In summary, the great intrusiveness of the dual-task paradigm can distort interpreters' note-taking process in CI. Therefore, researchers should find a more natural research method to measure the cognitive effort of note-taking without disturbing the interpreting process.

2.2.1.3 Pen-recording

Digital pens are widely adopted in designing industries where graphic designers have very high requirements about the naturalness of pen use and precise control of pen movement. Marc Orlando is the first researcher to apply this pen-recording technology to note-taking studies, for the purpose of improving note-taking training efficiency and developing a new hybrid mode of interpreting called Consec-simul with notes (see Orlando, 2010, 2014, 2015 for details). Later, Sijia Chen (2018) proposed that pen-recording, eye-tracking and voice-recording methods should be adopted at the same time to examine how interpreters write their notes, process the notes and produce the target speech in CI. This pen-eye-voice approach to the process of note-taking in CI is illustrated in Figure 2-2.

Figure 2-2. A pen-eye-voice approach to the investigation of the process of note-taking proposed by S. Chen (2018, p. 5)

As this figure shows, the participants would be asked to stand before an adjustable table and take notes on a Cintiq 13HD tablet. Their pen movements are recorded with a Wacom Pro Pen, and their eye movements are registered through a pair of SMI ETG eye-tracking glasses. The interpretations are automatically recorded by the eye tracker. A computer software called Eye and Pen is applied to control the experiment and synchronize the pen-recording and eye-tracking data. Ideally, with this experimental design, researchers are able to collect three streams of data from the participants: the distance, speed and duration of pen movements (i.e., handwriting); the duration and frequency of eye fixations on the notes; and the interpretation recordings. Overall, the pen-recording data point to the physical and temporal demands of note-writing, eye-tracking data indicate the cognitive load of note-taking and note-reading, and voice-recordings provide materials for interpreting performance evaluation. It is worth mentioning that, among the publications of S. Chen, most focus is put on pen-recording data (S. Chen, 2017a, 2020a, 2020b, 2022) while there are no descriptions regarding her use of the eye-tracking glass in exploring the process of note-taking. The two papers (S. Chen, 2020a; S. Chen et al. 2021) that touch on the eye-tracking data only present the results related to the process of note-reading. Therefore, while reviewing her work on the process of note-taking in this section, all results concerning the process-oriented data were generated from the pen-recording method, rather than the eye-tracking one.

S. Chen (2017b) measured the physical, temporal, and cognitive demands of taking notes in different forms and languages: the physical demand of note-writing was indicated by the distance of pen movement during handwriting; the temporal demand of note-taking was reflected by the duration and speed of handwriting; and the cognitive demand of note-taking was indicated by the interpreters' EPS. Based on five professional interpreters' data, she found that, regardless of how great the physical and temporal demands of note-writing were, the participants always preferred note forms and note languages that entailed the shortest EPS, i.e., the least cognitive demand of note-taking (Table 2-3). For instance, although symbols were easier to write than language notes, the participants opted to take more notes in language than in symbols. The same findings are reported in S. Chen (2020b) when the sample expanded to 22 professional interpreters. These results indicate that the cognitive demand for note-taking could be a decisive factor in the selection of note forms and note languages. However, as discussed in Section 2.2.1, solely relying on the EPS to measure the cognitive effort of note-taking can be problematic. Hence, other reliable measures of cognitive effort should also be available to ensure the observed differences in EPS found in S. Chen (2017b) across different note categories were caused by the differences in the cognitive demands for note-taking.

Table 2-3. The physical, temporal and cognitive demands of note-taking in different note categories reported in S. Chen (2017b)

Data	Indication	Form		Language
		Language vs. Symbol	Abbreviation vs. Full word	Chinese vs. English
Distance and speed of handwriting	Physical and temporal demands	Symbol<Language	Abbreviation < Full word	Chinese ≈ English
EPS	Cognitive demand	Language<Symbol	Abbreviation < Full word	English < Chinese
The written notes	Note preference	Language	Abbreviation	English

S. Chen (2020a) explored the impact of interpreting directionality on professional interpreters' cognitive load during the input and output phases of CI. The former was indicated by the participants' EPS during note-taking, and the latter was reflected by their mean fixation duration (MFD) on the notes during the process of note-reading. The results show that, compared with interpreting into their L1, the professionals presented significantly longer EPS when interpreting into their L2. Since there were no significant differences between the two interpreting directions in terms of note quantity, the increase of EPS in the L1-to-L2 task could have been caused by the use of a larger proportion of symbols and of a smaller percentage of notes in English. According to S. Chen (2017b), language notes and English notes required less cognitive effort from interpreters than symbols and Chinese notes. Therefore, when the participants listened to their L2, which was generally more demanding than listening to their L1, they preferred to take notes in the forms and languages that were less effortful during note-taking. This finding demonstrates that the cognitive demand for listening comprehension in the input phase of CI plays an important role in affecting interpreters' note-taking effort and note choices. These effects further decide the interpreters' cognitive load during the output phase of CI, which directly exerts impacts on their interpretation quality. Taken together, a CI task is an integrated process that includes receiving the source speech, producing the notes, receiving the notes and producing the target speech. All of these are integral components of CI that should not be studied separately.

With 22 data points collected from professional interpreters in S. Chen (2020b), S. Chen (2022) added another set of data points collected from 22 student interpreters to examine the effects of interpreter work experience and interpreting directionality on the process and product of CI. In terms of between-subject comparisons, she found that the professionals had a shorter EPS, a longer distance of pen movement, a faster writing speed and a smaller pen-tip pressure during note-taking than the students. These results imply that, compared with the student group, the professional group experienced less cognitive load and psychological pressure during the process of note-taking but bore heavier temporal and physical demands during the process of note-writing. These group differences suggest that, as interpreter work experience accumulates, interpreters can speed up the shift from listening comprehension to note production, better manage the psychological tension during the process of note-taking, and accelerate their writing speed. However, the two groups presented similar patterns of EPS differences among various note categories, with numbers being the shortest, symbols being the longest, and language notes in the middle. In other words, interpreter work experience did not affect the difference in the EPS across different note forms and note languages. As for within-subject comparisons, S. Chen found that both groups presented longer EPS in the L1-to-L2 direction than in the L2-to-L1 direction, and that the two groups' difference in interpreting quality was more obvious in the former direction than in the latter. Specifically, the professionals successfully maintained their interpreting performance at a high level in the two interpreting directions, whereas the students' performance was easily affected by the changes in the cognitive demand for listening comprehension in the bidirectional CI tasks. From this perspective, interpreting directionality and interpreting experience are both found to be important variables of the cognitive load of interpreting and the quality of interpreting products.

Overall, three essential findings can be summarized from S. Chen's series of research. Firstly, as interpreter work experience accumulates, interpreters can accelerate note-planning (indicated by EPS) and note-writing (indicated by the distance and speed of handwriting) procedures. Secondly, an increase in the cognitive load of listening comprehension can lead to an increase in the cognitive load of note-taking. Thirdly, the cognitive demand for note production can affect interpreters' note choices, and these note choices further affect their cognitive load of note reception, all of which contribute to the quality of the final interpreting output. However, it is worth mentioning that these findings are reported based on the fact that EPS is adopted as the sole indicator to reflect interpreters' cognitive load of note-taking. As discussed above, a longer EPS does not necessarily represent a greater amount of cognitive load. In essence, EPS only reflects how interpreters distribute their time to note-planning and note-writing procedures during the process of note-taking. S. Chen (2020b, p. 134) also points out that EPS only "potentially indicates the combined cognitive effort required to analyse the source speech unit, to decide whether or not to write a note, and if so, which form and language to use". Therefore, these findings based on the EPS data need to be further validated with other reliable indicators of cognitive effort.

In addition, although all the participants in S. Chen's study had Chinese as their L1 and English as their L2, the professional interpreters had around 7.4 years of working and living experience in an English-speaking country (Australia). Compared with general bilinguals whose L2 proficiency is usually lower than that of their L1, the professional interpreters in S. Chen's experiments were very proficient L2 users. This could directly affect their note-taking effort and note choices. Furthermore, the student interpreters in S. Chen (2022) were English-major undergraduates in a Chinese city. With such considerable differences in the two groups of participants' language background and life experience, findings based on these data should be interpreted cautiously. Meanwhile, the interpreting materials (one in English and one in Chinese) in S. Chen's research were designed for Australian contexts. Hence, the professionals could be more familiar with the topics than the students. All in all, the observed professional-novice differences in the note-taking behaviour in CI need be re-examined by recruiting professional and student interpreters with a more balanced language backgrounds and using materials that participants are equally familiar with.

In summary, process-oriented note-taking research is at a take-off stage. With pen-recording and eye-tracking methods, researchers are able to visualize the process of note-taking and quantify the effort of note-taking. However, how to select the most appropriate pen-recording and eye-tracking indicators to measure interpreters' cognitive, physical and temporal aspects of the note-taking effort remains underexplored.

2.2.2 Note-taking product

Compared to the process of note-taking, the product of note-taking (i.e., the notes) has attracted much more attention from researchers since the 1950s (e.g., Rozan, 1956). Seventeen studies that investigate the three aspects of interpreters' note-taking product, note language, note form, and note quantity are reviewed in this section. All of these studies attempt to answer two questions: 1) what are interpreters' preferences for different note languages, note forms and note quantities? 2) how do these note preferences enhance or undermine their interpreting quality? To answer these questions, researchers first categorize notes into language notes, symbols, and numbers (e.g., S. Chen 2020b). Then, they further classify languages notes according to the form of notes (e.g., Cardeon, 2018) and the language of notes (e.g., Arumí Ribas, 2012) (Figure 2-3). Note forms include full words and abbreviations; and note languages include the SL or TL of the interpreting task, the L1 or L2 of the interpreter (i.e., the mother tongue or the second language), and the A language or B language of the interpreter (i.e., the language the interpreter is more or less proficient in). In most cases, A language is equal to interpreters' L1, and B language represents their L2. Altogether, the identified 13 studies in this section cover 9 language pairs, 4 interpreter types, and 2 interpreting directions.

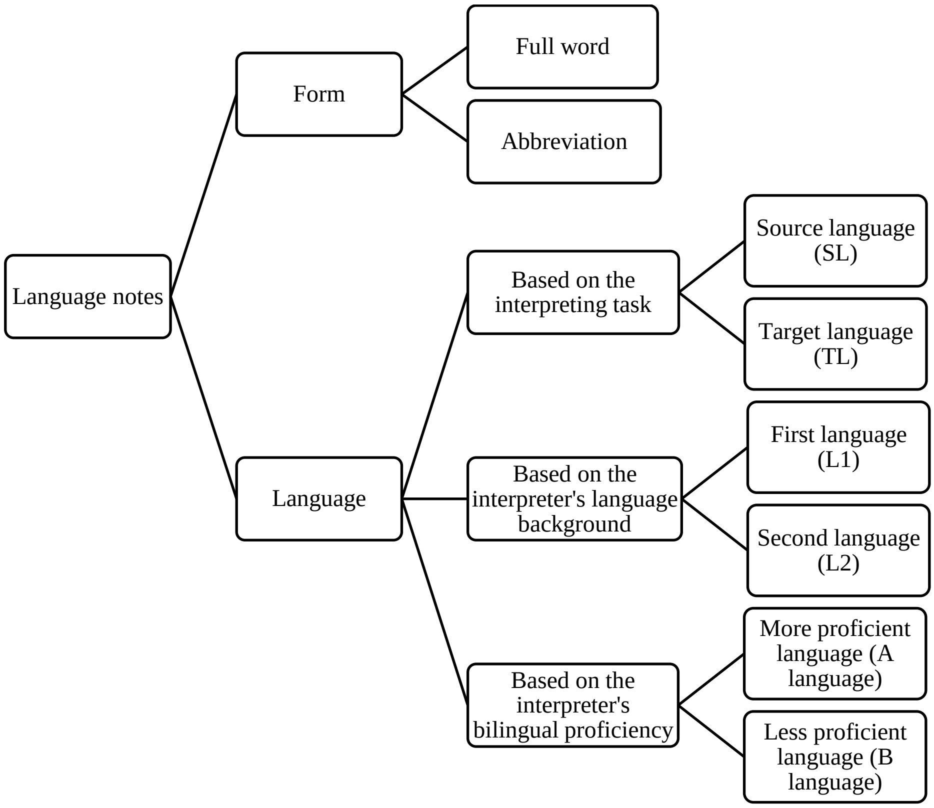


Figure 2-3. Categorization of language notes

2.2.2.1 Note language

Researchers have looked into both student interpreters' and professional interpreters' preferences for note languages. To present the findings clearly, the first half of this section reviews studies that focus on student interpreters, and the second half of the section is devoted to studies that involve professional interpreters. Student interpreters are further classified into beginner student interpreters, and advanced student interpreters with one year's interpreter training experience as the bar (Wen & Dong, 2019) (as shown in Table 2-4). As for professional interpreters, researchers have adopted different criteria to define "professional" in their studies (e.g., at least two years of working experience in Liu et al., 2004, and NATTI-accredited interpreters in S. Chen 2020a, 2020b). Some professional interpreters in one study would only be considered as advanced student interpreters in another (Wen & Dong, 2019). To avoid confusion, the review here follows the original expressions in the articles used to name the interpreters, but will also provide specific details about the training and working experience of the so-called "professional interpreters".

Table 2-4. Studies identified with student interpreters as participants

Literature	Language pair	Interpreting direction	Participant type	Sample size	Choice of note language	Choice of note form
Lung (2003) (Inaccessible)	Chinese -English	L2-L1	Beginner	21	SL>/TL	Few SYM
					L2 >L1	FW>AB
Dai & Xu (2007)	Chinese -English	L1-L2	Beginner	12	SL>TL	More AB and SYM in trained students
					L1>L2	AB>FW
Liu (2010)	Chinese -English	L1-L2	Beginner	120	SL>/TL	LG>SYM
					L1>L2	FW>AB
Wang et al. (2010)	Chinese -English	Both	Beginner	12	SL>TL	LG>SYM
						AB>FW
Gao (2019)	Chinese -English	Both	Beginner and advanced	46	SL>TL	NA
					Shift as interpreting experience	
Dam (2004a)	Danish -Spanish	Both	Advanced	4	L1>L2	LG>SYM
					AL>BL	
Szabó (2006)	Hungarian -English	Both	Advanced	8	L2>L1	LG>SYM
					BL>AL	

Note.
1. Studies are listed according to the researched interpreter type(s) and the year of publication.
2. AL=A language, BL=B language, LG=language notes, SYM=symbol, FW=full word, AB=abbreviation, NA=not available. This applies to all tables.
3. ">" represents that more notes are taken in the former note language/form than in the latter one, and "<" means that less notes are taken in the former note language/form than in the latter one.

- ***Beginner student interpreters***

For beginner student interpreters, Dai and Xu (2007) compared the language of notes created by six students who had received three months' interpreting training and six students who learned interpreting by themselves. The self-taught students had obtained an important certificate in interpreting in China, suggesting a certain degree of interpreting capability. The source material consisted of five typical political speeches

(529 words in Chinese), a text type the students were familiar with. It was found that, although both groups of students' notes were dominated by the SL of the interpreting task (also the participants' L1), the trained students used more TL in their notes than the untrained ones, indicating an effect of interpreter training experience on the preference of note language during note-taking. This also implies that, when interpreters' L1 coincided with the SL of the interpreting task, they preferred to take notes in their L1 because this can save them some effort in SL processing. However, these comparisons were made based on the original counts of the notes in different languages. As the two researchers report in the article, the untrained group had a bigger note quantity than its counterpart across the five source interpreting segments. Therefore, using percentages to compare the two groups' choice of note language is more reasonable than using sheer counts. Moreover, since note-taking is featured with high individuality (e.g., S. Chen, 2016), the fact that the two researchers simply added the participants' note counts together to represent a whole group's note preference could obscure individual differences. One solution for this issue is to increase the sample size in each participant group and conduct statistical analysis to confirm that the observed group differences also exist statistically.

By comparison, Liu (2010) had a much larger sample size by collecting notes from 120 English-major fourth-year undergraduates who had just completed one semester's CI training. The adopted L1-to-L2 CI task was described as being of a medium level of difficulty, and the delivery rate was 200 words per minute (wpm). After scoring the students' performance from four aspects, information completeness (25%), fluency of delivery (25%), logic (24%) and accurate expressions in the TL (25%), Liu (2010) compared the notes created by the 31 best performers and the 31 worst performers. In accordance with Dai and Xu (2007), Liu (2010) found that both high- and low-score students preferred the SL over the TL (i.e., their L1 over their L2) in note-taking. Except for the reasons mentioned by Dai and Xu (2007), Liu (2010) adds that this language preference could also be attributed to the students' lack of processing capacity and inadequate language proficiency. As they were overwhelmed by the information delivered by the speaker, they had no spare cognitive resources to complete the SL-to-TL transfer during note-taking. In addition, Liu (2010) reports that the statistical analysis shows there were no significant correlations between the students' interpreting scores, on the one hand, and their use of note language, note form and note quantity, on the other hand. He thus concludes that interpretation quality was not decided by note-taking. However, since Liu (2010) only investigated the relationship between interpreting performance and the product of note-taking, more exploration is needed to examine the relationship between interpreting performance and the process of note-taking.

Wang et al. (2010) studied the notes of 12 student interpreters who had the same interpreting backgrounds as those in Liu (2010). The results show that, on the one hand, regardless of interpreting directionality, most of the students took significantly more notes in the SL of the interpreting task than that of the TL, suggesting a more important role for the input language of the interpreting task than the students' L1 in deciding their note language. On the other hand, there were two students who always took more notes in their L1 than in their L2 in the bidirectional interpreting tasks, which implies an opposite preference to the previous finding. Noting these individual differences in note languages, Wang et al. (2010) concludes that there are no absolute rules behind the interpreters' choice of note language. Both bilingual proficiency and the features of the interpreting task could exert impacts on their language choice. In addition, in line with Liu (2010), Wang et al. (2010) observed no significant correlations between the use of note languages and the students' interpreting performance, again throwing doubt on the role of note-taking in attributing CI quality.

Gao (2019) analysed the notes of three groups of student interpreters: 12 finalists from a national interpreting contest (described as high-level interpreters), 18 interpreting-major masters (described as medium-level interpreters), and 16 Bachelor students who had completed one semester's interpreting training (described as low-level interpreters). The two source materials lasted 1 minute and 40 seconds (195 wpm) in Chinese and 1 minute and 53 seconds (119 wpm) in English, and the topics (smart city and machine learning) were described as being in common for the participants. The results show that, when interpreting into their L1 (English-to-Chinese), the high-level students had a more even distribution of the SL and TL in their notes. By comparison, the medium- and low- groups used more SL (the L2) than TL (the L1) in note-taking; while it was among the low-level student interpreters that the proportion of notes in the SL was the largest. When it came to Chinese-to-English interpreting, the high-level group presented the greatest percentage of notes in their L1 and SL (Chinese) and the smallest percentage of notes in their L2 and TL (English). The medium-level students' note language shifted along with the SL of the interpreting task; while the low-level students showed the smallest proportion of notes in TL across the three groups. All of these results suggest that the students' notes were dominated not by their L1 (also their A language) but by the SL of the interpreting task. Improvement in interpreting expertise thus could help interpreters to decrease the effort of listening and analysis, leading to faster language transfer during note-taking and a larger proportion of notes in TL. However, it is worth mentioning that the high-level group completed the two CI tasks in an interpreting contest, while the other two groups finished the task in a lab. The greater pressure that the high-level group bore during the contest thus could affect their note-taking choice and interpreting performance. Therefore, the conclusions related to the high-level interpreters need to be cautiously interpreted and generalized.

Overall, three major findings are reported in previous studies on beginner student interpreters' note languages. Firstly, the SL of the interpreting task plays an important role in deciding the beginners' note language. Secondly, their preference for the SL of the interpreting task is reinforced when it coincides with their L1 (also the A language in the reviewed studies). Thirdly, the choice of note language does not directly affect interpreting performance. When explaining the existing findings, researchers often resort to concepts such as cognitive resources and processing capacity, which are difficult to verify without process-oriented research methods. Therefore, more process-oriented research is needed to confirm whether speculations about the cognitive mechanisms underlying beginners' choice of note languages are shown to be valid.

- ***Advanced student interpreters***

Two studies specifically probed into advanced student interpreters' choice of note language. The first one was conducted by Dam (2004a) who collected notes from four master's students in Conference Interpreting, with three having Danish as their L1 and Spanish as their L2, and the last student having the inverse language combination background. The source material was about unfair competition in EU competition policy, which was described as common for the students. The students were trained with Rozan's (1956) principles of note-taking, which assumes that taking notes in the TL of the interpreting task can save interpreters effort in target speech production. However, regardless of interpreting directionality and the students' language background, all of the participants in the Dam (2004a) study preferred to take notes in the language they were more proficient with (A language) than the language they were less proficient with (B language). In other words, "interpreters are likely to take notes in whichever language is easier (to access) and therefore faster" (p. 13). Although the sample size is very small in Dam (2004b), her results bring a new perspective to the investigation of note language.

Szabó (2006) followed Dam (2004a) to test the impact of language status on eight second-year interpreting students' choice of note language. The students had Hungarian as their A language and English as their B language. It turned out that, regardless of interpreting directionality, the students always took more notes in their B language than in their A language. This result is seemingly opposite to that of Dam (2004b), but the two studies can be compared directly. In Dam (2004a), the participants were interpreting-major master's students who had no professional experience. However, in Szabó (2006), the student interpreters, as described in the literature, were actually "quasi professionals" (p. 133). Most of the latter had been working as interpreters in high-level international organizations and interpreter trainers in universities. Their average age was 33 years old and their average working experience was 3.5 years. They were only described as "students" because they were the first class of the EMCI[2] course. The observed "contradictory" findings above thus could be attributed to these notable differences in the interpreting background of the participants. Moreover, the participants in Szabó (2006) received education in English. In their professional life, they interpreted mostly from Hungarian (A language) to English (B language). Therefore, the "subjects must have a very strong B language, practically equivalent in standard with their A language" (p. 142). This also explains why they preferred their B language over their A language during note-taking. In addition, concerning the specific language pair in Szabó (2006), it was explained that, compared with Hungarian, "English has much shorter linguistic units, is much easier to abbreviate, and possesses a good number of internationally accepted contractions and acronyms" (p. 141). Hence, taking notes in English was comparatively easier than in Hungarian. Taken together, to provide a more comprehensive investigation of the evolvement of interpreters' choice of note language, researchers can collect notes from beginner student interpreters, advanced student interpreters, and/or professional interpreters with the same language background to conduct group comparisons. They can also follow the same group of interpreters along the course of their training experiences and professional life to observe how their note language evolves.

[2] For details of the European master's in Conference Interpreting program, please visit https://www.emcinterpreting.org/emci/

- ***Professional interpreters***

A variety of findings are reported about professional interpreters' preferences of note languages (Table 2-5). The first half of this section focuses on three studies that only involve professional interpreters as participants (S. Chen, 2017b; Dam, 2004b; Seleskovitch, 1975), and the second half reviews studies that report on the comparisons between professional interpreters and student interpreters.

Table 2-5. Studies with professional interpreters as participants (in chronological order)

Literature	Language pair	Interpreting direction	Participant type	Choice of note language	Choice of note form
Seleskovitch (1975)	English -French	L2-L1	12 professionals	Varied	Varied
Andres (2002)	French -German	L2-L1	14 professionals and 14 students	SL>TL	LG>SYM
				More TL uses in the professionals	
Dam (2004a)	Danish -Spanish	L2-L1	5 professionals	TL>SL	LG>SYM
				AL>BL	FW>AB
Abuín González (2012)	English -Spanish	L2-L1	10 beginners, 10 advanced students, and 10 professionals	Shift as experience	NA
S. Chen (2017b)	Chinese -English	Both	5 professionals	L2>L1	LG>SYM
				BL>AL	FW<AB
Cardoen (2018)	Dutch -English	L2-L1	5 first-year masters' students, 5 second-year master's students, and 5 professionals	NA	Easy task: Professionals: LG>SYM Students: SYM>LG
					Difficult task: LG>SYM
Wang (2018) (Inaccessible)	Chinese -English	Not known	Beginner and professional interpreters	NA	Better use of SYM in the experienced group
S. Chen (2022)	Chinese -English	Both	22 professionals and 22 beginner students	Professionals: L2 and BL	Both groups: LG>SYM
				Students: SL	Professionals: FW ≈AB (from S. Chen, 2020a)

Seleskovitch (1975) observed a combinational use of SL and TL in an L2-to-L1 CI task among 12 professionals' notes. Extreme variations were found among the participants' note choices of forms and languages. She explained this phenomenon with the theory of sense, which claims that interpreters use notes to store the sense they extract from the source speech rather than the linguistic expressions in the speech (see Section 2.1.1). In this case, the appearance of the notes does not matter. What truly affects interpretation quality is the completeness and accuracy of the "sense" that is stored in the notes. Seleskovitch's (1975) conclusion is in a sense echoed in the finding of Liu (2010) and Wang et al. (2010) that there are no significant correlations between the features of the note-taking product and the quality of the participants' interpreting product. However, this finding is not in accordance with the majority of product-oriented note-taking studies, where notes are frequently found with strong linguistic features of the source speech.

In Dam (2004a), 5 professional interpreters (with 2 to14 years of interpreter work experience) took 58%-87% of their notes in their A language or the TL of the L2-to-L1 CI task, corroborating her earlier findings with advanced student interpreters that the A language played a decisive role in deciding interpreters' note language (Dam, 2004b). She also found that the participants shifted to their B language or the SL of the interpreting task when the source material became difficult. This shift of note language implies that, when dealing with easier interpreting tasks, interpreters tend to note more in the TL to reduce their burden of language transfer in target speech production; and that, when the interpreting task became difficult, they opt for a less effortful note-taking strategy by noting more in the SL of the CI task.

Different results are reported in S. Chen (2017b) with 5 NATTI-accredited professionals (5.4 years of interpreter work experience on average) working between Chinese (L1) and English (L2). S. Chen (2017b) found that the participants preferred to take notes in their B Language (L2) over their A Language (L1) in both interpreting directions. With the collected pen-recording data, she explains that this result could be caused by the fact that English notes entailed shorter EPS than Chinese notes, which means that it was faster for the participants to come up with English notes than Chinese notes. In the retrospective interview, the participants also reported that English was easier to write than Chinese. All of these could lead to the participants' preference for the B language in their note-taking.

Based on the three studies discussed above, four factors that affect professional interpreters' choice of note languages can be identified: 1) the balance of their bilingual proficiency; 2) the difficulty level of the source speech; 3) the inherent features of the language pair; and 4) the individuality of note preferences. To explore how these preferences of note languages develop during professional interpreters' educational and professional experience, a series of horizontal studies were conducted to examine the differences across student interpreters and professional interpreters in the choice of note language.

Andres (2002) found a clear preference for the SL over the TL in an L2-to-L1 CI task among 14 professionals' and 14 students' notes. However, she noted that the professionals took more notes in the TL than the students did, suggesting that the former group completed more language transfer during note-taking than the latter group. Similar effects of interpreter work experience on note languages are reported in Abuín González (2012). She discovered that there was a shift from the SL to the TL of the interpreting task in the notes created by beginner student interpreters (third-year students in Translation and Interpreting), advanced student interpreters (fourth-year students in the same program) and professional interpreters (five years' work experience on average). This again demonstrates that, as interpreting experience increases, interpreters can complete the language transfer during note-taking more easily and produce more notes in the TL of the interpreting task.

Different from the findings reported for European languages (Abuín González, 2012; Andres, 2002), S. Chen (2022), who looked into Chinese-and-English interpreting, found that professional interpreters (5.7 years' experience on average) preferred to take notes in their L2 in both interpreting directions, and that the students' choice of note languages shifted as the input language of interpreting changed. One reason for this inconsistency in the research findings could be the differences in the studied language pairs. In addition, the professional participants in S. Chen (2022) had very strong L2 as they lived and worked in an English-speaking country, while the students lived and studied in a Chinese environment. Therefore, S. Chen's results could not be directly compared with other studies and should be cautiously interpreted. However, despite the varied research findings, there appears to be a consistent advantage among professional interpreters over student interpreters in the ease of completing language transfer during the process of note-taking. To find out the reasons behind this advantage, it is essential to probe into the process of note-taking to examine how professional interpreters manage to complete language transfer in the speaker-paced input of CI.

2.2.2.2 Note form

In terms of the form of notes, the majority of previous studies have found that interpreters prefer language notes over symbols, but mixed results have been reported about their preferences for full words or abbreviations (see Table 2-4 and Table 2-5).

Between language notes and symbols, the latter are believed to be easy in handwriting, fast in note-reading, and independent of specific language interference (Gillies, 2017). Therefore, many researchers recommend taking notes in symbols rather than in language (e.g., Matyssek, 1989). However, symbols can be challenging for interpreters who have not fully mastered the use of this note form. For instance, Alexieva (1994) found that beginner student interpreters expended much cognitive effort on selecting or improvising symbols during the process of note-taking. As a result, there was insufficient processing capacity for listening comprehension, which further led to poor interpreting performance. This might also be the reason why beginner student interpreters rarely take notes in symbols (Chmiel, 2010; Lung, 2003). By contrast, research has found that advanced student interpreters (Dai & Xu, 2007) and professional interpreters (Wang, 2018) can use symbols more easily during the process of note-taking. However, even so, professional interpreters generally prefer to take notes in language rather than in symbols (e.g., Andres, 2002; S. Chen, 2017b, 2020a). Sometimes, advanced student interpreters have been shown to even use more symbols during note-taking than the professionals (Cardoen, 2018). Therefore, it appears that an increase in interpreting experience can lead to a decrease in the effort of noting symbols but not necessarily an increase in the use of symbols. In terms of the relationship between the use of symbols and interpreting performance, Liu (2010) found a significant positive correlation between the two among beginner student interpreters in his study. Similarly, Cardoen (2018) found that advanced student interpreters' interpreting fluency improved when they used more symbols, while by contrast, the professional interpreters' interpreting fluency decreased as they used more symbols. Moreover, when the CI task was easy, all the participants' interpreting accuracy was higher when they noted fewer symbols. However, the situation was the opposite for the difficult tasks (Cardoen, 2018). These findings suggest that interpreter training experience, interpreter work experience and source speech difficulty can interactively affect interpreters' note preferences and the relationships between the use of note forms and interpreting performance. Therefore, controlling the variables of the speaker and the features of the interpreting task is of great importance in conducting note-taking experiments.

Between full words and abbreviations, Wang et al. (2010) and Cardoen (2018) observed more use of abbreviations than full words among, respectively, beginner student interpreters and advanced student interpreters; whereas Dam (2004a) and Liu (2010) report the opposite preferences for the same two types of student groups. As for professional interpreters' notes, Dam (2004b) and Cardoen (2018) observed bigger proportions of full words than abbreviations, but S. Chen (2017b) found the opposite in her data. These apparently contradictory findings could be caused by the fact that these investigations varied considerably in source speech difficulty, participant recruitment method, the presence of the audience, and the studied language pair.

Cardoen (2018) was the only study that included three interpreter types (two student groups and one professional group) and two task conditions (easy and difficult) to comprehensively examine the effects of interpreting experience and source speech difficulty on the interpreters' preferences for note forms and note languages. Her results show that, in the easy segment of speech, the professionals used many full words whereas the beginners and advanced students mainly used symbols. She found this result "surprising" (p. 299), because the professionals in her study, who also worked as interpreting lecturers in universities, had attached great importance to the use of symbols in their teaching experience. However, they themselves did not put that into practice, indicating the great "disparity between theory and practice" (p. 299) in note-taking. One of the reasons for this result could be that the professionals in Cardoen's study rarely worked in the interpreting market, which means that they were not familiar with the use of symbols in real interpreting tasks. Furthermore, Cardoen noted that the interpreters adapted their choice of note forms within different task settings. Specifically, as source speech difficulty increased, the interpreters had larger note quantities, more frequent use of abbreviations and symbols and fewer notes in full words. She explained that this adaption was activated by the principle of economy of note-taking (Ilg & Lambert, 1996, p. 81). Compared with full words, symbols and abbreviations are generally easier to write. Therefore, within the same given time, interpreters can produce more notes in symbols and abbreviations than in full words. With increased information stored in the notes, the pressure on the interpreters' short-term memory decreases. Hence, they can allocate more cognitive resources to comprehend the source speech in the difficult task condition. Altogether, these results indicate that interpreter work experience and source speech difficulty can have great impacts on interpreters' note choices and interpreting performance. However, to prove Cardoen's explanations of her results, which include many speculations about what happened during the processes of note-taking and note-reading stages, process-oriented research methods and empirical evidence are needed.

2.2.2.3 Note quantity

The quantity of notes has always been a controversial topic in note-taking studies.

Scholars who opt for a large note quantity argue that this helps interpreters to release the pressure on their short-term memory (e.g., Dam et al., 2005), whereas those who recommend a limited number of notes claim that this can save interpreters some processing capacity for listening comprehension (e.g., Ilg & Lambert, 1996). To find out the answer to this question, one group of researchers chose to compare the notes created by interpreters with different levels of interpreting expertise, aiming to figure out how experienced interpreters deal with this issue. Another group of researchers has aimed to answer this question by revealing the relationship between note quantity and interpretation quality (Table 2-6).

Table 2-6. Studies on note forms and note quantity

Literature	Language pair	Interpreting direction	Participant type	Note quantity and CI quality	Note forms and CI quality
Her (2001)	Chinese -English	Both	Students with 8 weeks' and 4 weeks' training	Quantity+	NA
Dam et al. (2005)	Danish -Spanish	L2-L1	1 professional	Quantity+	Abbreviation+
Dai and Xu (2007)	Chinese -English	L1-L2	6 professionally-trained students and 6 non-professionally-trained students	No significant correlations	NA
Dam (2007)	Danish -Spanish	L2-L1	5 professionals	Quantity+	Abbreviation+ Full word-
Liu (2010)	Chinese -English	L1-L2	62 trained undergraduates	NA	Symbol+
Wang et al. (2010)	Chinese -English	Both	12 trained undergraduates	No significant correlations	NA
Salaets and Theys (2016)	Dutch -French	L2-L1	13 students	NA	Notes for links+
S. Chen (2017b)	Chinese -English	Both	5 professionals	No clear relationships	No clear relationships
Cardoen (2018)	Dutch -Spanish	L2-L1	5 first-year master's students, 5 second-year master's students, and 5 professionals	NA	Interpreting accuracy: Note quantity + Symbol+ in easy tasks and - in difficult tasks Interpreting fluency: Note quantity + Full word + Abbreviation –
S. Chen (2020b)	Chinese -English	Both	22 professionals	Quantity+ in L1-to-L2 interpreting	Symbols+ in L2-to-L1 interpreting
S. Chen (2022)	Chinese -English	Both	22 professionals and 22 students	Professionals> Students	Use of symbols: students>professionals

Note. "+" represents a positive relationship with interpretation quality and "–" means a negative relationship with interpretation quality.

Mixed results were reported in the studies that compared note quantities in different interpreter types to investigate the optimal note quantity in note-taking. For instance, Cardoen (2018) found that it was the second-year master's students that noted the most, followed by the professional interpreters and then the first-year master's students. There thus appears to be a U-shape of note quantity changes as interpreting experience accumulates. However, Abuín González (2012), who also recruited beginner student interpreters, advanced student interpreters and professional interpreters in his investigation, found that the professional interpreters noted more than the students, presenting different results to Cardoen (2018). In studies attempting to reveal the relationship between note quantity and interpretation quality, most draw their conclusions based on the researchers' own subjective observations. For instance, some researchers reported an observed positive relationship between note quantity and interpretation quality among beginner student interpreters (e.g., Her, 2001), advanced student interpreters (e.g., Cardoen, 2018) and professional interpreters (e.g., Dam, 2007; Dam et al., 2005), but others conclude there is no clear relationship between the two (e.g., Dai & Xu, 2007; Wang et al., 2010). For the small proportion of the researchers who conducted correlation tests to decide the coefficients between the two variables, Liu (2010) detected no significant correlations between note quantity and interpretation quality among beginner student interpreters in an L1-to-L2 CI task, but S. Chen (2020b) discovered a positive correlation in the same interpreting direction among professional interpreters. At the same time, the positive relationship S. Chen (2020b) found in the L1-to-L2 interpreting task was not observed in the other interpreting direction. These findings suggest that the relationship between note quantity and interpretation quality could be jointly decided by interpreter training experience, interpreter work experience and interpreting directionality.

As mentioned, Cardoen (2018) is the only study that looked into the features of effective and ineffective notes in three types of interpreters under two task conditions. In terms of interpreting accuracy, Cardoen (2018) found that an increase in note quantity was associated with an increase in interpreting accuracy. As for interpreting fluency, it was found that beginner student interpreters' fluency was adversely affected by their note quantity. According to Cardoen's explanation, this was because, when note quantity increased, the number of notes that the students needed to process within a given time also increased. In that case, more processing capacity was devoted to note-reading and less processing capacity was available for target speech production and speech monitoring, which finally lead to undermined interpreting fluency. However, such relations between note quantity and interpretation quality were not found in the advanced student interpreters and professional interpreters. Instead, she observed a positive relationship between the two in the advanced student group in an easy CI task and among the professional group in a difficult one. These results suggest that the associations between note quantity and interpreting fluency are of great complexity and that relevant conclusions should only be drawn with a clear explanation of the interpreter-related and task-related features. Moreover, the fact that interpreting accuracy and interpreting fluency have presented different associations with note quantity also indicates that the note-taking product can exert different impacts on the various dimensions of interpretation quality. Therefore, when investigating the relations between note activities and interpretation quality, it is essential to apply a comprehensive assessment method to evaluate interpreters' performance. All in all, the existing research has not come to a widely accepted conclusion about the appropriate quantity of notes in CI. However, a consensus can be found in these studies that, regardless of large or small note quantity, there are upper and lower limits for this quantity (e.g., Cardoen, 2018; S. Chen, 2017b; Dam, 2007). Taking excessive notes will consume a large amount of effort, whereas taking a limited number of notes will pose challenges to interpreters' short-term memory in source speech retrieval.

2.2.3 Note-reading process

Compared to the product of note-taking, the process of note-reading is rarely explored in empirical interpreting studies (listed chronologically in Table 2-7). One important reason for the lack of research in this regard is that there is a shortage of appropriate apparatus that can directly visualize the note-reading process and quantify interpreters' note-reading effort. In 2017, Sijia Chen innovatively adopted eye-tracking to record interpreters' eye fixations on the notes during the output of CI. Her publications, S. Chen (2020a) and S. Chen et al. (2021), which were based on the same data set of 18 professional interpreters, reveal a series of processing patterns during the process of note-reading. This section sets out to review these findings to summarize what has been found about the cognitive aspects of note-reading and discuss these results with reference to two other studies (Abuín González, 2012; Xu & Chai, 2008) which investigate interpreters' note-reading difficulties through questionnaires. In addition, the discussion also extends to the methodological considerations that need to be included when adopting eye-tracking as a research method, as in the present study, to examine the process of note-reading.

Table 2-7. Studies on the process of note-reading

Literature	Language pair	Interpreting direction	Participant type	Research method	Main findings
Xu & Chai (2008)	Chinese -English	L1-L2	12 (non) professionally -trained students	Stimulated recall	• Professionally-trained interpreters have fewer note-reading problems than their counterparts
Abuín González (2012)	English -Spanish	L3-L1	8 beginner and 7 advanced students	Question-naire	• Advanced students are more bothered by unclear notes and memory problems
					• Beginner students are more troubled by understanding and connecting the notes
S. Chen (2020a)	Chinese -English	Both	22 professionals	Eye-tracking	• Note-reading is more effortful in the L1-to-L2 task than in the L2-to-L1one
S. Chen et al. (2021)	Chinese -English	Both	22 professionals	Eye-tracking	Note-reading effort: • FW≈AB • CN>EN in the early stage of note-reading • EN>CN in the late stage of note-reading (only in the L2-to-L1 interpreting)

Note. CN=Chinese, EN=English.

2.2.3.1 Reading patterns and cognitive effort

The process of note-reading is visualized in S. Chen et al. (2021) with a figure of the scan path on the notes. The researchers first numbered the notes one by one and drew AOIs on the individual notes. Then, they exported the time point that the interpreters' eyes started to fixate on the notes with the help of the eye-tracking software. With these two sets of data, they created a scan path of the interpreters' fixations on the notes, with the *x*-axis representing the passage of time and the *y*-axis showing the note that was being fixated on at the time.

As Figure 2-4 presents, the participant read the notes numbered between 17 to 20 from time 0 to 8,750 milliseconds (ms), and then moved his/her eyes to notes 22-26 from time 8,750 ms to 21,250 ms. After that, the participant moved to notes 28-33 and read them for the rest of the time. In other words, the interpreters did not read notes one by one but group by group (Figure 2-5). In addition, after checking the participants' interpretations, S. Chen found that, when the participants were going to finish the current note group's interpretation, they would read ahead to the next one, indicating a procedure of interpretation preparation during the process of note-reading. Overall, S. Chen et al. (2021) found that note-reading was a non-linear reading process where interpreters would decipher their notes in groups and prepare for the following interpretations by reading across these note groups.

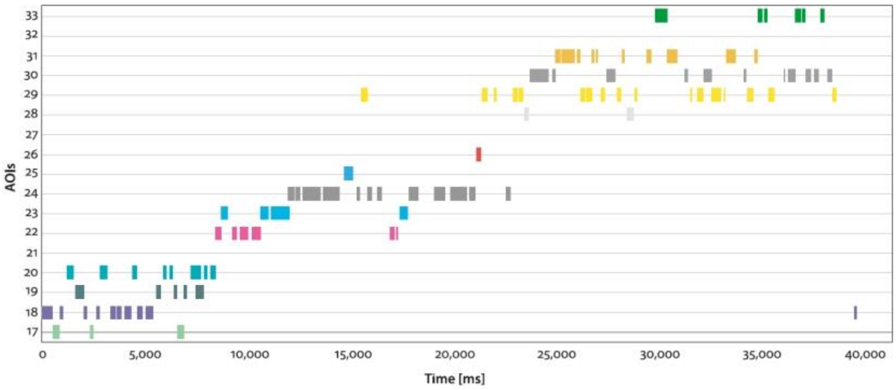


Figure 2-4. The scanpath of note-reading presented in S. Chen et al. (2021, p. 13)

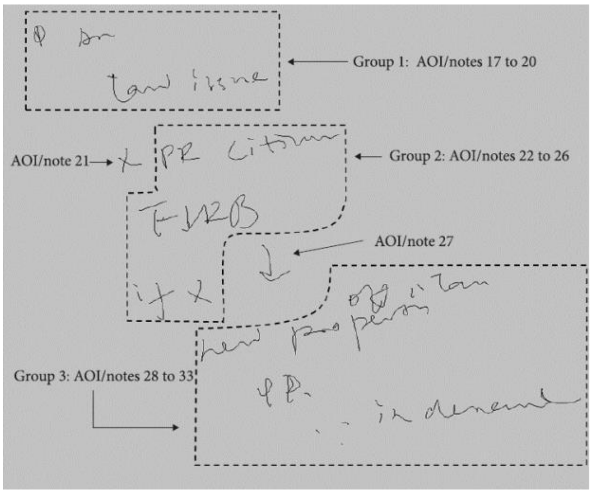


Figure 2-5. The note groups that match Figure 2-4 (retrieved from S. Chen et al., 2021, p. 14)

S. Chen measured the cognitive effort of note-reading by calculating the participants' MFD on the notes, which is a frequently adopted indicator of the depth and effortfulness of cognitive processing in translation and interpreting studies (e.g., Hvelplund, 2019; Rayner, 1998; Stachowiak-Szymczak & Korpal, 2019). The results show that note-reading (277ms) entailed longer MFD than silent reading for comprehension (225ms in Rayner, 1998), reading in preparation for translation (245ms in Dragsted, 2010; 205ms in Jakobsen & Jensen, 2008), and sight translation (252ms in Dragsted & Hansen 2009; 235ms in Jakobsen & Jensen, 2008). Despite the differences in the experimental settings and participant backgrounds across these reading studies, the long MFD found in the note-reading process could still indicate that this is a cognitively taxing experience. Such a high cognitive demand for note-reading could be attributed to its special underlying cognitive mechanism. Compared with those tasks that only require visual or auditory processing, note-reading requires interpreters to transform what they see in the notes into what they speak in the target speech. The visuospatial sketchpad and phonological loop in their WM, which are, respectively, responsible for visual and auditory processing (Baddeley & Hitch, 1974), are thus impacted with high cognitive demands (see Chapter 3, Section 3.1.3). In addition, unlike the materials in normal reading tasks, notes as the reading material for interpreters in CI have highly condensed or even fragmentary surfaces (Albl-Mikasa, 2008). More often than not, beginner student interpreters find themselves stuck in note recognition and advanced student interpreters are troubled by memory retrieval (Abuín González, 2012; Xu & Chai, 2008). All of these results indicate that note-reading is a cognitively complex activity that can directly exert impacts on the production of the target speech in CI.

Another study which could bring abundant implications for understanding of the note-reading process is S. Chen (2020a). The reason why this study is introduced here after S. Chen et al. (2021) is that the former actually focused on the impact of interpreting directionality on interpreters' cognitive load of interpreting during the input and output stages of CI, rather than that of note-taking and note-reading. However, S. Chen in fact adopted two note-related measures to indicate the interpreters' cognitive load across the two stages of CI: the participants' EPS during note-taking, and their MFD on the notes during note-reading. Technically speaking, the collected data should have been interpreted within the scope of note-taking and note-reading. Specifically, S. Chen's results show that, at the input stage of CI, the interpreters had shorter EPS in the L2-to-L1 interpreting task than in the other direction; whereas, at the output stage of CI, the interpreters had a longer MFD on the notes in the L1-to-L2 task than in its reverse-direction counterpart. From a note-taking perspective, this means that the interpreters needed more time to plan for the notes in the L2-to-L1 interpreting, and that they expended more cognitive effort on reading the notes in the L1-to-L2 interpreting. In the literature, S. Chen (2020a) explains that these directionality effects on the process of interpreting were caused by the different cognitive demands of listening comprehension in the tasks. Compared with listening to one's L1, comprehending a speech that was delivered in the interpreter's L2 required a longer processing time and greater cognitive effort. Therefore, instead of generating a thorough understanding of the source speech, the interpreters chose to follow the source micropropositions closely and take notes mechanically. As for the shorter MFD on notes during the output stage of the L1-to-L2 CI task, S. Chen explains that this could be caused by the fact that L1 production was less demanding and effortful than that of L2 production (see also, e.g., Lin et al., 2018 on SI; Mead, 2005 on CI). These findings on this part of the eye data are in accordance with the finding that higher interpreting fluency was achieved in the L2-to-L1 interpreting than in the other direction. Taken together, these results indicate that the cognitive demand for listening comprehension at the input stage of CI could affect the time that interpreters spend on note-planning and the cognitive effort they expend on note-reading, all of which in turn exert further impacts on the performance of interpreting.

2.2.3.2 Early processing and late processing

Looking into the process of note-reading, S. Chen et al. (2021) measured interpreters' cognitive effort in the early and late stages of note processing, with a variety of eye-tracking indicators. Altogether, these indicators point to six aspects of the note-reading effort: 1) note recognition (first fixation duration) and the integration of note meanings at a lexical level (first-pass dwell time) during early processing; 2) re-analysis of notes and the integration of note meanings at a syntactic and a textual level (total dwell time and second-pass dwell time) during late processing; 3) incomplete note processing (number of revisits); 4) the degree of non-linearity during note-reading (regression rate); 5) the predictability of the notes during reading (skip rate); and 6) the overall cognitive effort invested in note-reading (number of fixations and average fixation duration). The results show that, regardless of interpreting directionality, reading language notes was more cognitively demanding than reading symbols in both stages. Between full words and abbreviations, the former demanded more cognitive resources from the interpreters than the latter, but such differences disappeared when the length of the notes was controlled. On Chinese (L1) and English (L2) notes, the two sorts of notes entailed a similar level of cognitive load in early processing, but English notes required a greater amount of cognitive effort than Chinese notes in later processing (only in the L2-to-L1 interpreting). Important implications can be inferred from the combined results of S. Chen (2020b) and S. Chen et al. (2021). At first, the cognitive demand for listening comprehension during the input stage of CI can affect interpreters' note-taking effort and note choices. Such impacts on the use of note forms and note languages further influence the cognitive demand for note-reading. As the cognitive load of note-reading constitutes a part of the cognitive load at the output stage of CI, it exerts further impacts on interpreter' accuracy and fluency of delivery. Taken together, the process of note-taking, the product of note-taking and the process of note-reading are three integral note activities in CI that should not be discussed independently. To provide a full account of interpreters' note-taking behaviour, researchers need to cover every stage of these note activities as an integrated process.

One concern about the measurement of note-reading effort is that, unlike other reading tasks where participants are presented with the same material, interpreters are reading different notes that are created by themselves. Since the reading material for interpreters is not identical, it might not be rigorous to compare one interpreter's note-reading effort to that of another. Neither is it precise to measure a group of interpreters' note-reading efforts simply by calculating the mean of these note-reading efforts. One solution for this is that, when examining the effort of note-reading, researchers should also look into the process and product of note-taking. As S. Chen presents in her Ph.D. thesis (see S. Chen, 2017a, p. 152), there appears to be a trade-off between the cognitive effort of note-taking and that of note-reading. Therefore, only by ensuring that the interpreters have adopted the same note-taking approach and presented the same note patterns can researchers compare their efforts in note-reading. All in all, it is essential to uphold an integrated view on the note activities in CI. After all, the process of note-taking, the product of note-taking and the process of note-reading are interconnected with each other, and any or all of these processes could enhance or undermine interpretation quality.

2.3 Summary

The first half of this chapter was devoted to theoretical discussions concerning the cognitive and linguistic attributes of notes in CI; while the second half reviewed empirical studies that have explored the process and product of note activities in CI. Overall, existing studies have shown that note-taking and note-reading are two complex tasks that demand cognitive, physical and temporal aspects of effort devotion. Such activation and distribution of effort are affected by the experience level of the interpreter and the difficulty level of the interpreting task. More importantly, these experience- and difficulty-related effects on the process of note-taking extend to the product of note-taking and the process of note-reading, all of which, finally, contribute to the interpretation quality. Hence, to fully understand the cognitive mechanisms underlying interpreters' note-taking behaviour in CI, researchers should: 1) adopt both process- and product-oriented methods to visualize the production and reception phases of note activities in CI; 2) include diversified interpreter types to examine the experience effects on note-taking; 3) design both easy and difficult interpreting tasks to observe the adaptation of interpreters' note-taking behaviour; and 4) cover all the three stages (i.e., note-taking, notes and note-reading) of note activities in CI to provide a full and integrated account of interpreters' note-taking behaviour.

For process-oriented note-taking studies, the selection of the apparatus and the adoption of the indicators can be notably tricky. Video-recording (Andres, 2002) can be a challenging method because it can lead to tedious data processing and imprecise data analysis (e.g., S. Chen, 2017b). Meanwhile, the dual-task paradigm (Hu, 2008) is considered too intrusive to allow interpreters to complete note-taking naturally. As for pen-recording (e.g., S. Chen, 2017b), this achieves much higher ecological validity as compared to the dual-task paradigm. However, a digital pen alone is not powerful enough to measure the cognitive effort of note-taking. One solution for this is to combine pen-recording with eye-tracking to collect interpreters' pen movements and eye fixations during the note activities. With these two streams of data, researchers would be able to infer how interpreters distribute their limited processing capacity to complete visual processing (fixation distributions on notes), cognitive processing (fixation durations and counts on notes), physical handwriting (pen movement-related indicators) and temporal management (EPS of note-taking) during note-taking and note-reading. Although these observed behavioural changes are only by-products of cognitive processing (Hvelplund, 2011), they can help researchers to probe into the intangible cognitive processes underlying the production and reception phases of the note activities.

Aside from these methodological concerns, the control of interpreter-related variables and task-related variables is also of great necessity in note-taking research. Studies have shown that the accumulation of interpreting experience can shorten interpreters' EPS (Andres, 2002; S. Chen, 2022; Hu, 2008), accelerate their handwriting speed (S. Chen, 2022), facilitate the use of the TL in note-taking (e.g., Abuín González, 2012), and increase the number of notes (e.g., S. Chen, 2022). A high comprehension demand during the input stage of CI can lead to a shorter (e.g., S. Chen, 2020a) or longer EPS (Hu, 2008), more use of the SL in note-taking (Dam, 2004b), and a larger note quantity (Cardoen, 2018). Therefore, interpreting experience and task difficulty are essential factors that should be considered when investigating and explaining interpreters' note-taking behaviour. More importantly, since the studies reviewed in this chapter adopted different participant recruitment criteria, sometimes a professional interpreter in one study would be only classified as an advanced student interpreter in another. Therefore, to ensure the reliability and generalisability of research findings, it is important to adopt an appropriate standard for participant selection.

Furthermore, the examination of experience and difficulty effects on interpreters' note-taking behaviour should not be restricted to the production stage of the notes. Instead, the process of note-taking, the product of note-taking and the process of note-reading should be studied in an integrated way. After all, notes are created based on the note-taking strategy, and the features of notes in turn affect the cognitive demand for note-reading. Moreover, previous literature has identified a trade-off between the cognitive effort of note-taking and of note-reading among professional interpreters (S. Chen, 2017a), which has significant implications in terms of the way interpreters effectively allocate their limited processing capacity across the two stages of the note activities. However, since the adopted indicators of the cognitive effort of note-taking (EPS and response time in task switching) can be problematic, it is necessary to re-examine the claimed relationship between the two stages of noting effort with more reliable indicators and diverse interpreter types.

Chapter 3: From Note-taking Process to Interpreting Product: A Cognitive Approach

Understanding that note-taking is a cognitively demanding activity that is conducted under taxing conditions in CI serves as an important foundation for the departure of the present study. Adapting from Paas and Van Merriënboer's (1994) construct of cognitive load (Figure 3-1), this chapter aims to explore the cognitive aspects of note activities in CI from source speech reception to target speech production. The original construct consists of three parts: the causal factors that can exert impacts on cognitive load; the cognitive load posed on the task performer; and the assessment factors that can be affected by an increase or decrease in the cognitive load. Accordingly, this chapter will explore the processes of note activities in CI within three modules (Figure 3-2): task-, environment-, and interpreter-related factors that contribute to the cognitive load of note-taking (Section 3.1); the construct of visual, cognitive, physical and temporal demands of note production and note reception (Section 3.2); and the relationships between different note-taking stages and the associations between interpreters' note-taking behaviour and interpretation quality (Section 3.3). Finally, a summary of this chapter is provided in Section 3.4.

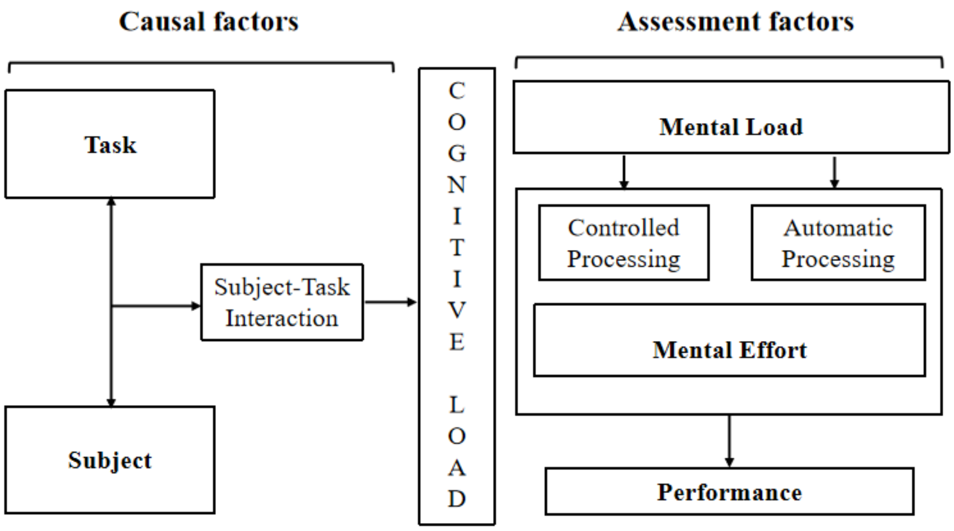


Figure 3-1. The construct of cognitive load proposed by Paas and Van Merriënboer (1994), reproduced with permissions from the authors; copyright 1994 Springer.

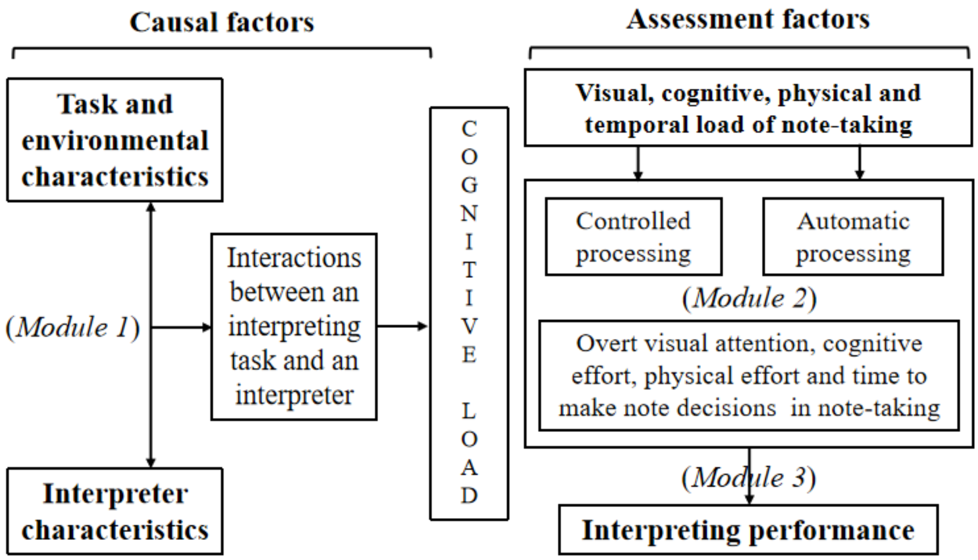


Figure 3-2. The adapted construct of cognitive load in note-taking

3.1 Identifying factors of the cognitive load of note-taking

Note-taking is an integral subtask of classic CI (Pöchhacker, 2004). Therefore, factors influencing the cognitive load of interpreting might also lead to changes in the cognitive load of note-taking. Through illustrating the task characteristics, environmental characteristics and interpreter characteristics that have been identified as important factors of the cognitive load of interpreting, this section aims to explain how these factors can exert potential impacts on the process of note-taking.

3.1.1 Task and environmental characteristics

Based on Meshkati's (1988) comprehensive mental workload model and Paas and Van Merriënboer's (1994) construct of cognitive load, S. Chen (2017c) proposes a model of cognitive load that is dedicated to interpreting activities. As presented in Figure 3-3, S. Chen classifies the construct of cognitive load in interpreting into two categories: task and environmental characteristics that jointly decide the inherent "input load" (Johannsen, 1979, p. 4) or "mental load" (Paas & Van Merriënboer, 1994, p. 353) of an interpreting task; and interpreter characteristics that decide the "operator effort" (Johannsen, 1979, p. 4) or "mental effort" (Paas & Van Merriënboer, 1994, p. 353) of an interpreter.

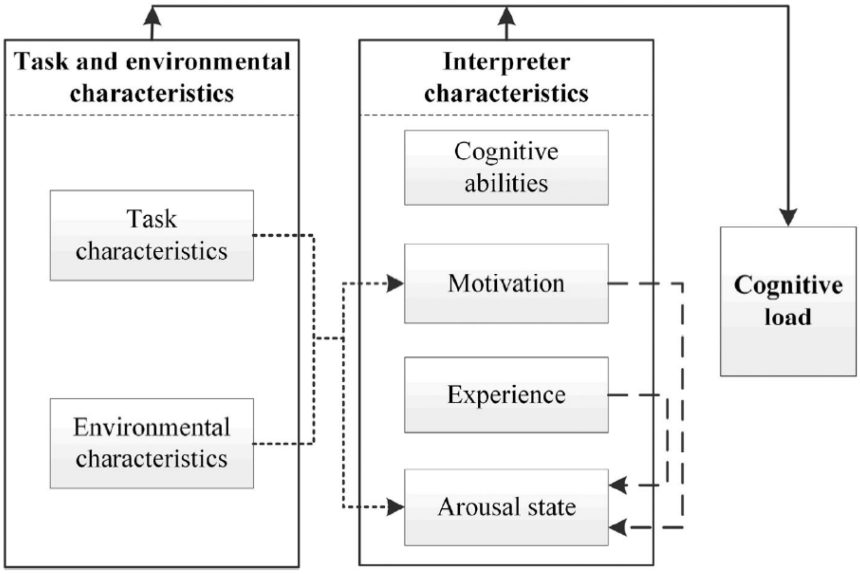


Figure 3-3. The construct of cognitive load in interpreting (S. Chen, 2017c, p. 644)

Specifically, task characteristics include interpreting mode, language pair, interpreting direction, features of speech, features of the speaker, expected response, time on task, preparation time, task criticality and task novelty. Existing interpreting studies have repeatedly demonstrated that these task characteristics can exert considerable influence on interpreters' cognitive load and performance during interpreting, such as syntactical asymmetry (Chmiel & Lijewska, 2019) and word order asymmetry (Ma et al., 2022) in sight translation, terminology uses, syntactical complexity and lack of redundancy in SI (Díaz-Galaz et al., 2015), and language specificity of Chinese-English interpreting in CI (Wang & Zou, 2018). However, only two studies (S. Chen, 2020a and Hu, 2008) have examined these task characteristics in the subtask of CI of note-taking (Chapter 2, Section 2.2.1), with both confirming that the cognitive demand for listening comprehension during the input stage of CI could affect interpreters' allocation of time and effort during the process of note-taking. Digging into the causes behind these findings, one important reason could be the concurrency of listening analysis and note-taking in the first stage of CI. Basically, a substantial input on listening analysis will limit the availability of cognitive resources for note-taking. Therefore, when the source speech contains many difficult lexes and complex syntactic structures, interpreters will need more time to decode the information and thus will postpone note-taking. Moreover, while producing notes, "note takers must maintain an active representation of what they are hearing in order to get sufficient time to exploit and to transcribe a portion" (Piolat et al., 2005, p. 297). This means that interpreters will have to constantly refresh and activate the information they store in their WM until they finish note-writing (Piolat, et al., 2005). Research has found that people's short-term memory for verbal resources can be significantly influenced by the frequency (e.g., Roodenrys et al., 2002) and length (e.g., Baddeley & Andrade, 1994) of words. For words that people are familiar with, they can easily activate relevant information in their long-term memory (LTM) and store this in their short-term memory, both lexically and sub-lexically (Nimmo & Roodenrys, 2002). However, for unfamiliar words, they may not achieve equal memory strength even when they invest in longer processing time and greater processing effort. Similarly, compared to words containing only one syllable, people perform much worse in recalling the words that contain several syllables from their short-term memory (Baddeley et al., 1975). A combination of these findings suggests that interpreters can have larger memory spans for words with higher frequency and less syllables during the process of note-taking than those with lower frequency and more syllables. Therefore, when taking notes for a source speech that contains many uncommon and polysyllabic words, interpreters will have to either speed up note-writing in order to avoid the decay of memory or give up note-taking to retain the source information solely by their short-term memory. In a word, listening analysis and note-taking are two almost concurrent tasks in the input stage of CI. Therefore, the task characteristics that contribute to the cognitive load of listening analysis can directly decide the availability of cognitive resources left for note-taking.

In addition to task characteristics, environmental characteristics are also important in shaping the "input load" (Johannsen, 1979, p. 4) of an interpreting task. Specifically, environmental characteristics include the physical environment of the workplace, the selection of the equipment, and the visibility of the speaker and the audience. Compared with on-site interpreting, all these environmental characteristics are more difficult to control in RI, where interpreters are not physically in the same room as the speaker and the audience. Researchers have found that the impact of such physical separation in RI goes well beyond physiology. For instance, interpreters frequently report feeling isolated, stressed, anxious and unmotivated (e.g., Mouzourakis, 2006; Roziner & Shlesinger, 2010; United Nations, 2001) in RI because they know that the situation is "not being in control" (Moser-Mercer, 2003, p. 11) of. Increased cognitive load (Moser-Mercer, 2003) and decreased interpretation quality (Roziner & Shlesinger, 2010) also demonstrate that interpreting at a distance can exert detrimental impacts on the process and product of interpreting.

When it comes to the consecutive mode of VRI, the situation is even more complex because interpreters will have to conduct screen-viewing and note-taking at the same time. On the one hand, as their vision of the venue is significantly restricted, interpreters have to expend extra effort to find the visual information they need for interpreting on the computer screen (Mouzourakis, 2006). On the other hand, in consecutive VRI tasks where notes are involved, interpreters have to further distribute a greater amount of overt visual attention and processing capacity to completing note-taking and note-reading tasks. For interpreters who are already working close to saturation during interpreting (e.g., Gile, 1995/2005), such a heavy cognitive demand for visual processing can easily make them experience cognitive overload. Indirect evidence for this visual conflict in VRI can be found in Doherty et al. (2022) where the researchers compared interpreters' eye movements during a remote SI task and a remote CI task in a stimulated police interview. They found that the participants' gaze time on the screen was significantly shorter in the consecutive mode of interpreting than in the simultaneous one; which was caused by the fact that much of the participants' overt visual attention was occupied by off-screen note-taking. Moreover, the remote CI task entailed a higher fixation count, longer MFD and more frequent shifts of overt visual attention on the screen than the remote SI task, implying a higher cognitive load in the consecutive mode of interpreting than in the simultaneous one. Again, the reason for this mode-to-mode difference is attributed to the greater proportion of the interpreters' time and energy that the note-taking occupied. According to the literature, "although participants engaged in note taking to facilitate the interpreting task, they may have had to catch up on the visual events that they missed, at least visually, while looking off screen" (p. 10). Such constant repositioning, while receiving the visual information on the screen and processing the written notes on the notepad, further increases the cognitive load of interpreting, resulting in longer fixations and shifts of viewing in CI than in SI. These findings not only imply that remote CI could be more challenging than remote SI but also illustrate the important role of note-taking in shaping the cognitive load of interpreting in a VRI setting.

Taken together, compared with on-site interpreting, RI can lead to a significant increase in the cognitive load of interpreting: compared to VRI without note-taking, including note-taking into VRI can contribute to a substantial growth in visual and cognitive processing during interpreting. All in all, effective distribution of limited overt visual attention and cognitive resources is essential in ensuring that screen-viewing and note-taking can be conducted fluently during VRI without being impeded by the potential conflicts in visual and cognitive processing.

3.1.2 Interpreter characteristics

In comparison to task and environmental characteristics which decide the input load of interpreting, interpreter characteristics are responsible for determining the operator load during interpreting. S. Chen (2017c) has identified four major interpreter characteristics from previous research on cognitive load and studies on interpreting activities, which includes cognitive abilities (Hoffman, 1997), motivation (Moser-Mercer, 2008), experience (O'Donnell & Eggemeier, 1986) and arousal state (Yerkes & Dodson, 1908) (Figure 3.3). Among these, motivation and the arousal state can vary across different task conditions and environmental settings. For instance, "increased task criticality could motivate an interpreter, putting the interpreter in a better state to marshal cognitive resources" (S. Chen, 2017c, p. 646). However, the enhancement of cognitive abilities and the accumulation of interpreting experience cannot be achieved overnight, because they root in the two essential memory systems of human beings: WM and LTM.

To be specific, the development of cognitive abilities intertwines with the enhancement of interpreting experience because the former includes not only general abilities such as "intellect, knowledge (both general knowledge and topical knowledge), language proficiency, cultural competence, and memory (especially working memory)" (p. 646) but also interpreting-specific skills such as note-taking in CI. The acquisition of these domain-general and domain-specific skills is closely related to the growth of WM and LTM. However, little interpreting research has explored the role of these memory systems in note-taking. Studies on note-taking for general purposes (such as lecture note-taking) have proven that a greater WM capacity is associated with a lighter cognitive load of note-taking (Piolat, 2007), and people with larger WM spans are able to create higher-quality notes than those with smaller spans (Peverly et al., 2014; Piolat, 2007). All of these results suggest a close relationship between WM and note-taking, which could be attributed to the inherently cognitive nature of notes. At first, note-taking requires coordinated use of the two slave systems in WM: the phonological loop for audio information processing, and the visuospatial sketchpad for visual information processing (Baddeley & Hitch, 1974). When taking notes for audio materials, note-takers have to first retain the audio information in their phonological store by subvocal rehearsal (Baddeley & Hitch, 1974), then decode the information with their linguistic and extra-linguistic knowledge (Olalla-Soler, 2018), and finally generate the written notes with the help of the visuospatial sketchpad. When it comes to note-reading, interpreters have to first decode the written notes through their visuospatial sketchpad and then articulate the rendition with the help of the phonological loop. Without a coordinated operation of the two slave systems of WM, interpreters will not be able to produce or decipher notes smoothly in either note-taking stage.

Moreover, as discussed in Section 3.1.1, the WM span decides the amount of information that can be stored and processed at the same time. Therefore, an individual interpreter's WM capacity can directly affect the cognitive load of listening analysis and the cognitive demand for information retention during note-writing. Interpreting studies have repeatedly proven that the accumulation of interpreter training experience can improve WM span (e.g., Dong & Liu, 2016; Hiltunen et al., 2014; Tzou et al., 2012). At the same time, around half of the existing studies deny such beneficial effects of interpreter work experience on WM growth (e.g., Köpke & Nespoulous, 2006; Liu et al., 2004; Padilla, 1995), which finding could be the result of ageing among more experienced interpreters (e.g., Wen & Dong, 2019). Even so, professional interpreters are frequently found to be quicker in note-planning, faster in note-writing and more effective in note rendering than student interpreters (e.g., Andres, 2002; S. Chen, 2022; Hu, 2008). Taken together, these findings imply that the underlying cause for these professional-novice differences in note-taking expertise could also lie in their difference in LTM. According to the theory of expertise (Chase & Simon, 1973), experts can outperform novices because they have stored extensive knowledge of a given domain in their LTM. In other words, LTM decides how many knowledge schemas of interpreting and note-taking can be at their disposal while performing an interpreting task. Although experienced interpreters' WM span stops increasing at a certain age, they could have developed enormous interpreting and note-taking schemas to "counteract the burden on WM during interpreting" (Wen & Dong, 2019, p. 779), thus enabling effective note-taking without heavy cognitive constraints on their WM. Therefore, WM span alone does not decide the expertise of note-taking. Instead, an effective operation of WM and a sustained accumulation of LTM jointly contribute to effective note-taking.

In summary, the task characteristics, environmental characteristics and interpreter characteristics of the cognitive load of interpreting can serve as important factors in the cognitive load of note-taking. The first two types of characteristics determine the input load that is inherited in interpreting and note-taking tasks, and the last type of characteristics decides the actual amount of effort that an interpreter allocates to complete interpreting and note-taking.

3.2 Deciphering the resource demands in note-taking and note-reading processes

Having identified the potential factors of the cognitive load of note-taking, this section looks into the specific resource types that are demanded in note activities. By applying the multiple resource theory (Wickens, 1984), Section 3.2.1 illustrates the resource demands of note production and note reception from visual, cognitive, physical and temporal perspectives. With reference to the cognitive load theory (Pass & Van Merriënboer, 1994), Section 3.2.2 explores how interpreters meet these demands through developing recurrent component skills, non-recurrent component skills, and coordination ability.

3.2.1 Multiple resource demands in note production and note reception

Distinct from other subtasks in CI, note-taking requires not only cognitive processing for note-planning but also physical responses for note-writing. These two subtasks of note-taking frequently overlap with each other during the input phase of CI and they are conducted in parallel with the listening and analysis of the source speech. From this perspective, the multiple resource theory (Wickens, 1984), which is a model for visualizing the coordination and interferences among time-shared tasks, well fits the present study to decipher the constructs of load involved in the process of note-taking.

The core of multiple resource theory lies in a four-dimensional multiple resources model (Figure 3-4), which illustrates how humans' limited resources are allocated across different time-shared tasks in terms of four aspects: stages (perceptual/cognitive vs. response); sensory modalities (auditory vs. visual); codes (visual vs. spatial); and channels of visual information (focal vs. ambient). The efficiency in performing time-shared tasks depends on the interferences underlying the quantitative and qualitative demands for information processing structures involved in these tasks (Wickens, 2002). Such interferences can happen in any stage, sensory modality, code or channel of visual information. When the involved tasks do not overlap in the type(s) of demanded resource, the workload of performing the tasks simultaneously equals the ratio of the "(t)ime required (to perform tasks)/time available" (Wickens, 2002, p. 167). For instance, if the input stage of CI is interpreter-paced rather than speaker-paced, interpreters can take their time to fully comprehend the source speech and then start note-taking. However, if the tasks overlap in the demanded resource types, then the workload of executing the tasks in parallel will also depend on the degree of between-task interference. One example is the great visual conflict that interpreters encounter during a consecutive mode of VRI, where overt visual attention is demanded in both screen-viewing and note-taking. Overall, multitasking is better realized when the structures involved in the time-shared tasks are separate than when they are overlapping (Kantowitz & Knight, 1976; North, 1977; Wickens, 1976).

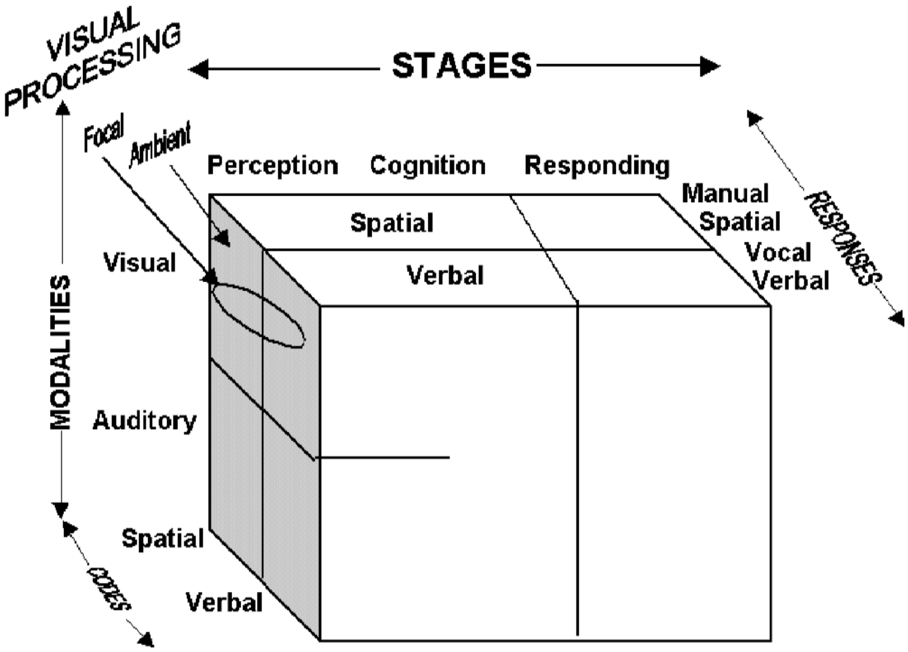


Figure 3-4. The four-dimensional multiple resources model (Wickens, 2002, p. 163)

Figure 3-5 illustrates how the four-dimensional multiple resources model can be applied to note-taking in VRI. During the input stage of VRI, interpreters first perceive the auditory utterance from the speaker and receive the visual information from the screen (*perception*). Then they elicit the message contained in the SL and plan for the notes through verbal processing (*cognition*). During this process, spatial processing is also required to help them arrange the layout of the notes on the notepad, which is found to be essential in indicating the logical development of the source speech (e.g., Chang, 2015; S. Chen, 2020b). Moreover, for interpreters working in a consecutive mode of VRI, they will have to rely on spatial processing to frequently reposition the notes on the notepad and locate the useful visual information on the computer screen (Doherty et al., 2022). Finally, notes are produced through the physical response of handwriting (*response*). In principle, interpreters can initiate the *response* procedure of note production after they have fully completed the *perceptual* and *cognitive* processing involved in visual processing, listening analysis and note-planning. However, as humans' writing speed is only one-tenth of their speaking rate (Foulin, 1995), the ratio of "(t)ime required (to perform tasks)/time available" (Wickens, 2002, p. 167) will be definitely larger than one, which can lead to a cognitive overload during note-taking. Therefore, interpreters have to conduct cognitive processing and physical handwriting at least partially in parallel to make sure they are not being left behind by the speaker. Then, the problem remains of the potential conflicts in the demanded resource types between these two concurrent tasks. Research has found that people tend to rehearse what they are going to write by "speaking" it out in mind (Chenoweth & Hayes, 2003; Locke & Fehr, 1972), and that this "subvocal articulatory rehearsal process (or inner voice)" (Chenoweth & Hayes, 2003, p. 99) during writing relies on the same phonological loop in WM as listening analysis does. Therefore, phonological interferences could happen during the process of note-taking, leading to a trade-off between the cognitive effort expended on listening analysis and that on note-taking. Moreover, as discussed in Section 3.1.2., visual conflicts are inherited in VRI with note-taking, where interpreters have to process visual information on the screen and take notes at the same time. A great amount of overt visual attention paid to the computer screen can possibly lead to a very small number of notes and incomplete memory recall during interpreting. With these identified resource types and possible resource conflicts during note-taking, researchers can collect four types of online data from interpreters to probe into the process of note-taking: the overt visual attention interpreters paid to the computer screen and the notes; the amount of cognitive processing that is involved in note-taking; the amount of physical effort that is invested in note-writing; and the time they distribute to note-planning and note-writing. In summary, the process of note-taking in CI can be explored in terms of visual, cognitive, physical and temporal dimensions.

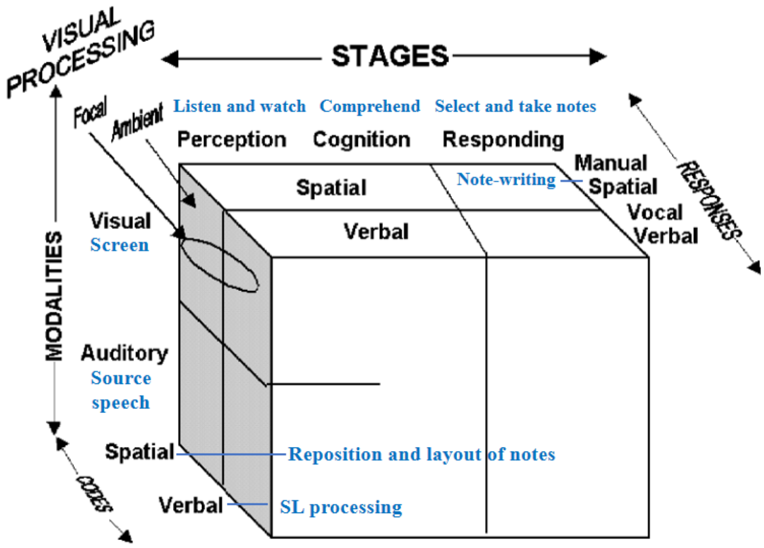


Figure 3-5. An illustration of resources demands during the process of note-taking

As Gumul and Łyda (2007) point out, CI is doubly simultaneous in that it consists of concurrent listening analysis and note-taking during the input stage as well as concurrent note-reading and target-speech production during the output stage. Similar to the first stage of CI, interpreters face both visual (notes) and auditory (their own interpretations) inputs during the second one. On the one hand, they recognize, comprehend and transform the notes into target speech content; on the other hand, they sometimes listen to their own interpretations for the purpose of monitoring the output quality. During the execution of these two subtasks, serious conflicts could happen in the demand for interpreters' limited cognitive and visual resources (see Figure 3-6). Firstly, the notation language differs substantially from normal language in that it consists of "lexis" in varied forms and "sentences" in non-linear structures (Kohn & Albl-Mikasa, 2002) (see Chapter 2, Section 2.1.2). Therefore, it requires a special type of verbal processing that decodes not only actual words but also all sorts of symbolic signs. From the recognition of notes in the early stage of note processing to the integration of note meanings at a textual level in the late stage of note processing, interpreters might get stuck in any of the procedures during note-reading and fail to proceed to target speech organization. In turn, the fact that target speech organization, articulation and monitoring also demand a certain amount of cognitive processing further limits the verbal resources and processing capacity left for note-reading.

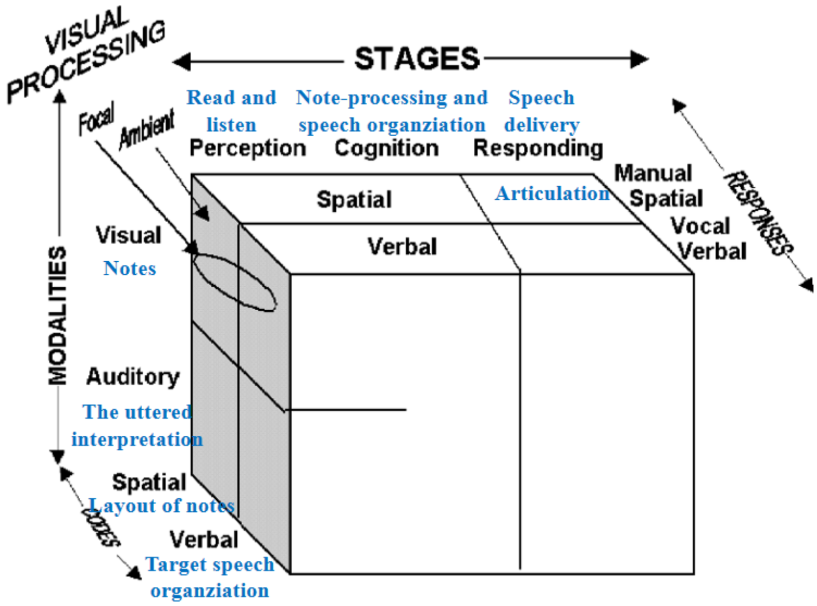


Figure 3-6. An illustration of resource demands during the process of note-reading

Secondly, visual and cognitive conflicts can happen even within the operation of note-reading. According to S. Chen et al. (2021), interpreters frequently revisit different notes in the same note group to decode the logical relations hidden in the layout and generate a holistic understanding of the meaning chunks in the source speech. At the same time, when they are going to finish the translation of the current note group, they would read ahead to the next one, in order to ensure a fluent delivery of the target speech. During this process, interpreters have to distribute their limited overt visual attention and cognitive resources to a large number of notes that are usually organized in non-linear structures (Albl-Mikasa, 2008). In traditional face-to-face CI, they may also have to allocate a part of their attention to keeping eye contact with the audience, if necessary, which further intensifies the conflicts of visual processing. Based on these identified resource types that are demanded during the process of note-reading and target speech production, consideration of the distribution of interpreters' overt visual attention and cognitive effort between these two tasks can be a valuable perspective for researchers to probe into the output stage of CI.

3.2.2 Parallel operation of controlled processing and automatic processing

Given the complex resource demands in note-taking and note-reading, the question remains to be how to meet these various demands with a limited pool of processing capacity (Kahneman, 1973). The dual-process theory from psychology studies can provide implications for this question from an attentional perspective. According to the dual-process theory, human “behavior is determined by the interplay of automatic and controlled processing” (Barrett et al., 2004, p. 553; see also Fiske & Neuberg, 1990 on person perception; Wenzlaff & Wegner, 2000 on mental control; and Baumeister & Heatherton, 1996 on self-regulation). Whereas automatic processing is a “default mode of processing” (Barrett et al., 2004, p. 554) that is difficult to be ignored or altered, controlled processing “arises from the central executive aspect of working memory…and occurs when attention is applied in a goal-directed, top-down, or endogenous fashion (p. 555). Comparatively speaking, automatic processing is attention-free and reliant on one’s LTM, while controlled processing is capacity-limited and executed through one’s WM. Therefore, while completing cognitive complex tasks such as note-taking in interpreting, the more automatic processing is involved, the less cognitive load the interpreter would experience during note-taking.

To transform controlled processing into automatic processing, the key lies in transforming non-recurrent component skills into recurrent component skills (Figure 3-7). According to Pass and Van Merriënboer (1994), recurrent component skills can be applied throughout a series of similar tasks, but non-recurrent component skills vary considerably across different tasks. For instance, interpreters can use “○” to represent “person” in their notes across different interpreting tasks, but they have to re-decode the information in the source speech every time when the speaker speaks. Specifically, recurrent component skills are developed through rule automation, which is a quantitative process that requires a considerable amount of practice. Hence, “automation” is repeatedly emphasized as the key to developing “expertise” in a given field in cognitive psychology (e.g., Ericsson & Charness, 1994; Pass & Van Merriënboer, 1994). By contrast, nonrecurrent component skills are acquired through schema construction, which is more of a qualitative process that aims at facilitating the execution of controlled processing in similar tasks. In general, after rounds of schema application in similar problem-solving activities, schemas can be advanced and applied to more complex situations.

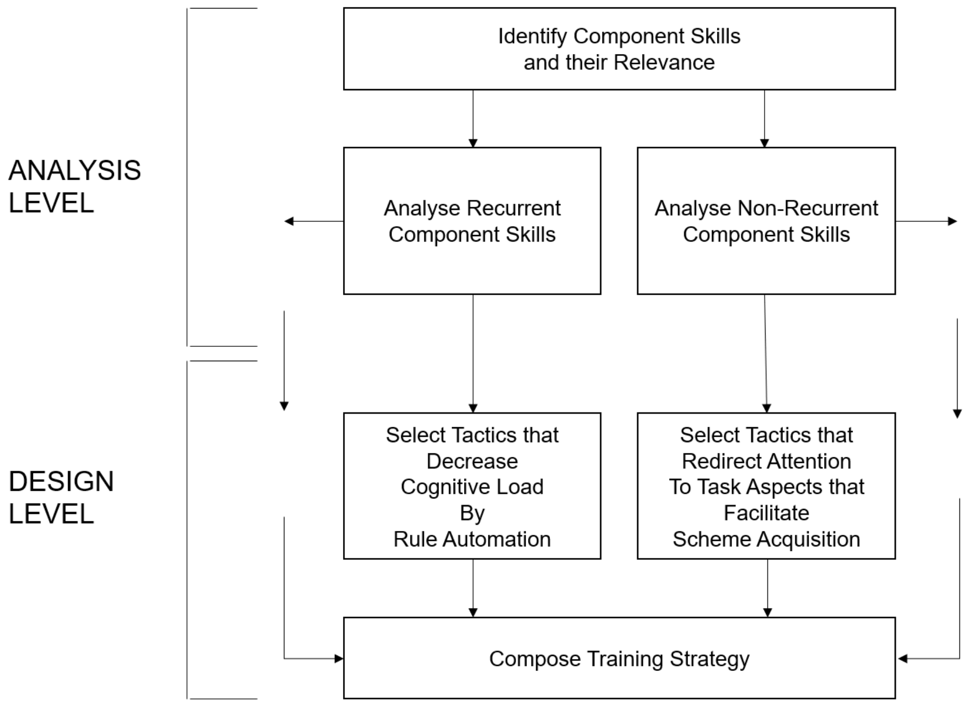


Figure 3-7. The instructional design model adapted to the cognitive load theory (Pass & Van Merriënboer, 1994, p. 360)

Similarly, researchers from the field of note-taking for general purposes have identified two types of skills that match the concepts of recurrent component skills and non-recurrent component skills in note-taking: the domain-specific, basic skill of handwriting; and the higher level cognitive skills of language comprehension and background knowledge accessibility (Peverly et al., 2014). Peverly et al. (2014) have summarized that the efficiency of note-taking:

> …depends on the simultaneous (in parallel) activation of a hierarchy of domain-specific and higher order cognitive skills, within a limited capacity working memory. Domains-specific basic skills must be sufficiently fluent or automatic so that most if not all of the limited space in working memory can be used for the application of the higher level cognitive skills needed to produce successful academic outcomes. Once lower level skills are sufficiently fluent or automatic, the quality of the product is strongly related to the quality of the higher level skills (p. 2).

Following Peverly et al. (2014), handwriting is the recurrent component skill that is shared by all note-taking tasks for different interpreting scenarios. Listening analysis and the application of knowledge schemas in LTM together constitute the non-recurrent component skills. In existing models and theories, regardless of whether for note-taking in CI (e.g., Gile, 1995/2009), note-taking in general (e.g., Ladas, 1980) or common writing tasks (e.g., Scardamalia & Bereiter, 1982), researchers have put much emphasis on the non-recurrent component skills that center on the higher-level cognitive processing, and have paid little attention to the application of the recurrent component skill of handwriting. It appears that researchers assume that handwriting is an easy skill that has been automated by all note-takers. However, more often than not, interpreters report great difficulties in following the speaker during note-taking (e.g., Arumí Ribas, 2012; Her, 2001) and are found to rush to take notes for the purpose of noting more without much listening comprehension (e.g., Albl-Mikasa, 2006; S. Chen, 2022). This implies that the striking gap between people's writing and speaking speed (around 1:10) (Foulin, 1995) still burdens interpreters heavily during note production. As discussed earlier, note-writing requires information retention through subvocal rehearsal in interpreters' WM. Therefore, the more note-writing occupies interpreters' processing capacity, the more cognitive constraints are posed on the completion of listening analysis and LTM retrieval.

Moreover, the concurrent application of recurrent and non-current component skills in note-taking requires effective coordination. In writing studies, Olive and Kellogg (2002) have shown that the coordination between transcription and composition in writing tasks can be cognitively very demanding. Specifically, these researchers measured adults' reaction time to auditory probes in three task conditions: "(1) copying in longhand a prepared text (transcription), (2) composing a text and pausing handwriting for longer than 250 msec (composition), and (3) composing and currently handwriting (transcription + composition)" (p. 594). By calculating the reaction time that the subjects needed to complete the secondary task of sound detection across the three situations, the researchers compared the involved level of cognitive processing under the three task conditions. The results show that, for adult writers, the interference was greatest in the compound task, medium in the composition-only task, and smallest in the transcription-only task, suggesting that the synchronization of composition and transcription is cognitively very demanding. Hence, aside from enhancing the efficiency of applying recurrent and non-recurrent component tasks, interpreters should also pay close attention to their coordination ability, which is essential in reducing the cognitive load of note-taking and interpreting.

3.3 Exploring the relationships between note production, note reception and interpretation quality

As discussed, listening to and analysis of the source speech, production and reception of the notes as well as generation and delivery of the target speech are all integral components of a CI task, which means that they should not be discussed separately. This section aims to explore how interpreters should make maximal use of their limited processing capacity in the two stages of note activities to enhance their interpretation quality. Discussions will firstly be devoted to exploring the relationships between the two stages of note activities in CI, namely note-taking and note-reading (Section 3.3.1), and then to the complex associations between interpreters' note-taking behaviour and interpreting performance (Section 3.3.2).

3.3.1 Note-taking effort and note-reading effort

The relationship between the effort of note-taking and the effort of note-reading varies according to the interpreter's perception of the nature of note-taking (Figure 3-8). As introduced in Chapter 2, Section 2.1, notes can serve as the storage of "ideas" or the storage of "words". The former centres on using notes to record the sense that interpreters have extracted from the source speech (e.g., Seleskovitch, 1975), whereas the latter follows source expressions closely and aims at reminding interpreters of the detailed information of the source speech (e.g., Albl-Mikasa, 2008). Accordingly, these two types of perception of notes lead to two general strategies of note-taking: "a higher-level comprehension strategy" and "a lower-level comprehension strategy" (Albl-Mikasa, 2006, p. 11). The former strategy requires interpreters to devote a great amount of cognitive effort to complete listening comprehension during the input stage of CI, and notes only serve as a physical repository of what is comprehended. By contrast, the latter does not pose high requirements for listening comprehension but demands much physical effort of handwriting. The higher-level comprehension strategy is widely encouraged in interpreter training programs and note-taking guidelines (e.g., Mackintosh, 1985; Rozan, 1956; Seleskovitch, 1975), because it follows the standard conduct of meaning-based interpreting (e.g., Dam, 2001). By comparison, the latter strategy can entail great risks of form-based interpreting and is thus much less encouraged in interpreter training (Albl-Mikasa, 2008).

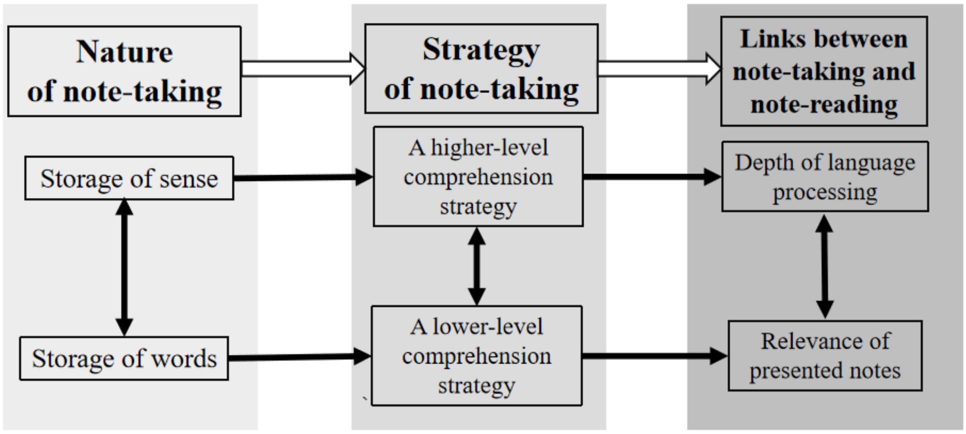


Figure 3-8. A graphic illustration of the association between the nature of note-taking and the adoption of note-taking strategies, and the impact of the latter on the relation between note production and note reception

While adopting a higher-level comprehension strategy, little processing capacity is left available for note-taking and only a small number of notes can be produced (Albl-Mikasa, 2006). Interpreters can face great pressure during note-reading because they have to recall the source information with a very limited number of notes. However, according to levels of processing model (Craik & Lockhart, 1972), humans' memory trace is "a by-product or record of normal cognitive processes such as comprehension, categorization, or discrimination" (Lockhart & Craik, 1990, p. 90). This means that a deeper processing level of the SL during note-taking can lead to a stronger memory trace of the source speech during note-reading. In this case, a higher-level comprehension strategy during note-taking can decrease the demand for memory retrieval during note-reading (Albl-Mikasa, 2006), leading to a trade-off between the effort interpreters devote to the two noting stages. This assumed trade-off between note production and note reception can be further grounded by the concept of an episodic buffer, which is introduced by Baddeley (2000) in his multi-component WM model. The episodic buffer is a limited and temporary storage system that binds information received from different sources "by using a common multi-dimensional code" (p. 421). According to Van Dijk and Kintsch (1983), a list of components from the memory system needs to be activated to enable text comprehension: perceptual features, linguistic features, propositional structure, macrostructure, situation model, control structure, goals, lexical knowledge, frames, general knowledge, and episodic memory for prior text (see Van Dijk & Kintsch, 1983, p. 347). Therefore, if interpreters intend to adopt a higher-level comprehension strategy of note-taking, they need to activate, consult and store a series of linguistic and extra-linguistic information in their episodic buffer to comprehend the source speech. The greater the amount of information that can be integrated and stored in the episodic buffer, the more thorough the source speech comprehension can be achieved. When it comes to note-reading, as long as the "multi-dimensional code" (p. 421) can be retrieved and recognized from their notes, interpreters will be able to activate the whole set of information that is bonded within the codes. Under this situation, a negative relationship can be observed between the note-taking effort and the note-reading effort

On the other hand, a lower-level comprehension strategy requires a smaller amount of information to be integrated into the episodic buffer and involves less SL processing. Although it entails a higher possibility of form-based interpreting (e.g., Dam, 1998; Gran, 1989), it is what interpreters apply most commonly in their interpreting practice (e.g., Albl-Mikasa, 2008; S. Chen, 2022; Dam, 1998). This strategic choice of interpreters is considered to be caused by the great demand for performing listening comprehension and note production at the same time during the speaker-paced input stage of CI (e.g., Albl-Mikasa, 2008; S. Chen, 2020a). Since interpreters usually find it difficult to achieve in-depth source speech comprehension during the input (e.g., Albl-Mikasa, 2008; Alexieva, 1994; Arumí Ribas, 2012), they choose to focus on note production to ensure that there are enough notes for them to recall the source contents during note-reading. Albl-Mikasa (2008) finds this note-taking strategy to be reasonable, by applying Relevance Theory to analyse the intrapersonal communication between the interpreter during note-taking and the same interpreter during note-reading (see Chapter 2, Section 2.1.2). However, notes created with this lower-level comprehension strategy demand much note-reading effort. During the output stage of CI, interpreters are expected to deliver a complete and fluent interpretation. With a great number of notes at hand, they have to decode, integrate and translate the meaning of these notes within seconds. Any hesitation in note recognition or misinterpretation can thus lead to inaccurate translations or disfluencies in speech delivery. From this perspective, with a lower-level comprehension strategy of note-taking, an increase in the effort of note production leads to an increase in the effort of note reception. In this situation, there will be a positive relationship between the effort of note-taking and that of note-reading.

In principle, regardless of a higher- or lower-level comprehension strategy of note-taking, notes can help interpreters to enhance interpretation quality so long as they can facilitate source speech recall during the output stage of CI. Based on the adopted note-taking strategy, the effort of note-taking and effort of note-reading, together, will present an accordingly positive or negative relationship.

3.3.2 Note-taking behaviour and interpretation quality

Note-taking has long been considered as an important factor in CI performance (e.g., Abuín González, 2012; Gile, 1995/2009; Gillies, 2017). The following discussion in this section introduces three models that can be referred to while exploring the relationship between interpreters' note-taking behaviour and interpretation quality.

Based on O'Donnell and Eggemeier's (1986) model of workload and task performance, S. Chen (2017c) proposes a hypothetical relationship between the cognitive load of interpreting and interpreting performance. As shown in Figure 3-9: in Region A, interpreters can maintain high interpretation quality because their cognitive load is at a very low level; in Region B, interpreting performance starts to decline with the increase in the cognitive load of interpreting; and in Region C, interpreters' performance stays poor as their WM is already saturated. This hypothetical relationship between the cognitive load of interpreting and interpreting performance can also be applied to a note-taking context in two ways. Firstly, note-taking and note-reading are two subtasks at the input and output stages of CI. Therefore, the cognitive load of the note activities can contribute to that of the whole interpreting task, which further exerts impacts on interpreting performance. Secondly, note-taking and note-reading per se are cognitively demanding activities. Hence, the cognitive load of the two noting processes and the "performance" of the notes (i.e., the effectiveness of the notes in target speech production) could also present such a staged relationship. Specifically: in Region A, interpreters can take effective notes without bearing much cognitive load of note-taking or note-reading; in Region B, the effectiveness of notes decreases as the cognitive load of note-taking or note-reading increases; and in Region C, the cognitive load of either noting process is too high for interpreters to achieve effective note-taking. While applying this model to note activities in CI, it is important to notice that notes are only intermediate products of CI and will not be presented to the audience. Therefore, interpreters can flexibly shift between more and less demanding note-taking strategies during the input phase of CI to maintain their overall cognitive load of interpreting within an acceptable range. However, the manipulation of note-taking effort during the input stage of CI can lead to different note-reading demands during the output (see

Section 3.3.1), which can be further reflected in the accuracy and fluency of the final interpreting product.

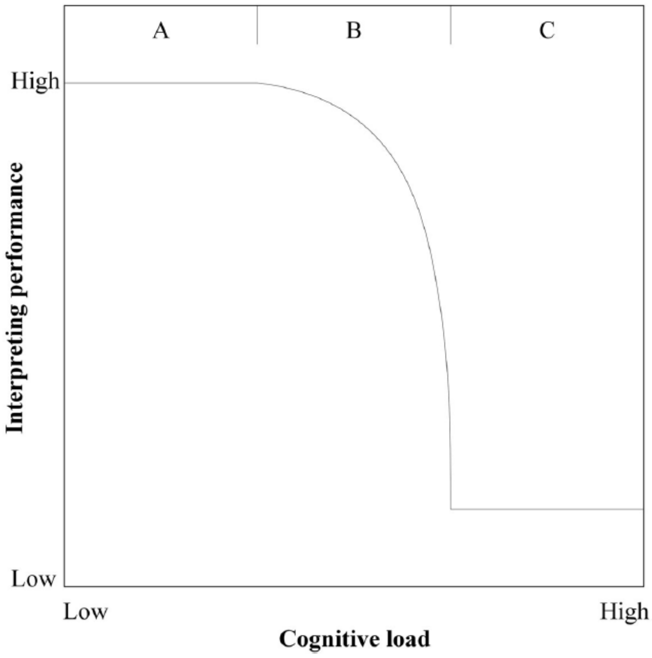


Figure 3-9. A hypothetical relationship between the cognitive load of interpreting and interpreting performance (S. Chen, 2017c, p. 650)

Different from S. Chen (2017c) where interpreting performance is discussed with reference to the overall cognitive load of interpreting, Gile's (1995/2009) effort model looks into the different subtasks of interpreting. In the effort model of CI, note-taking (referred to as note production, NP for short) is considered a subtask in the comprehension phase that competes for interpreters' limited cognitive resources with other subtasks including listening and analysis (L), short-term memory operations (M) and coordination (C). Accordingly, note-reading (NR) is recognized as a subtask in the reformulation phase of CI that accompanies speech reconstruction from memory (SR), target speech production (P) and coordination (C). As computer technology is increasingly popular in interpreting practice, Gile (2021) introduces another potential subtask that can be required during the process of interpreting, man-machine interaction (MMI); and the "manipulations of screen, keyboard and other non-automatic human-machine interactions" (p. 35) can all be referred to as the effort of MMI. Applying this notion to VRI with note-taking where interpreters have to constantly shift between the computer screen and the notepad to produce and read their notes, they also bear an extra cognitive load of MMI during interpreting. Therefore, the effort model for VRI with note-taking should be:

1) The comprehension phase: L + M + NP + C + MMI

2) The reformulation phase: NR + SR + P + C + MMI

According to Gile (1995/2009), to achieve high interpretation quality, interpreters need to fulfil two conditions. Firstly, the total cognitive load of interpreting should not surpass the interpreter's total processing capacity. Secondly, the cognitive load of each individual subtask of a CI task should not exceed the interpreter's processing capacity at any given time. From a note-taking perspective, this means that interpreters should maintain their effort of NP and NR at a minimal level. At the same time, they need to ensure that there is an adequate number of notes that can help them to retrieve source speech memory during the output stage of CI. To decrease the cognitive load of interpreting, Gile (2020) proposes a concept called language availability, which refers to "the time it takes to find/understand a word/linguistic structure" (p. 19) in language comprehension and language production. Basically, the higher the availability of a word in one's LTM, the easier the word can be activated in one's WM. Figure 3-10 visualizes how two interpreters with different levels of language availability comprehend the source speech. During the first time zone (t1), the high availability listener has finished the processing of two words while the low availability listener has only completed that of one. When the speaker says the seventh word in t2, the high availability listener has finished the processing of six words but the low availability listener has only processed the first two. When it comes to t3, the high availability listener can continue to follow the speaker but the low availability listener's WM is already saturated. This example clearly illustrates the gap between interpreters' processing capacity and the cognitive demand of listening in different situations.

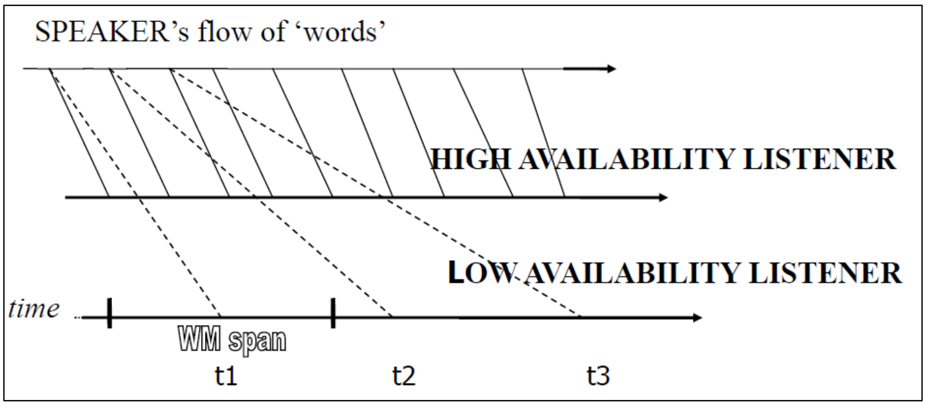


Figure 3-10. The gravitational model of language availability in Gile (2020, p. 22)

This notion of language availability can be applied to note-taking in two ways. Firstly, during the input stage of CI, increased language availability allows interpreters to quickly understand the source speech and proceed to the note-taking procedure; while, during the output stage of CI, high language availability helps interpreters to quickly transform the intended meaning of notes into the TL and leave cognitive resources for target speech monitoring. Gile (1999) also points out that, for unskilled interpreters, they have to make extra effort to understand the source speech, decide what to write, how to write, and control the writing operation. This great time lag between source speech delivery and note-taking completion can lead to a saturated WM, which further undermines interpreting quality. Secondly, the notation language is also a special kind of language that can be applied to the concept of "language availability". Specifically, if an interpreter is equipped with a high level of notation language availability, then he or she can flexibly utilize the note-taking resources (such as symbols and abbreviations) stored in the LTM. This can be reflected as quick and effortless note-planning during note production as well as fast and accurate note recognition, integration and translation during note reception. Based on these discussions, to conduct interpreting successfully, interpreters should be equipped with high language availability not only in the SL and the TL of the interpreting task but also that of the notation language.

Despite the implications of Gile's Effort Model for the relationship between note-taking behaviour and interpretation quality, there is an issue that it has not touched on much in the model: the "mechanical" (Gile, 2020, p. 13) effort of note-writing. As discussed in Section 3.2.1, the heavy physical demands of handwriting can also impose great cognitive load on the brain, which will further affect interpretation quality. In studies on note-taking for general purposes, researchers have investigated a series of factors that can contribute to one's note-taking ability, among which handwriting ability is found to be essential in affecting note quality. In a study with 85 undergraduates (Peverly et al., 2007), researchers assessed the correlations between note quality, on the one hand, and participants' transcription fluency (letter fluency[3] and compositional fluency[4]), spelling skills, verbal WM[5] (VWM) and ability to identify main ideas, on the other hand. The results show that letter fluency was the only effective predictor of note quality, and the quality of notes was the only effective predictor of test performance (summary writing). By contrast, VWM was not as expected an effective predictor of note quality. In a later study, Peverly et al. (2014) tested students with another series of tasks that investigate the two types of skills in note-taking: the domain-specific basic skill related to handwriting, and the higher-level cognitive skills such as listening comprehension. Specifically, the adopted tasks assessed the students' handwriting speed (writing letters alphabetically in 45 seconds), fine motor fluency (moving fingers as required), speed access to verbal codes (naming the presented letters in 15 seconds), language comprehension ability (reading tests), WM capacity (listening span test), executive control of attention (Color Naming, Word Reading, Color-Word Interference, Color-Word Interference with Switching), sustained attention (recalling items ended up with 55 in a 10-min listening task), written recall performance (writing a summary of the lecture), and note quality (taking and explaining notes correctly). The results from 71 undergraduates show that note quality was significantly correlated with their handwriting speed, WM capacity, and speed access to verbal codes in a positive way (see Figure 3-11). Written recall performance was only positively correlated with handwriting speed (Figure 3-11). Having a closer look at the observed correlations, the researchers found that faster verbal access was correlated with a higher handwriting speed, and quicker handwriting could release more space for WM to execute higher-level cognitive processes. These findings indicate that the utilization of higher-level skills during note-taking might affect note quality through the execution of lower-level skills of handwriting. In other words, skills at a lower level mediate the impacts of skills

---

[2] Letter fluency is measured as the number of letters that can be written alphabetically in capital letters in 30 seconds.
[4] Compositional fluency is measured as the number of logical sentences that the participants can compose based on three given words and a picture of a stimulus in seven minutes.
[5] Verbal WM is measured by the listening span task developed by Daneman and Carpenter (1980), which requires participants to make semantic judgments after listening to a sentence and recall the last words of all the sentences in the same set.

at a higher level during the process of note-taking. Therefore, the physical effort of note-writing can contribute substantially to the cognitive load of note-taking, and the fluctuance in the cognitive load of interpreting can further lead to enhanced or undermined interpretation quality.

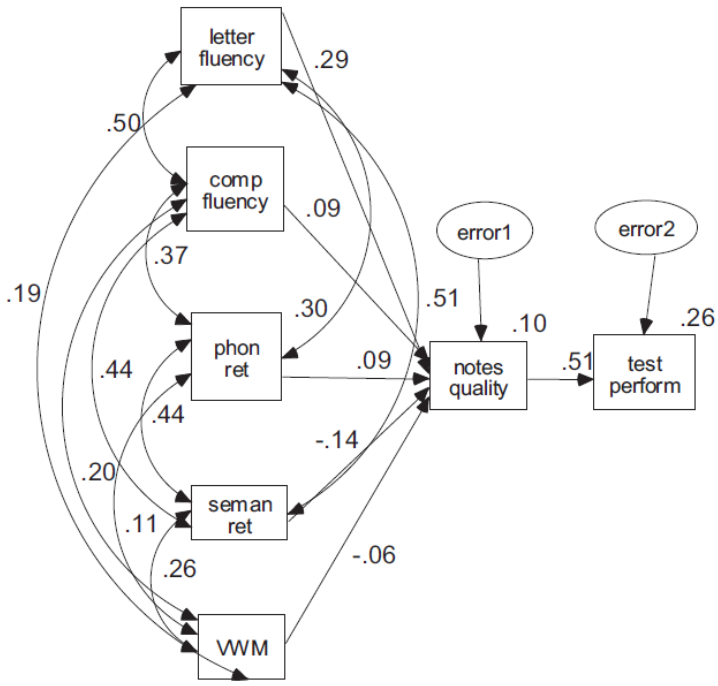


Figure 3-11. Correlations among letter and composition (comp) fluency, verbal WM (VWM), phonetic retrieval (phon ret), and semantic retrieval (seman ret) to notes and the relationship of notes to test performance (test perform), in Peverly et al., (2007, p. 166)

3.4 Summary

This chapter has explored every component of note activities in CI, from source speech comprehension to target speech quality, by examining and illustrating relevant theoretical models and concepts. It firstly introduced the factors that can contribute to the cognitive load of note-taking, including task characteristics, environmental characteristics and interpreter characteristics. It then illustrated the resources and skills that note-taking and note-reading demand with reference to the multiple resource model and the cognitive load theory. Finally, it probed into models that can be consulted while exploring the relationships between the effort of note-taking and that of note-reading, as well as the associations between interpreters' note-taking behaviour and interpretation quality.

In summary, task characteristics and environmental characteristics determine the objective difficulty level of note-taking, and interpreter characteristics decide the actual amount of effort that an interpreter devotes to note production and note reception. Complex source speech and a remote mode of CI can seriously limit the number of cognitive resources that are available for note-taking and note-reading, leading to a substantial increase in an interpreter's cognitive load of interpreting. Therefore, an effective allocation of WM resources and a long-term accumulation of LTM knowledge schemas are both important in achieving effective note-taking.

In terms of the resource demands for note activities in CI, it is important to first recognize that CI consists of concurrent listening comprehension and note-taking during the input stage as well as concurrent note-reading and target-speech production during the output. Therefore, the note activities in CI can be well applied to the multiple resource theory, which is designed for identifying the resource demands in concurrently-performed tasks. Overall, note-taking in VRI is identified with four types of resource demands: overt visual attention (screen viewing and note processing), cognitive effort (listening analysis and note-planning), physical effort (note-writing) and temporal management (time to make note decisions). Note-reading is recognized as having two sorts of resource demands: overt visual attention (note processing and potential eye contact with the audience) and cognitive effort (early- and late-stage note processing). Hence, the processes of note-taking and note-reading can be explored from visual, cognitive, physical and temporal perspectives.

At the same time, the demands of note production and note reception heavily depend on interpreters' approach to note-taking. When interpreters adopt a higher-level comprehension strategy during the input stage of CI, much of their processing capacity is occupied by listening and analysis of the source speech. This can lead to a more thorough understanding of the speech but a very limited number of notes. When interpreters choose to adopt a lower-level comprehension strategy of note-taking, only a surface level of language processing is achieved during the input stage of CI. Therefore, the main problem for interpreters during the process of note-reading would be making up for the insufficient source speech comprehension by quickly digesting a large number of notes. Based on Craik and Lockhart's (1972) levels of processing model and Baddeley's (2000) multi-component WM model, a higher-level comprehension strategy of note-taking is expected to entail stronger memory links during the process of note-reading. In this case, the effort of note-taking and note-reading would present a negative relationship with each other. On the other hand, if notes are produced with little source speech comprehension, the relationship between the effort interpreters devote to the two noting stages can be a positive one.

The key to meeting these demands of note activities in CI remains to be increasing the proportion of automatic processing and decreasing that of controlled processing. By automating the basic skill of handwriting, interpreters not only reduce their physical load of note-writing but also release the cognitive resources that are occupied for the subvocal rehearsal. By facilitating the execution of higher-level cognitive skills such as language comprehension and language production, interpreters can considerably decrease the time and effort that are demanded during listening analysis and target speech production, thus focusing more on the production and reception of notes in interpreting. Research on general note-taking has proven that handwriting speed is a decisive factor in written recall tasks, and that letter fluency is the only effective predictor of note quality. At the same time, since the participants' performances in the tests of lower- and higher-level skills correlate with each other significantly, this indicates that the execution of the lower-level skill of handwriting can affect the operation of higher-level cognitive processing during note-taking. Altogether, both the mechanical and cognitive aspects of note-taking must be taken into consideration when examining interpreters' note-taking behaviour in interpreting.

Overall, note activities in CI can be associated with interpretation quality both directly and indirectly. On the one hand, note-taking and note-reading per se are two cognitively taxing activities that can cause a cognitive overload of note processing. Poor note-planning and slow note-writing can result in inaccurate and incomplete notes. Ineffective recognition, integration and translation of notes directly affect the accuracy and fluency of interpreting. On the other hand, the cognitive load of note-taking and note-reading contributes to the cognitive load during, respectively, the input and output of interpreting. Therefore, note-taking behaviour can also be associated with interpretation quality via its impacts on the overall cognitive load of interpreting.

Chapter 4: Research Methodology

This chapter presents three aspects of the research methodology: the manipulation of source speech difficulty; the measurement of the note-taking and note-reading effort; and the analysis of note patterns and interpretation quality. Since there can be discrepancies between the designed difficulty level and the perceived difficulty level of the interpreting material, both objective and subjective assessment measures are included to determine source speech difficulty. The specific measurements and results are introduced in Section 4.1. In addition, apparatus selection and indicator adoption are introduced in Section 4.2. The apparatus mainly includes an eye-tracker and a digital pen. The indicators consist of a range of eye-tracking and pen-recording measures that point to the overt visual attention, cognitive effort, physical effort and temporal management during the note-taking and note-reading processes. In addition, a descriptive method of analysis of the participants' note choices and a summative assessment approach to their interpretation quality are introduced in Section 4.3. Finally, Section 4.4 provides a summary of this chapter.

4.1 Assessment of source speech difficulty

The source speech in interpreting studies is sometimes referred to as the "source material" (e.g., Liu & Chiu, 2009), the "stimulus" (e.g., S. Chen, 2017a) or the "source text" (e.g., Szabó, 2006). The term "source speech" is used in the present study because, except for its written nature as a text, the source material in CI is also an orally-delivered speech. As introduced in S. Chen's (2017c) construct of cognitive load in interpreting, "features of the speech (formal features such as length, speed, and scripted/spontaneous speech, and content features such as topic, lexical and syntactic complexity)" (p. 644) are essential task characteristics that contribute to the difficulty of an interpreting task. Therefore, the assessment of source speech difficulty should concern two aspects: "what is written" and "how it is delivered". In the present study, the features of the speech are controlled in the selection and edition of the speech videos (see Chapter 5, Section 5.2.1), and the content features are examined with readability scores and idea density indices. In addition, subjective ratings including expert judgment and the participants' feedback in the pilot study are collected and evaluated to ensure that the levels of the designed (objective) difficulty and of the perceived (subjective) difficulty of the source speech segments are consistent.

4.1.1 Objective measurements

Readability and idea density are adopted as the objective indicators of source speech difficulty in the present study. Readability refers to a group of indices that quantify "the ease with which the text is likely to be read and comprehended" (Jensen, 2009, p. 63), and most of the calculations of these measures are rooted in word frequency and sentence length (Klare, 1963). These indexes have been adopted in over 1000 published studies and are proven to be effective in indicating text complexity[6] (DuBay, 2004). Among the 200 readability formulae, seven are most frequently adopted (Table 4-1) in translation (e.g., Cui & Zheng, 2020, Hvelplund, 2019) and interpreting (e.g., Liu & Chiu, 2009; Liu et al., 2004) studies. This is because these formulae concern a variety of "countable properties (of a text), such as characters, syllables, words, and/or sentences combined with one or more constants" (Jensen, 2009, p. 64) that can evaluate the complexity level of a text from multiple aspects. For instance, measures related to syllable count can help researchers to control the effects of speaking duration, where recall accuracy decreases as the number of syllables in words increases (Baddeley et al., 1975). Measures concerning word length and sentence length, both of which have been found to be negatively correlated with human memory trace (Baddeley & Andrade, 1994) and ear-voice span in interpreting (Timarová et al., 2014), are also included in the assessment. Moreover, these readability indices have been tested as effective in reflecting the difficulty level of comprehending and interpreting a text. For example, the Flesch Reading Ease index has been correlated with measures such as cloze and teacher judgment with the coefficients ranging between 0.64 and 0.70 (Harrison, 1980, cited in Liu & Chiu, 2009), and negatively related to interpreters' performance in simultaneous interpreting (Liu et al., 2004). Based on these references, the present study adopts the seven most frequently adopted readability indices (Table 4-1) in translation and interpreting studies to control for the objective difficulty level of the source speech in the investigation.

Table 4-1. The seven most frequently-adopted readability indexes[7] in Translation and Interpreting Studies

[6] Text complexity refers to the objective and relative difficulty of a text based on certain criteria, whereas text difficulty points to the subjective and perceived difficulty for a text reader (Jensen, 2009).
[7] The formulas and the interpretation of the results are summarized based on the information presented on https://readabilityformulas.com/.

Indices and formulas	Result interpretations
Flesch Reading Ease: 206.835 - (1.015 * average sentence length) - (84.6 * average number of syllables per word)	90-100: very easy 80-89: easy 70-79: fairly easy 60-69: standard 50-59: fairly difficult 30-49: difficult 0-29: very confusing
Gunning Fog: 0.4 * (average sentence length + percentage of hard words)	>12: too hard 7-8: ideal
SMOG: 3 + square root of polysyllable count	The result approximates a U.S. grade level to understand the text.
Coleman-Liau: 00588 * the average number of letters per 100 words- 0.296 * the average number of sentences per 100 words - 15.8	The result approximates a U.S. grade level to understand the text.
Flesch-Kincaid: (0.39 * average sentence length) + (11.8 * average number of syllables per word) - 15.59	The result approximates a U.S. grade level to understand the text.
LIX: number of words/number of periods (defined by period, colon or capital first letter) + ((number of long words (more than 6 letters)*100) / number of words	20-25: very easy 30-35: easy 40-45: medium 50-55: difficult 55-60: very difficulty
Automatic Readability Index: 4.71 * (characters / words) + 0.5 * (words/sentences) - 21.43	The result approximates the age needed to understand the text (can be transferred to the U.S. grade level).

As presented in Table 4-1, among the seven readability indexes, the Flesch Reading Ease index and the LIX index return numerical scores. A higher score in the former index indicates less difficulty in text comprehension, whereas the second index indicates the opposite. The rest of the indices, including the Gunning Fog index, the SMOG index, the Coleman-Liau index, the Flesch-Kincaid index and the Automatic Readability Index (ARI), return the years of education or the U.S grade level that readers must have completed to comprehend a text. Hence, a higher grade level indicates a higher difficulty level of comprehending the text. Results on the readability of the source speech segments used in this thesis study are presented in Chapter 5, Section 5.2.

In addition to readability indices, another measure that is adopted in the present study to assess the objective difficulty level of the source speech is idea density. Idea density is a measure that aims to approximate the extent to which propositions are closely distributed in the source speech. The core of this measure lies in the concept of proposition (Van Dijk & Kintsch, 1983), which is the smallest meaning unit in a text (1998). “(T)he main verb and all of its arguments (subject, object, indirect object, etc.)

are one proposition. Additional descriptive elements, such as adjectives, adverbs, and qualifier phrases are additional propositions" (Turner & Greene, 1977, as cited in Brown et al., 2008). A specific example is provided below:

The old gray mare has a very large nose.
breaks up into:
(HAS, MARE, NOSE)
(OLD, MARE)
(GRAY, MARE)
(LARGE, NOSE)
(VERY, (LARGE, NOSE)) (Brown et al., 2008, p. 541)

For this sentence, the idea density would be 9/5, which is obtained by dividing the total number of the words (9) by the total number of the propositions (5). Similarly, the idea density of a text is the result of dividing the count of words by the count of propositions (Brown et al., 2008). A computer program called Computerized Propositional Idea Density Rater 5 (CPIDR 5) is adopted in the present study to measure the idea density of the speech transcripts. The program is designed to automatically assign the part-of-speech tags to the source speech units (Figure 4-1), identify the potential propositions, and calculate the number of real propositions after applying a series of "later rules" (Brown et al., 2008, p. 542). For instance, one of the later rules is that "either…or" should be counted as one proposition rather than two. These rules are designed to ensure that no "false" propositions are counted and no "true" propositions are missed.

The Main Part-of-Speech Tags Used by MontyLingua and CPIDR

Tag	Interpretation
.	sentence-ending punctuation
CC	coordinating conjunction
CD	cardinal number
DT	determiner
IN	preposition, except *to*
JJ, JJR, JJS	adjective (positive, comparative, superlative)
MD	modal verb
NN, NNS	noun (singular, plural)
PDT	predeterminer
POS	possessive *'s*
PP$, PRP$	possessive pronoun
RB, RBR, RBS	adverb (positive, comparative, superlative)
TO	to (preposition or infinitive)
VB, VBZ, VBD, VBN, VBG, VBP	verb (various forms)
WDT, WP, WPS, WRB	interrogatives and relatives (e.g., *which*)

Note—For the full set used by MontyLingua and CPIDR, see Santorini (1995).

Figure 4-1. Main tags used in CPIDR to identify different parts of speeches based on Santorini (1995)

When all the propositions are identified, the left side of the program's interface will present the number of propositions (ideas), the number of words and the result for idea density (Figure 4-2). On the right hand of the interface, it will present the details of each proposition and the part of speech that is assigned to each word for manual checking. Results concerning the idea density of the source speech segments in this study are presented in Chapter 5, Section 5.2.

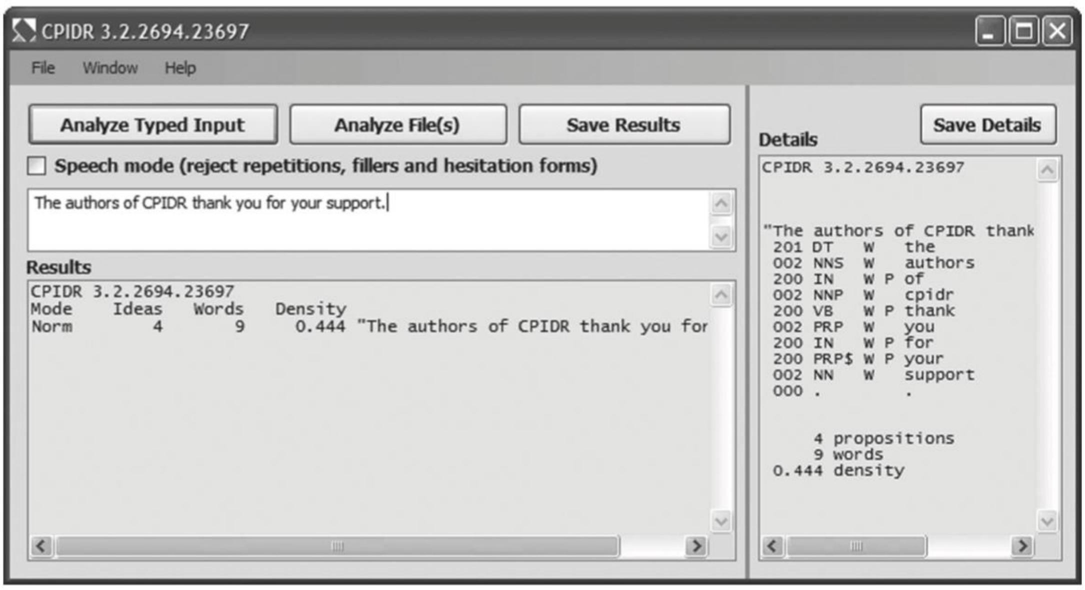


Figure 4-2. A sample of the result of propositional analysis in CPIDR

The concept of idea density can be confused with terms such as lexical density and informational load in Plevoets and Defrancq (2016), as well as information density and new concept density in Liu and Chiu (2009). Specifically, Plevoets and Defrancq (2016) operationalize informational load as a general term to cover four features of source speech in interpreting: delivery rate, lexical density, percentage of numerals, and average sentence length. All of these factors are found to be influential in deciding the amount of information that interpreters have to deal with during the process of interpreting. This information-centered idea of Plevoets and Defrancq (2016) is different from the proposition-based measurement in the present study. Although all of the four aspects of speech features are controlled in the present study, they are not categorized or referred to as informational load. In this study, the delivery rate of the source speech is controlled and balanced during the selection and edition of the speech videos. The lexical density can be reflected with the measured idea density, as both measures are developed from the identification of part-of-speech tags that can consist of propositions in a text, such as "verbs, adjectives, adverbs, prepositions, and conjunctions" (Brown et al., 2008). As for the percentage of numerals, the source speech in the present study does not contain many numbers and this is not a critical concern in this investigation. Finally, since the average sentence length is an essential component of the readability tests (Flesch, 1948), this is controlled by balancing the readability levels of the speech segments.

By comparison, the concept of idea density that is adopted in the present study is more similar to that of information density and new concept density in Liu and Chiu (2009). All of these measures are based on propositional analysis which aims to estimate the degree to which information is densely distributed in the source speech. Propositions in Liu and Chiu (2009) are recognized according to the manual created by Kieras and Bovair (1986), whereas propositions in this study are identified through CPIDR5 which is based on Kintsch's (1974) guidelines on proposition identification. These two approaches to propositions are described as "close relatives" (Weaver & Kintsch, 1990, p. 235) to each other, because both follow the same principle that "(e)ach proposition consists of a set of concepts" (Turner, 1987, p. 3). "The first of these (concepts) is the predicate of the proposition. The others are the proposition's arguments" (p. 3). From this perspective, the core of information density in Liu and Chiu (2009) is in line with that of idea density in the present study. As for the new concept density, which is defined as "the proportion of the number of new arguments to the total number of propositions in each source material" (p. 249), this has been found to be not effective in predicting interpreters' CI performance. Neither are the results concerning the new concept density in accordance with the results of other difficulty measures in Liu and Chiu (2009). Therefore, the new concept density is not included in the present study to assess the difficulty level of the source speech in interpreting.

Despite their advantages outlined above, readability and idea density tests also have their drawbacks. For instance, the readability indices "do not distinguish between sense and nonsense, i.e., the semantic acceptability of a text is ignored" (Jensen, 2009, p. 68). In addition, they are generally adopted to assess the difficulty of monolingual reading rather than translation and interpreting (Vanroy et al., 2019). Furthermore, these measures point to the objective and comparative level of "text complexity" rather than the perceived level of "text difficulty" (Jensen, 2009, p. 62-63). Therefore, in addition to adopting these two objective measures, it is necessary to include subjective assessment to further examine the difficulty level of the source speech.

4.1.2 Subjective measurements

Expert judgment is a widely-adopted method of difficulty assessment in language testing, translator training and interpreter training (Liu & Chiu, 2009). Liu and Chiu (2009) propose eight aspects of difficulty assessment in interpreting: words, syntactic structure, information density, coherence, logic, clarity, abstractness, and required background knowledge. In addition, they show that the experts' scores on three source speeches in these eight aspects were consistent with the student interpreters' feedback after interpreting. Therefore, in the present study, ten interpreting trainers and freelance interpreters were invited to assess the difficulty level of the source speech segments from these eight aspects, for the purpose of ensuring that the designed difficulty level of the source speech is in accordance with the experts' judgments (results presented in Chapter 5, Section 5.2). An introduction to the eight aspects of source speech difficulty assessment is presented below.

Words and syntactic structures are important factors of readability assessment. At word level, the effects of word length (Baddeley & Andrade, 1994) and word frequency (Roodenrys et al., 2002) have been repeatedly demonstrated in studies on language processing efficiency and memory recall performance. On a syntactical level, sentence length, sentence duration and syntactical complexity are frequently found to be influential in deciding interpreters' performance in CI (Liu & Chiu, 2009), SI (T. Lee, 2002; Timarová et al., 2014) and sight translation (Chmiel &Lijewska, 2019; Ma et al., 2022). From this perspective, word- and sentence-related properties of a source speech are essential factors for experts to consider when evaluating the difficulty level of an interpreting task.

Information density, as introduced in the previous section, represents the degree to which information is densely distributed in a text. In interpreting studies, information density is found to be the most important factor in interpreting performance in CI (Liu & Chiu, 2009). In early research on lecture note-taking, it was also found that note quality decreases considerably when the class is highly informative (Russell et al., 1984). Taken together, these findings imply that the information density of a source speech in a CI task might also exert great impacts on the process and product of note-taking, making this a necessary consideration while assessing the difficulty level of the source speech in interpreting.

Coherence refers to “the ways in which the components of the textual world, i.e., the configuration of concepts and relations which underlie the surface text, are mutually accessible and relevant” (De Beaugrande & Dressler, 1981, p. 4). It is closely related to the *Logic* of the speech, which centers on “express(ing) the logical consistency of utterances” (Kostopoulou, 2007, p. 146). Since interpreters are usually expected to deliver a coherent and logical target speech, the coherence and logic of the source speech are thus essential in organizing the contents of the target speech.

Clarity and *Abstractness* consist of another group of confusing concepts which can be differentiated by considering their antonyms. The opposite of clarity is ambiguity (e.g., Van der Schoot et al., 2009), whereas the antonym of abstractness is concreteness (e.g., Sherman & Kulhavy, 1978). *Clarity* focuses on whether the “underlying logic” (Gile, 2008, p. 3) is clear, whereas *Abstractness* refers to items or even thoughts that cannot be specified easily in terms of concrete entities. Sometimes even when the words and sentences are not complex in the source speech, if the contents are too ambiguous and abstract, this can cause interpreters great difficulties in source speech comprehension and target speech production. Therefore, *Clarity* and *Abstractness* of the source speech are important considerations in the assessment of source speech difficulty.

Required background knowledge in interpreting includes but is not limited to syntactic knowledge, semantic knowledge, associated knowledge and background experience, cultural awareness and contextual knowledge (Russel, 2005, p. 145). For instance, advanced preparation for technical speeches has been found to be effective in decreasing the ear-voice span and improving the interpreting performance in SI (Díaz-Galaz et al., 2015). After intensive training (48 hours in two months) on a text type, students' cognitive effort of note-taking could reduce significantly when interpreting the same text type, because of their increased familiarity with the typical linguistic expressions (Hu, 2008). These findings suggest that background knowledge can contribute substantially to the cognitive demand of an interpreting task and the cognitive load of an interpreter, and thus should not be excluded from the assessment of a source speech in interpreting tasks.

In addition to expert judgment, the participants' feedback in the pilot study was also collected through an adapted NASA-TLX questionnaire to examine whether there were discrepancies between the designed and the perceived difficulty levels of the source speech segments. NASA-TLX is a multidimensional scale for task performers to report their subjective workload. It has been proven reliable in assessing translation difficulty (Sun & Shreve, 2014) and suggested for measuring the interpreters' perceived difficulty during interpreting (for details, see S. Chen, 2017c, p. 650). Following Sun and Shreve (2014), the present study collects the participants' ratings in the four aspects of NASA-TLX after completing the interpreting task (Table 4-2).

Table 4-2. NASA-TLX items and explanations

Assessed item	Description
Mental demand	How much mental and perceptual activity was required (e.g., thinking, deciding, remembering, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration level	How insecure, discouraged, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?
Performance	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Mental workload refers to the input load (Johannsen, 1979) of a task, pointing to the objective difficulty level of interpreting. Basically, a task with a higher mental workload requires a higher proportion of controlled processing (Meijman & O'Hanlon, 1984). By comparison, *Effort* focuses on the actual allocation of time and resources for task completion and is decided by the interpreter subjectively. This reflects how, after perceiving such a degree of *Mental workload* in the task, interpreters resolve the difficulties by increasing or decreasing the devotion of their limited *Effort*. In brief, *Mental workload* points to the perceived task difficulty, and *Effort* indicates the extent to which interpreters attempt to complete interpreting.

Frustration level, which emphasizes interpreters' feelings during the process of interpreting, is found to be negatively related to interpreting performance (e.g., Chang & Schallert, 2007). This is because emotions can be associated with the cognitive load of task completion in many ways (Plass & Kalyuga, 2019). For instance, emotional control requires extra cognitive effort, thus decreasing the availability of cognitive resources for other subtasks during interpreting. Emotions can also influence interpreters' motivation level during interpreting, which further affects the amount of effort that is devoted to the interpreting task and the quality of the interpreting product. More than often than not, interpreters express that "I was frustrated but I still wanted to interpret it. I didn't really hear clearly what the [following segments] was about. I was still frustrated over my inability to interpret this sentence" (Chang & Schallert, 2007, p. 159). Therefore, examining interpreters' frustration level during interpreting is important in helping researchers to understand the observed changes in interpreters' cognitive load of interpreting.

Performance is a frequently adopted measure of cognitive load in interpreting (e.g., Mead, 2000 on CI; Bendazzoli et al., 2011 on SI), with the notion that higher cognitive load would lead to poorer interpretation quality. However, self-reported *Performance* can be very different from objective performance ratings. For instance, Roziner and Shlesinger (2010) found that professional interpreters rated their performance significantly poorer in remote SI than in on-site SI, which were actually scored similarly by raters. This inconsistency in performance evaluation could be caused by the fact that the interpreters' judgments were distorted by the greater cognitive load they experienced while interpreting remotely. Therefore, self-reported *Performance* can provide important implications about the interpreters' cognitive load during interpreting and is thus worth investigating while assessing the cognitive demand of an interpreting task.

After collecting feedback from the participants in the pilot study, this study finally followed Sun and Shreve (2014) to design the rating scale (Figure 4-3). "Each subscale is presented as a line divided into 20 equal intervals anchored by bipolar descriptors" (p. 102). Specific results concerning the participants' NASA-TLX scores in the pilot and formal studies can be found in Chapter 5, Section 5.2 and Chapter 6, Section 6.1.

Figure 4-3. An example of the NASA-TLX rating scale

4.2 Measurement of note-taking and note-reading effort

The diverse research methods discussed in Chapter 2, Section 2.2 provide a solid foundation for the present study's methodology. In this project, eye-tracking and pen-recording methods are adopted in combination to explore interpreters' note-taking and note-reading effort from visual, cognitive, physical and temporal aspects. In addition, a descriptive analysis of the product of note-taking and a summative assessment of the product of interpreting are conducted to explore the associations between interpreters' note-taking behaviour and interpretation quality. This section provides an overview of the adopted process- and product-oriented methods and introduces how these methods are applied to answer the proposed research questions.

4.2.1 Eye-tracking method

4.2.1.1 Selection of eye-trackers

To find out the appropriate eye-tracker for the present study, 14[8] interpreting studies are identified from three major databases (CIRIN Bulletin, TSB and CSSCI) with "eye-tracking", "eye-tracker", "fixation", "saccade" and "pupil" as the keywords for searching. As Table 4-3 presents, there are three major types of eye-trackers (Figure 4-4) that have been applied in interpreting studies: remote eye-trackers (e.g., Tobii TX300 and Eyelink 1000 plus); head-mounted eye-trackers (e.g., SMI ETG glasses and Tobii glasses eye-tracker); and eye-trackers that require participants to keep their heads stable on a chin-rest (e.g., SR Research's EyeLink 1000 Head Supported). In previous literature, Tobii TX300 and EyeLink 1000+, both of which are remote eye-trackers, are most frequently adopted. These eye-trackers feature comparatively low intrusiveness, as the participants do not have direct contact with the eye-trackers; and they allow researchers to collect the participants' eye-tracking data without intervening in the process of translation or interpreting. Therefore, the validity of the eye-tracking data in these studies are maintained at a high level, ranging from 72.00% (Zheng & Zhou, 2018) to 100% (Su & Li, 2020). However, one issue with this type of remote eye-trackers is

[8] Papers based on the same research project are counted only once.

that they can only record the eye movements on the screen but not off the screen. This is also the reason why these eye-trackers are mostly adopted in sight-translation experiments where the stimuli can be simply presented on a screen (see Table 4-3). By comparison, head-mounted eye-trackers can record all kinds of eye movements without restrictions on body movements. Despite this, this type of eye-trackers can be quite intrusive to participants as they are fixed on the participants' heads. In the three studies that adopted eye-tracking glasses, S. Chen et al. (2021) report that the validity rate of the data was only 69.23%, while the other two did not report data quality assessments (Dragsted & Hansen, 2009; Tiselius & Sneed, 2020). For the same reason, eye-trackers that require fixed head positions are least adopted in interpreting studies (e.g., Wan & Qian, 2020).

Table 4-3. The adoption of eye-trackers in interpreting studies (in chronological order)

Literature	**Interpreting mode**	**Eye-tracker description**
Dragsted & Hansen (2009)	Sight translation	"A Tobii eye-tracker" (p. 591)
Seeber & Kerzel (2012)	Sight translation	"An EyeLink II head-mounted binocular eye tracker at 250 Hz" (p. 234)
Zhao & Xu (2018)	Sight translation	EyeLink 1000+
Zheng & Zhou (2018)	Sight translation	Tobii TX300
Korpal & Stachowiak-Szymczak (2018)	SI	"EyeLink 1000+ remote eye-tracker…accuracy was up to 0.50º" (p. 342)
Wang et al. (2018)	Reading for sight translation	TobiiTX 300
Chmiel & Lijewska (2019)	Sight translation	EyeLink 1000+
Su & Li (2020)	Sight translation	Tobii TX300
Korpal & Stachowiak-Szymczak (2020)	SI	EyeLink 1000+
Wan & Qian (2020)	Sight translation	No specifics about the eye-tracker, but the participants kept their heads on a holder
Tiselius & Sneed (2020)	Dialogue interpreting	SMI glasses
S. Chen (2020a) and S. Chen et al. (2021)	CI	A pair of SMI ETG glasses that is "lightweight (47g)" and featured with "tracking accuracy of 0.5°", "sampling rate of 60 Hz" and "a built-in high-definition camera" (p. 8)
Ma et al. (2021)	Sight translation	EyeLink 1000+
Lu & Zheng (2021)	Sight translation	Tobii TX300

Note. The validity rates of Seeber and Kerzel (2012) and Chmiel and Lijewska (2019) are calculated based on the number of sentences rather than the number of participants.

Figure 4-4. Examples of the three types of eye-trackers (a remote eye-tracker on the left, a head-mounted eye-tracker in the middle, and an eye-tracker with a chin-rest on the right)

The accuracy of an eye-tracker plays an important role in deciding the eye-tracking data quality, and the intrusiveness of the device is essential for the experiment’s ecological validity. These two requirements of data quality and ecological validity are difficult to be met simultaneously with one eye-tracker. For instance, in Zheng and Zhou (2018), a remote eye-tracker of Tobii TX300 was adopted to record students’ eye movements during sight translation, which ensured a high level of ecological validity during the experiment. However, it was also a matter of fact that the eye-tracker did not have exceptionally high sample rates. In the study, the researchers filtered the collected data with a comprehensive set of criteria proposed by Hvelplund (2014), which includes gaze time on the screen (GTS), gaze sample to fixation percentage (GFP), and mean fixation duration (MFD) (see Table 4-4). These three measures, respectively, focus on the proportion of gaze on screen, the proportion of fixation in all eye movements, and the average length of fixation. Data points that were one standard deviation (SD) below the mean of each criterion were excluded; and only those participants that had met at least two of the three criteria were included in the analysis. In this way, the researchers ensured that the remaining data in the analysis were of high quality. By contrast, there are also studies that did not report procedures of data quality assessment (e.g., Wang et al., 2018) or simply relied on one criterion (e.g., Su & Li, 2020) in the assessment. Specific details are summarized in Table 4-4.

In terms of studies on interpreters’ note-taking behaviour in CI, S. Chen et al. (2021) recorded interpreters’ eye movements on the notepad during note-reading with a pair of SMI glasses. Since the eye-tracking glasses can be intrusive, the researchers designed a practice session, allowing the interpreters to familiarise themselves with the experiment setting. Only those who rated a comfort level in the experimental setting higher than 50% (from 0 to 100%) would proceed to the formal experiment. Adding to those data losses after the assessment of eye-tracking data quality, roughly 31% of the participants were excluded from data analysis. Issues such as "the eye tracker did not work well with participants who wore bi-focal glasses" (p. 8) were raised as causes that led to the comparatively high rate of data loss. These results indicate that adoption of an eye-tracker that requires direct contact with participants needs exceptionally cautious operation and well-designed pre-test training.

Table 4-4. Assessment of eye-tracking data quality

Literature	Assessment methods	Valid rate
Dragsted & Hansen (2009)	None	NE
Seeber & Kerzel (2012)	"When blinks masked more than 50 per cent of the pupil size measurements… the item was not included in the analysis" (p. 236).	80.15%
Zhao & Xu (2018)	None	NE
Zheng & Zhou (2018)	Excluded data below one SD of the mean of the participants' GTS, GFP and MFD based on Hvelplund (2014)	79.17%
Korpal & Stachowiak-Szymczak (2018)	"…. needed to be excluded from the analyses due to low eye-tracking accuracy, i.e. an observed difference between the position of a test stimulus on the screen and the measured gaze position" (p. 341)	72.00%
Wang et al. (2018)	None	NE
Chmiel & Lijewska (2019)	Viewing measures: "removed all observations longer than 4000 ms (4.5% of data)" (p. 387)	95.3%
	Sentence viewing times: "all observations below 250 ms and above 20,000 ms were excluded from the analysis (1% of the data)" (p. 388)	99.0%.
Su & Li (2020)	"…recordings with at least 80% of valid gaze samples were used for further analysis" (p. 1007) based on Hvelplund (2014)	100%
Korpal & Stachowiak-Szymczak (2020)	"low eye-tracking accuracy" (p.132) (without further explanations)	79.63%
Wan & Qian (2020)	None	NE
Tiselius & Sneed (2020)	Not specified	NR
S. Chen et al. (2021)	Not specified	69.23%
Ma et al. (2021)	"excluded participants with half of their fixations bring shorter than 200 ms" (p. 11) based on Hvelplund (2011) and Pavlović & Jensen (2009)	77.27%
	exclude potential measurement errors: "fixations shorter than 80 ms and longer than 1200 ms" (p. 11) based on Drieghe et al. (2008) and White (2008)	
	"examined both fixation duration and the degree of fixation drift" (p. 11) to exclude serious fixation drifts	
Lu & Zheng (2021)	excluded the fixations that were shorter than 180 ms based on Sjørup (2013) and Da Silva et al. (2015)	NR
	excluded outliners with box plots based on Feng (2018)	

Note. NE=not evaluated; NR=not reported.

Concerning that the present study focuses on VRI where interpreters deliver interpretations via a video link, a remote eye-tracker, Tobii Pro Fusion 250, that is attached to the computer screen was adopted in the formal experiment. This eye-tracker records eye movements on the screen remotely without physical contact with the participants. This allows the present study to present the source speech on the screen and collect the participants' eye movements during interpreting at the same time. In addition, this eye-tracker features light weight (168g) and high portability. This is especially important for this investigation as it was conducted in different cities during the difficult times of COVID-19. However, this eye-tracker cannot record off-screen eye movements, which means that interpreters had to keep their eyes on the screen to conduct note-taking and note-reading. Therefore, the participants were asked to write with a smartpen on a tablet, where all handwriting on the notepad was synchronized on the screen in real-time. Details about the smart pen and experimental setting can be found in Section 4.2.2 and Chapter 5, Section 5.3.

In addition, Hvelplund's (2014) assessment method of eye-tracking data quality was adopted in the present study, for two reasons. Firstly, Hvelplund's criteria include the gaze time on the screen, which allows the researcher to check whether the participants looked down on the tablet too much in order to position their pen. Secondly, since the total time that the participants looked at the screen differed to some degree in the experiment, it was necessary to include a measure that did not presuppose the same amount of screen-gazing time to judge eye-tracking data quality. GFP can well serve this purpose and is thus adopted in the present study. Following Zheng and Zhou (2018), data that were one SD below the mean of each criterion were excluded from the data pool; and only those participants with data points meeting at least two of the three criteria were included in the data analysis. Data screening results are presented in Chapter 5, Section 5.4.

4.2.1.2 Adoption of eye-tracking indicators

Among the 15 identified interpreting studies in the three databases, eye-tracking methods are frequently adopted to investigate two aspects of the interpreting process:

processing patterns and cognitive load[9] (Table 4-5).

Table 4-5. Eye-tracking measures adopted in previous interpreting studies

Literature	Interpreting mode	Research focus	Eye-tracking measures
Dragsted & Hansen (2009)	Sight translation	Differences between written translation and sight translation in terms of the processing patterns	• Heat map • Fixation count • Mean fixation duration
Seeber (2011, 2013), Seeber & Kerzel (2012)	SI	Cognitive load of interpreting sentences that require syntactic restructuring	• Pupil diameter
Zhao & Xu (2018)	Sight translation	Cognitive load of interpreting logic conjunctions	• Mean fixation duration • Fixation count • Revisit count
Zheng & Zhou (2018)	Sight translation	Processing patterns and cognitive load of interpreting metaphorical expressions in sight translation	• Eye-voice span • Total processing time • Total fixation duration
Korpal & Stachowiak-Szymczak (2018)	SI	Cognitive load of interpreting numbers	• Mean fixation duration
Wang et al. (2018)	Reading for sight translation	Processing depth and cognitive load of reading	• Total reading time • Fixation count • Mean fixation duration
Chmiel & Lijewska (2019)	Sight translation	Processing patterns and cognitive load of interpreting sentences with (a)symmetrical structures	• Sentence viewing time • Gaze duration • Regression path duration • Percentage of dwell time on each sentence
Su & Li (2020)	Sight translation	Processing patterns in different translation directions	• Preparatory reading time • Total translation time • Mean fixation duration • Eye-voice span
Korpal & Stachowiak-Szymczak (2020)	SI	Cognitive load of interpreting numbers at different delivery rates of the source speech	• Fixation count per minute on numbers • The percentage of gaze time devoted to numbers
Wan & Qian (2020)	Sight translation	Effects of interpreting experience and task difficulty on the cognitive load of sight translation	• Mean fixation count • Mean fixation duration
Tiselius & Sneed (2020)	Dialogue interpreting	Cognitive load of dialogue interpreting	• Gaze patterns while listening to different interlocutors

[9] "Cognitive load" and "cognitive effort" are often used interchangeably in these studies to suggest the amount of cognitive processing that is involved in interpreting. The difference between the two terms lies in the expressions: "the associated cognitive load imposed by each mode of interpreting" (Doherty et al., 2022, p. 10) and "the high cognitive effort experienced (by the interpreters)" (Korpal & Stachowiak-Szymczak, 2018, p. 340).

S. Chen (2020a) and S. Chen et al. (2021)	CI	Processing patterns and cognitive load of note-reading during the output phase of CI	• Total fixation duration, mean fixation duration and duration count • First dwell duration and second dwell duration • Revisit count • Regression rate • Skip rate
Ma et al. (2021)	Sight translation	Effects of word order asymmetry between the SL and the TL on the cognitive load of sight translation	• Dwell time • Fixation count • First fixation duration • Regression path duration
Lu & Zheng (2021)	Sight translation	Cognitive resource allocation and cognitive load during metaphor translation	• Heat map • Total fixation duration • Regression duration • Fixation count • First fixation duration

Processing patterns, which mainly refer to the allocation of time and effort to different subtasks during the process of interpreting, can be investigated with visualization tools and temporal measures. The heat map is an important visualization method that allows researchers to quickly understand where the participants have been looking during the experiment. As presented in Figure 4-5, a heat map consists of "hot" and "cold" areas. In general, the "hotter" the spot is, the longer the eyes stay in the area. Based on the eye-mind hypothesis (Just & Carpenter, 1980), there are close connections between what people are looking at and what they are processing in the mind. Therefore, heat maps can provide researchers with a basic idea about the interpreters' allocation of processing capacity to different parts of the stimulus.

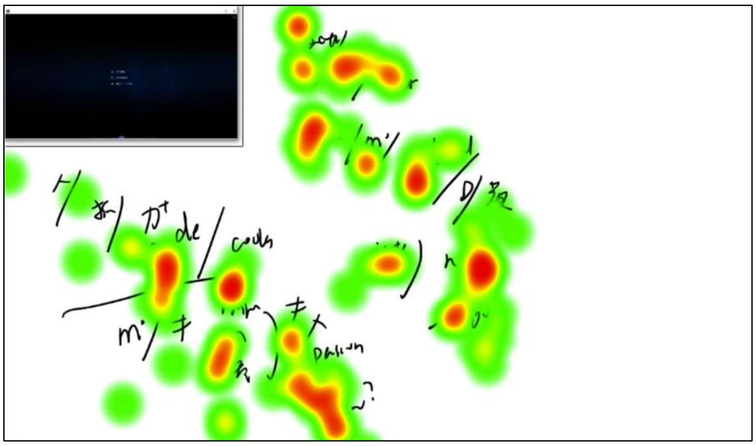


Figure 4-5. An example of heat maps based on the note-reading process of P06 in the present study

However, since heat maps are static pictures, they cannot illustrate what happens during the process of interpreting in a dynamic way. Researchers sometimes also turn to scan paths, regression-related measures and gaze patterns to see how interpreters shift their attention to different parts of the stimulus while interpreting. For instance, revisit count (S. Chen et al., 2021; Zhao & Xu, 2018) and regression path duration (Chmiel & Lijewska, 2019; Ma et al., 2021) have been adopted to reveal the cognitive effort of incomplete lexical processes during note-reading and the sequence of word processing in sight translation. Gaze patterns are collected to examine how interpreters shift their attention among different interlocutors in dialogue interpreting (Tiselius & Sneed, 2020) and switch among the prepared materials, the speakers' slides, the partner's actions and other visual information in the meeting room in SI (Lei & Li, 2019). In addition, Su and Li (2020) divide the time of reading before interpreting (the preparatory reading time) by the total time that the participants devoted to a sight translation task (the total translation time), to explore the extent to which participants prioritize source text comprehension over target speech production. Zheng and Zhou (2018) used the eye-voice span during sight translation to examine the demand for coordinating source-text processing and target-text delivery. Overall, interpreters' processing patterns can be visualized through comparatively static heat maps which picture the interpreting process at a global level, and quantified with relatively dynamic viewing measures that reveal the interpreting process at a local level.

On the other hand, the cognitive load of interpreting emphasizes the depth and amount of cognitive processing that are involved in an interpreting task. There are generally four types of indicators for the cognitive load of interpreting. The first one, pupil diameter, does not require interpreters to gaze at a certain area as in normal eye-tracking studies. Therefore, it is mostly adopted in SI studies, where there is no visual input for interpreters to process (e.g., Seeber, 2011, 2013). However, since pupil diameter can be highly sensitive to many factors such as age (Van Gerven et al., 2004), intelligence (Ahern & Beatty,1979), fatigue (Lowenstein & Loewenfeld, 1962) and luminosity (Clarke et al., 2003), changes in the pupil diameter do not necessarily point to the change in interpreters' cognitive load. For the same reason, pupil dilation is not widely adopted in interpreting studies. The second type of indicator is the sum of fixation durations or fixation counts during the process of interpreting, such as total fixation duration (S. Chen et al., 2021; Lu & Zheng 2021; Zheng & Zhou, 2018), total fixation count (S. Chen et al., 2021; Dragsted & Hansen, 2009; Ma et al., 2021) and total gaze duration (e.g., Chmiel & Lijewska, 2019; Tiselius & Sneed, 2020). These measures can be greatly affected by the size of the AOI. For instance, full words in notes entail significantly longer fixation durations than abbreviations in note-reading, but such significant differences disappear when the length of the two types of notes is controlled (e.g., S. Chen et al., 2021). Hence, when explaining the results of these measures, it is important to recognize that they actually point to the total amount of cognitive processing that has been involved in the task rather than the extent to which the task is cognitively demanding (e.g., Cui & Zheng, 2020; Hvelplund, 2019). By comparison, the third type of indicator, which includes mean fixation duration (Dragsted & Hansen, 2009; Korpal & Stachowiak-Szymczak, 2018, Wang et al., 2018; Zhao & Xu, 2018), mean fixation count (Wan & Qian, 2020) and fixation count per minute (Korpal & Stachowiak-Szymczak, 2020), is measured in units. It is frequently adopted in interpreting studies to indicate the degree of cognitive demand of an interpreting task. Sometimes researchers even explore further to measure the local cognitive load of different types of cognitive processing involved in interpreting, which consists of the fourth type of indicators. For instance, in S. Chen et al. (2021)[10], first fixation duration (FFD: the duration of the first fixation on a note) is used to indicate the cognitive effort of recognizing a note. First pass duration (FPD: the total fixation duration within the first visit to a note) is adopted to reflect the cognitive effort of integrating the meaning of notes at a lexical level. In addition, total fixation duration (TFD) and second pass duration (SPD: TFD except for the FPD) are used to measure the cognitive effort of late-stage note processing, such as integrating the meaning of notes at a sentence and textual level. In this way, researchers are able to investigate how interpreters distribute their limited cognitive resources to different parts of cognitive processing that are involved in an interpreting task.

Based on the application of eye-tracking indicators in previous interpreting studies, a range of eye-tracking tools and measures is adopted in the present study to investigate the processing patterns and cognitive load of note-taking and note-reading in VRI (Table 4-6).

[10] In Chen et al. (2021), the concept of “dwell” is equal to that of “pass”. Therefore, total dwell duration is the same as total fixation duration.

Table 4-6. Eye-tracking measures adopted in the present study

Tool/measure	Operational definition	Corresponding activities	Previous applications
Heat map	A visual representation of attention distribution on a given area	The distribution of overt visual attention on the screen	Processing patterns in sight translation (Dragsted & Hansen, 2009)
Gaze plot	A visual representation of the sequence of fixations	The sequence of information processing on the screen	Similar to the scan path for investigating word processing sequence in sight translation (Su & Li, 2020)
Event log	A chronological and detailed record of each fixation and saccade in AOIs		
Total fixation duration (TFD)	The total time of fixations on an AOI	Overt visual attention	Overt visual attention on consultation in written translation (Cui & Zheng, 2020; Hvelplund, 2019)
Total fixation count (TFC)	The total count of fixations in an AOI	Overt visual attention	Cognitive effort of note-reading (S. Chen et al., 2021)
Mean fixation duration (MFD)	The average time of fixations on an AOI	Overall cognitive effort	Cognitive effort of reading for comprehension (Jakobsen & Jensen, 2008) and note-reading (S. Chen et al., 2021)
Visit count (VC)	The number of times that eyes enter and leave an AOI	The shift of the interpreter's overt visual attention to different areas on the screen	Similar uses as the "shifts of overt visual attention" in Doherty et al. (2022)
Revisit count (RVC)	The number of times that eyes re-enter and leave an AOI	Cognitive effort of incomplete note processing	Cognitive effort of incomplete processing in note-reading (S. Chen et al., 2021)
First fixation duration (FFD)	The duration of the first fixation in an AOI (note)	Cognitive effort of note recognition in early processing	Cognitive effort of note recognition (S. Chen et al., 2021) and word recognition (Ma et al., 2021)
First pass duration (FPD)	The total fixation duration between eyes entering and leaving an AOI (note) for first time	Cognitive effort of integrating the meaning of notes at a lexical level in early processing	Cognitive effort of integrating the meaning of words in reading (Rayner, 1998)
Second pass duration (SPD)	The total fixation duration that eyes stay in an AOI except for the FPD	Cognitive effort of integrating the meaning of notes at a sentence and textual level in late processing	Cognitive effort of sentence processing (Murray, 2000) and global text processing (Hyönä et al., 2003)
Skip rate	The percentage of AOIs that do not receive any fixation	Predictability of notes during note-reading	Predictability of notes during note-reading (S. Chen et al., 2021)

Firstly, visualization tools including heat maps, gaze plots and event logs are adopted to indicate how interpreters allocate their overt visual attention on the screen and process the notes in sequence. In brief, the heat map provides an overall picture of attention distribution on the screen. The gaze plot and event log show the detailed information of each fixation on the designed AOIs (as presented in Figure 4-6 and Figure 4-7).

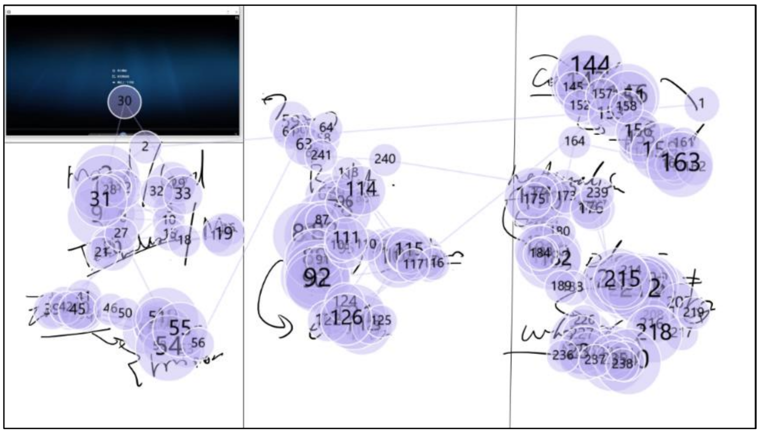


Figure 4-6. An example of the gaze plot exported from Tobii Pro Lab

Participant	TOI	Interval	Event_type	Validity	EventIndex	Start	Stop	Duration	AOI
P01	P01_P1_NR1	1	Fixation	Partial	1	0	88	88	RS%mix∗pen+〇^2$1illustrator
P01	P01_P1_NR1	1	Fixation	Whole	2	104	521	417	E%fw_en∗ted^1$1ted
P01	P01_P1_NR1	1	Fixation	Whole	3	562	754	192	Screen
P01	P01_P1_NR1	1	Fixation	Whole	4	787	946	158	Screen
P01	P01_P1_NR1	1	Fixation	Whole	5	962	2021	1058	E%fw_en∗ted^1$1ted
P01	P01_P1_NR1	1	Fixation	Whole	6	2037	2254	217	Screen
P01	P01_P1_NR1	1	Fixation	Whole	7	2279	2496	217	Screen
P01	P01_P1_NR1	1	Fixation	Whole	8	2521	3112	592	E%fw_en∗ted^1$1ted
P01	P01_P1_NR1	1	Fixation	Whole	9	3154	3312	158	Screen
P01	P01_P1_NR1	1	Fixation	Whole	10	3346	3612	267	Screen

Figure 4-7. An example of the eye-tracking event log exported from Tobii Pro Lab

Secondly, measures that are related to the size of an AOI (such as TFD and TFC) are used to indicate the overt visual attention that the participants have paid to the notes during interpreting, whereas MFD, which features a mean attribute, is adopted to indicate the average cognitive effort of taking or reading one note. Thirdly, measures concerning some specific aspects of fixations such as FFD, FPD and SPD are selected to investigate the cognitive effort of early- and late-stage note processing during note-reading. Finally, the skip rate of notes during note-reading, which represents the percentage of notes that have not been fixated at all during the output stage of CI, is used to indicate the predictability of notes.

4.2.2 Pen-recording method

4.2.2.1 Selection of digital pens

Insofar as note-taking in interpreting is concerned, two researchers have adopted digital pens to investigate interpreters' note-taking behaviour. Marc Orlando applies a Livescribe smartpen which features a built-in speaker and an infrared camera (Figure 4-8) (see Orlando, 2010, 2014, 2015) in the interpreting class to record and observe the process of note-taking. With the so-called dot-paper technology, students could replay the corresponding source speech audios simply by tapping on the written notes. In this way, they could better identify which part of the speech hindered their listening comprehension and note production. Sijia Chen, on the other hand, focuses on measuring the physical, temporal and potential cognitive demands of note-taking (S. Chen, 2017c, 2020a, 2020b, 2022). She adopts a Wacom digital pen (Figure 4-9) which is able to record the distance, speed and duration of handwriting to probe into the process of note-taking. In both studies, the participants show a high level of acceptance of the digital pen, with only 4 participants' data (15.38%) being excluded in S. Chen (2020b) because of poor writing experience[11]. Overall, digital pens are not widely adopted in interpreting studies, but they have shown great potential in revealing the note-taking process in CI.

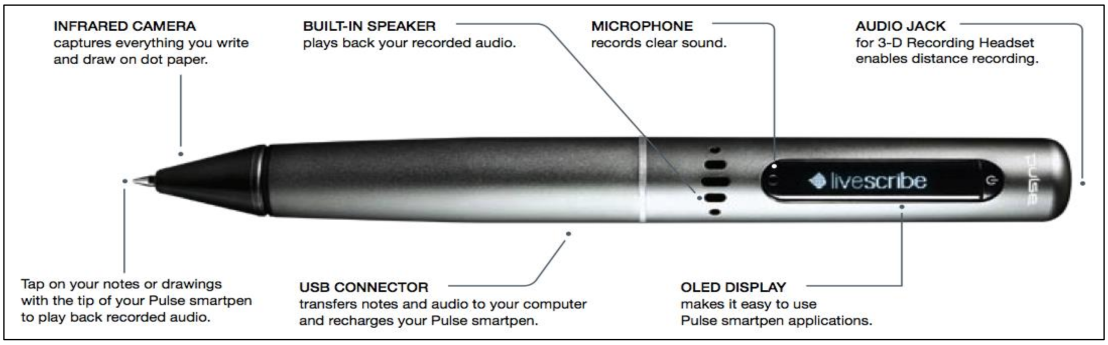


Figure 4-8. The Livescribe smartpen adopted by Orlando (2010)

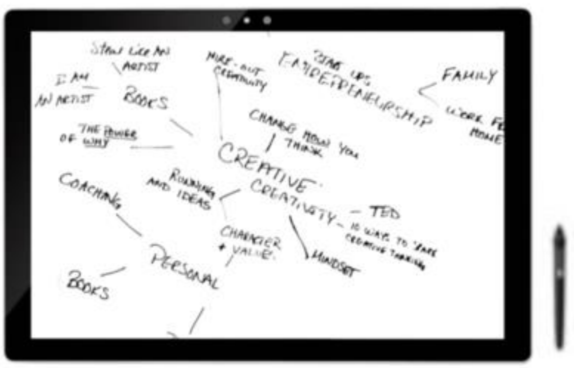


Figure 4-9. A digital pen with a Tablet PC

[11] The comfort level of the smartpen was rated by a 7-point scale, and ratings lower than 4 were considered as a poor experience.

One important research aim of the present study is to quantify interpreters' note-taking effort in VRI from visual, cognitive, physical and temporal aspects. This requires that the digital pen can record the writing scenes as in S. Chen's study for measuring the physical and temporal aspects of the note-taking effort. At the same time, the pen should be able to synchronize what is written on the tablet and what is presented on the computer screen for the sake of collecting eye-tracking data remotely. The Wacom CTL 672 digital pen and its equipped graphics tablet can well fit these research needs. Firstly, the digital pen can only be activated by connecting to a computer screen. All the writings on the tablet are automatically presented on the screen (Figure 4-10), and they can be screen-recorded for later processing. Secondly, the digital pen is pressure-sensitive (2048 LPI[12]), cordless and battery-free. Together with the graphics tablet, this set of devices only weighs 441g (9g for the smartpen). This not only secures high accuracy in data collection but also secures high portability for a multicentre study. Thirdly, handwriting on the tablet with this digital pen can be recognized as mouse clicks in the eye-tracking software of Tobii Pro Lab, which contains a series of click-based measures that allows the present study to measure the participants' physical effort in note-writing. Based on these reasons, the Wacom CTL 672 digital pen and its equipped tablet were selected and adopted in this investigation.

Figure 4-10. The experimental setting of the present study

[12] LPI: lines per inch. The higher the LPI, the better the resolution.

However, writing on a tablet with eyes fixating on the screen can cause hand-eye coordination problems. Loch (2005), who explored the usability of this pen-and-tablet setting in university lecturing, described her experience with the digital pen in detail:

> Although writing on a tablet requires a certain level of hand-eye coordination, as the user writes blindly on the tablet while looking up at the screen, I found it did not take long to get used to writing with the pen, both standing in front of the class and sitting at my desk. I was in fact so impressed by the ease of use of both the pen and the software (p. 234).

To ensure the applicability of this pen-and-eye setting in the present project, two pilot studies were conducted to check potential hand-eye coordination problems, and a practice session was designed in the formal experiment to ensure the participants' familiarity with this experimental design. In the end, 2 (out of 61) participants' data points were excluded from the analysis because of the digital pen. Details can be found in Chapter 5, Sections 5.3 and 5.4.

4.2.2.2 Adoption of pen-recording indicators

In the series of S. Chen's work, she explores a range of pen-recording indicators that can be adopted to visualize and quantify the process of note-taking in CI (Table 4-7). These measures can be divided into three kinds according to their indications: cognitive processing involved in note-planning (EPS); physical and temporal demand of note-writing (distance and duration of pen movements); and the interpreters' psychological tension during note-taking (pen-tip pressure).

Table 4-7. Pen-recording measures adopted in S. Chen (2020b)

Measures	Operational definition	Corresponding activities
EPS	"The time span between the moment a speech unit is heard and the moment it is written down as notes" (p. 125-126)	"The combined cognitive demand to analyse the source speech unit, to decide whether or not to write a note, and if so, which form and language to use" (p. 126)
Distance, speed and duration of writing	"How far the pen tip moved across the surface, reported in centimetres…(and) how long the pen moved, reported in milliseconds" (p. 125)	"The physical and temporal demands of different notetaking choices"(p. 126)

As discussed in Chapter 2, Section 2.2.1.3, a prolonged EPS does not necessarily correspond to an increase in the cognitive effort of note-planning. Instead, it might be a strategic decision of the interpreter to prioritize listening comprehension over note-taking during the input stage of CI. Hence, in the present study, the duration of EPS would not be over-interpreted as the amount of cognitive effort. Instead, it is only taken to represent the duration of note-planning, i.e., the time interpreters need to make note decisions. In addition, because of the functional differences in the adopted digital pens, click count (CC), which is the number of strokes during note-writing, is used in this study to indicate the physical effort of note-writing (see Table 4-8). This stroke-based indicator can fit well with the Chinese language where characters can differ significantly in the strokes of writing. For instance, "*饕餮盛宴*" and "*丰盛大餐*" both mean a "feast" in Chinese. Whereas the former requires 62 strokes of writing, the latter only involves 34. Therefore, writing the first phrase will be more physically demanding than writing the second one; and this difference could be indicated by using CC as the measurement. Finally, pen-tip pressure is not adopted as an indicator of psychological pressure in this study, as it can vary significantly across participants because of personal habits. Instead, the participants were asked to report their frustration level in a NASA-TLX questionnaire after interpreting, which is a reliable measure of workload that has been frequently adopted in translation studies (e.g., Sun & Shreve, 2014; Weng & Zheng, 2020).

Table 4-8. Pen-recording measures adopted in the present study

Measures	Operational definition	Corresponding activities
Ear-pen span (EPS)	The interval between the onset of a speech unit that is articulated and the start of the corresponding note-writing act	Time to make a note decision
Click count (CC)	The number of strokes involved in note-writing	The physical effort of note-writing

It is worth mentioning that, in S. Chen's work, EPS is calculated based on the offset of the source speech units and the onset of the note-taking acts to prevent potential word length effects. However, research has found that word processing can start from a sublexical level (e.g., Kolinsky, 1998). The acoustic properties of one phoneme can also be co-activated with the properties of adjacent segments, boosting "the mapping between the acoustic signal and the mental lexicon" (Kolinsky, 1998, p. 1). To avoid such potential impacts of sublexical processing on note-planning, the EPS data in the present study are calculated based on the onset of source speech units and the onset of note-taking acts. Specifically, as the pen tip on the tablet is recognized as a mouse click by the eye-tracking software, the measure of “Time to First Click” which records the time that the pen tip touched the table was adopted to indicate note-taking onset. In addition, the onset of source speech delivery was decided based on automatic sound recognition technology and manual corrections.

4.3 Analysis of note-taking and interpreting products

4.3.1 Descriptive patterns of note choices

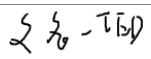
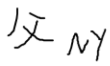
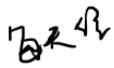
The participants’ notes were analysed with a descriptive approach from three aspects: note quantity, note form, and note language. Based on S. Chen’s (2017c) categorization of notes, which specifically looks into note-taking in Chinese-English interpreting (Table 4-9), notes are firstly classified into language, symbol and number in the present study. Then, language notes are further grouped into full words and abbreviations according to the form of notes, as well as Chinese and English notes based on the language of notes. Since the participants’ note quantities varied, the distribution of different note forms and note languages in the product of note-taking is presented in proportions rather than in counts.

Table 4-9. Classification of notes based on S. Chen (2017c, p. 10)

Category	Definitions from S. Chen (2017c)	Examples from this study
Full word	“A full word is a Chinese or English word written in full, including words both with and without morphemes of inflection”.	• “Father” and “父亲”
Abbreviation	“An abbreviation consists of parts of the letters of a long English word, or part of the characters of a long Chinese word, or the phonetic spelling of a word, including: (1) real abbreviations (i.e., units in which only part of a word is represented); (2) acronyms; (3) other short forms that cannot be characterised either as real abbreviations or as acronyms, but rather as something in between”.	• “Fa” / “父” for “father” • “NY” for “New York” • “THX” for “Thanks”
Symbol	“A symbol is a representation of (1) the underlying meaning of a word or expression rather than the actual word or expression; or (2) the relationship(s) between two units. Symbols are mostly pictorial, but they can also be a pair of letters, a single letter, or (part of) a Chinese character”.	• Signs like arrows (“→”) for casual relationships • Letter “%” for “a part of” • Chinese character “走” for “跑(run)”
Language	“The combination of full words and abbreviations. Further divided into Chinese and English”.	• Full words like “father” • Abbreviations like “fa” for “father”
Number	“A special language where its implied meaning and verbal surface detach from each other”.	• “18” for “the 18th century”

In addition, based on Albl-Mikasa (2006), three note-taking strategies are identified in the present study: ellipsis, replacement and restructuring (Table 4-10). As mentioned in Albl-Mikasa (2006), an ellipsis strategy refers to a simple reduction of source speech contents where notes follow the source expressions closely. In addition, a restructuring strategy refers to a replacement of the source syntactic structures with new ones in notes. As for the replacement strategy, this refers to using different lexis to represent the same meaning of a source speech unit during note-taking. For instance, in the source speech in the present study, "Ted" is the name of the speaker's father. In the collected notes, "he" and "父" (an abbreviation of "father" in Chinese) were frequently used in to represent "Ted". For this type of note, they are grouped as "replacement" during the analysis of the note-taking strategy.

Table 4-10. Examples of the ellipsis, replacement and restructuring strategies of note-taking

Strategy	Source speech	Notes	Explanation
Ellipsis	…because my father's name is Ted.		• 父: abbreviation of "father" (父亲) in Chinese • 名: abbreviation of "name" (名字) in Chinese • Ted: full name for "Ted"
Replacement	Now Ted was a New Yorker…		• 父: abbreviation of "father" (父亲) in Chinese • NY: abbreviation of "New York"
Restructuring	…what happens every day.		• 每天: "every day" in Chinese • 经: abbreviation of "happens" ("经过") in Chinese

4.3.2 Summative assessment of interpretation quality

There are four major assessment methods of interpreting performance: 1) impression scoring; 2) error counts; 3) checklists; and 4) analytic rating scales (S. Lee, 2015). Among these, checklists are frequently adopted in peer assessment and self-assessment to monitor progress. This helps interpreters to check what they have or have not achieved by ticking "yes or no" in a list of interpreting requirements (S. Lee, 2015). Impression scoring is also known as holistic scoring (e.g., Hunter et al., 1996), and can be easily conducted by simply asking raters to give a score to an interpretation. However, it is not widely adopted in academic research because the scores can be unreliable when the raters do not follow the same principles of rating. By comparison, error counts can be very "reliable", especially "when the assessment requires the 'principle-based' and 'mechanical' processing of responses" (S. Lee, 2015, p. 229), such as in a lexical accuracy test in Liu and Chiu (2009). However, an important issue with this assessment method is that it neglects the textual and pragmatic aspects of interpretation (Green, 2014, p. 143). Even if an interpreter can interpret every word accurately, the target speech might not make sense as a coherent whole. Therefore, this assessment method should be cautiously adopted in judging an interpreter's performance. By comparison, analytic rating scales can solve this issue by asking raters to give scores on each concerned aspect of performance based on certain pre-decided criteria (e.g, Angelelli, 2009; C. Han, 2015; S. Lee, 2015). However, it is worth mentioning that rating scales will always be confined to the pre-determined aspects of performance (Marquardt & Gillam, 1999). This requires researchers to include all the necessary aspects of performance assessment into consideration. At the same time, comprehensive instructions and adequate training should be provided for the raters to ensure the high reliability of the scores.

Based on the discussions above, rating scales were adopted in the present investigation to assess the participants' interpreting performance. C. Han (2018) provides a comprehensive review of the scale categories, scale bands, scalar descriptors and scale types that have been adopted in interpreting performance assessment, after reviewing relevant literature in 11 SSCI- and/or A&HCI-indexed translation and interpreting journals. Firstly, there are three aspects of performance that raters attach much importance to: content, delivery and language quality. Secondly, the number of scale bands is usually 7 plus or minus 2. Sometimes, the number can be as large as 20 when researchers intend to achieve a finer-grained differentiation (Rosier et al., 2011). Thirdly, there should be at least two raters (e.g., Angelelli, 2009; Campbell & Hale, 2003) who have extensive interpreting-related education and work backgrounds as well as interpreting assessment experience. Fourthly, rater training is indispensable in the application of analytic rating scales. It should include an introduction to the rating scales, a pilot session of rating, and a following norming session for benchmark setting, as well as a discussion session for rater-to-rater communication. Following these standards, the present study follows C. Han (2019) and invited two experienced language experts with abundant interpreting experience to assess the participants' interpreting performance from three aspects: information completeness (InfoCom), fluency of delivery (FluDel) and target language quality (TLQual) (Table 4-11). The raters participated in a training session to become familiar with the scale. Then, they went through a pilot rating session with five randomly-selected interpretations. After an online discussion about the discrepancies, they finally proceeded to the main scoring task. The total score for each aspect of quality assessment is eight, and it is evenly distributed into four bands. The raters were always asked to first decide which band the interpretation belonged to and then give a specific score. Finally, InfoCom was given a weight of 2, and the other two measures were each given a weight of 1 (S. Lee, 2015). The sum of these scores comprises the total score of interpreting.

Table 4-11. Descriptor-based rating scales for CI performance assessment proposed by C. Han (2019, p. 20)

Band/ Scoring criteria	**Information completeness (InfoCom)**	**Fluency of delivery (FluDel)**	**Target language quality (TLQual)**
Band 4 (Score range: 7–8)	A substantial amount of original messages delivered (i.e., > 80%), with a few number of deviations, inaccuracies, and minor/major omissions.	Delivery on the whole fluent, containing a few disfluencies such as (un)filled pauses, long silence, fillers and/or excessive repairs.	Target language idiomatic and on the whole correct, with only a few instances of unnatural expressions and grammatical errors.
Band 3 (Score range: 5–6)	Majority of original messages delivered (i.e., 60–70%), with a small number of deviations, inaccuracies, and minor/major omissions.	Delivery on the whole generally fluent, containing a small number of disfluencies.	Target language generally idiomatic and on the whole mostly correct, with a small amount of instances of unnatural expressions and grammatical errors.
Band 2 (Score range: 3–4)	About half of original messages delivered (i.e., 40–50%), with many instances of deviations, inaccuracies, and minor/major omissions.	Delivery rather fluent. Acceptable, but with regular disfluencies.	Target language to a certain degree both idiomatic and correct. Acceptable, but contains many instances of unnatural expressions and grammatical errors.
Band 1 (Score range: 1–2)	A small portion of original messages delivered (i.e., < 30%), with frequent occurrences of deviations, inaccuracies, and minor/major omissions, to such a degree that listeners may doubt the integrity of renditions.	Delivery lacks fluency. It is frequently hampered by disfluencies, to such a degree that they may impede comprehension.	Target language stilted, lacking in idiomaticity, and containing frequent grammatical errors, to such a degree that it may impede comprehension.

4.4 Summary

This chapter examines the overall research methodology adopted by the present study. Firstly, in order to control the difficulty level of the source speech segments, both objective and subjective methods were applied in the assessment of interpreting difficulty. The objective assessments included readability and idea density measures, and the subjective assessments consisted of expert judgment and participant feedback. Secondly, eye-tracking and pen-recording methods were applied in this study to measure interpreters' note-taking effort and note-reading effort. The desktop-based Tobii Pro Fusion 250 was selected to collect interpreters' eye movements on the screen in a remote manner. In addition, a Wacom CLT 672 digital pen that could synchronize the writings on the tablet with the computer screen was used to track interpreters' pen movements. A range of eye-tracking and pen-recording measures were adopted to indicate the overt visual attention, cognitive effort, physical effort and temporal management during the two noting processes. Thirdly, the product of note-taking is analysed based on the note choices (note quantity, note form, note language) and note-taking strategies (ellipsis, replacement and restructuring). Finally, the product of interpreting is assessed in a summative manner from three aspects (InfoCom, DelFlu and TLQual) with two experienced raters. The overall methodology of the present study is displayed in Figure 4-11.

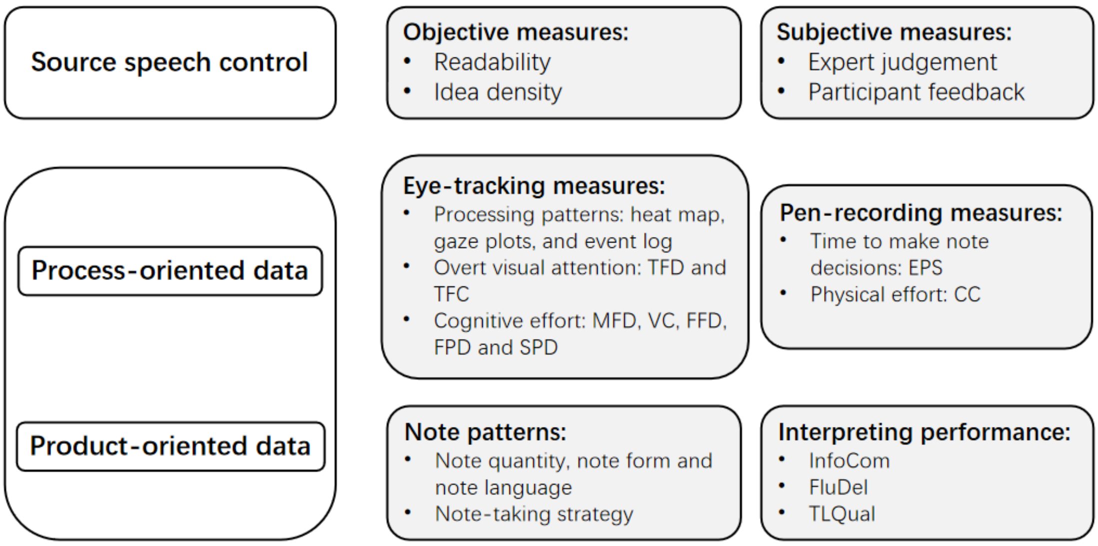


Figure 4-11. A graphic illustration of the methodology of the present study

Chapter 5: Experimental Setup and Data Processing

The present study adopted a between-subject and within-subject design. 20 professional interpreters and 29 student interpreters consecutively interpreted two easy segments and two difficult segments of a speech from English (L2) to Chinese (L1). Their eye fixations on the screen and pen movements on the tablet were registered, respectively, by a Tobii Pro Fusion 250 eye-tracker and a Wacom CLT 672 digital pen. This chapter presents the details of the experimental design, including the participants' background (Section 5.1), the manipulation of the source speech (Section 5.2), and the adopted apparatus and the experimental procedures (Section 5.3). In addition, details about the quality assessment of the eye-tracking data are provided in Section 5.4., and the statistical analysis methods are discussed in Section 5.5. Finally, Section 5.6 provides a summary of this chapter.

5.1 Participants

The participants were recruited through voluntary response and snowball sampling methods. This started from the researchers' professional and academic networks and developed in turn through the networks of the first group of qualified participants. Altogether, 24 professional interpreters (18 females and 6 males) with 7.29 years (*SD*=3.24 years) of interpreting work experience and 31 master's students (27 females and 4 males) from 11 Chinese universities participated in the formal experiment. All the participants had a normal or corrected-to-normal vision. The average age for the professional group was 33.58 years old (*SD*=6.16), and that for the student group was 23.22 years old (*SD*=2.17). Most of the professionals had a master's degree in interpreting (87.5%). Three who had learned interpreting from professional training courses (12.5%) had passed the Level-II test of the China Accreditation Test for Translators and Interpreters. On average, the professionals had received 1.64 years' (*SD*=0.68) interpreting training, including in systematic note-taking. The students had just finished their first-year master's program of Translation and Interpreting which included intensive CI training and note-taking practice. All the participants have Chinese as L1 and English as L2. They signed a consent form clarifying their full anonymity and confidentiality and were rewarded with a supermarket gift token upon task completion. The research was approved by the research ethics committee of Durham University, and the project ID is MLAC-2019-06-13T14:42:41-tzcw84.

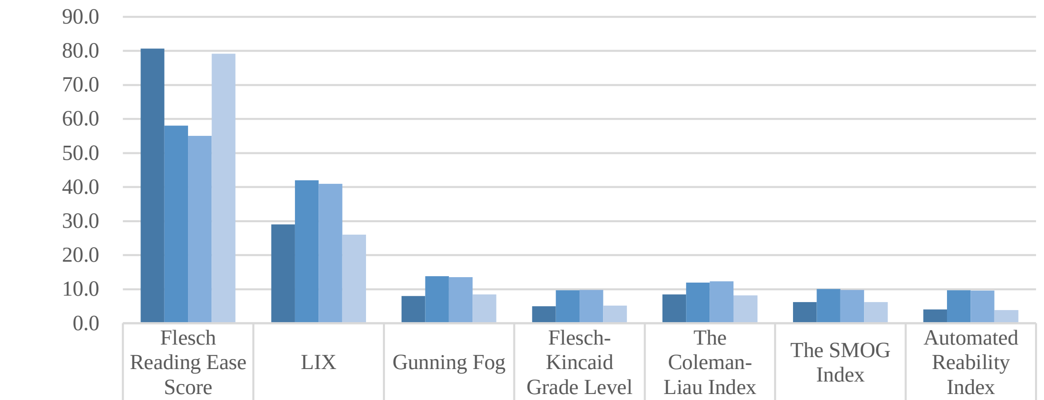
The COVID-19 pandemic broke out on 31 December 2019, and the formal experiment of the present study was conducted between July and December 2020. At the time of the experiment, all the professionals had started to conduct VRI in their work, and the students had attended video-mediated interpreting classes online for one semester. From this perspective, both groups of participants had extensive experience in interpreting speech presented through a computer screen.

The participants were grouped by their professional work experience rather than their educational backgrounds, because the present study aims to explore how interpreter work experience affects note-taking in VRI. As discussed in Chapter 3, Section 3.1.3, WM is an important factor in the cognitive load of note-taking and interpreting. Significant differences in the general WM capacity are more frequently observed between beginner student interpreters and expert interpreters, rather than between advanced student interpreters and professional interpreters (Wen & Dong, 2019). However, even with a similar capacity of WM, professional interpreters still can outperform student interpreters in note-taking by creating larger note quantities (S. Chen, 2022), using more notes in the TL (Abuín González, 2012), reacting to a secondary task faster during note-taking (Hu, 2008), and achieving better interpreting performance (S. Chen, 2022). Exploring the reason behind this professionalism is thus significant for interpreter training and interpreting practice. Based on Wen and Dong's (2019) criterion, the student interpreters in the present study should be categorised as advanced student interpreters, and the professional interpreters should be identified as expert interpreters. It was expected that there would be no significant differences between the two groups of participants' general WM capacity. In that case, the group differences observed in the note-taking behaviour would not be attributed to the differences in the participants' WM capacity. The two groups of participants' WM spans were tested with an English listening span task, which is adapted by Cai et al. (2015) from Daneman and Carpenter's (1980, 1983) reading span task. Test details and test results can be found in Sections 5.3 and 5.4.

5.2 Material

The source speech was selected and edited based on three principles. Firstly, since this study aims to investigate the effects of source speech difficulty on interpreters' note-taking behaviour, the source speech should contain segments that entail at least two levels of interpreting difficulty. Secondly, simply presenting easy segments before difficult segments, or the other way around, could lead to potential order effects and false experimental results. Therefore, the speech segments should be sequenced in a way that can minimize such order effects. Thirdly, to ensure the authenticity of the source speech, it should be a speech that is delivered on a real occasion, and the manipulation of the speech should be maintained at a minimum level. With these considerations in mind, a TED talk that introduces classical music in life in plain language was selected. The speech was originally 19 minutes and 21 seconds long, but it is very difficult for participants to stay still in front of an eye-tracker for such a long time. Therefore, around seven minutes of the speech were removed based on two experienced interpreter trainers' advice, with the principle of ensuring a logical development of the content. The remaining speech was divided into four segments with relatively independent topics. After transcription, an initial round of readability tests and idea density tests were conducted, and the results show that the four speech segments followed in an easy-difficult-difficult-easy sequence. Therefore, only minor adjustments were made to the video clips to further ensure that the induced difficulty of interpreting was as planned in the research design.

After adjustments to the source speech, the second round of reliability and idea density tests were conducted to ensure that: 1) Segment 1 and Segment 4 were easier to interpret than Segment 2 and Segment 3; and 2) the two segments that are designed at the same difficulty level (i.e., Segments 1 and 4; Segments 2 and 3) were comparable to each other in the test results. The readability tests show that, for all the seven tested measures, Segment 2 and Segment 3 returned values and grade levels that represent higher difficulty levels of comprehension than Segment 1 and Segment 4 (Figure 5-1). At the same time, the segments within each pair of interpreting difficulty, namely Segments 1 and 4 in the "easy" pair and Segments 2 and 3 in the "difficult" one, returned similar scores in all the measures of readability. In addition, the idea densities of the four segments are 0.54, 0.57, 0.57 and 0.55, respectively, suggesting that the distribution of information is denser in the middle segments than in the segments at the two ends (Table 5-1). Overall, both readability and idea density test results indicate that the four segments as mentioned above follow in an easy-difficult-difficult-easy sequence, which was in accordance with the research design. Detailed information about the speech segments can be found in Table 5-2 (see Appendix 1 for transcripts). Overall, the four segments had balanced word count (M=399.00, SD=12.68), speech duration (M=163.75 seconds, SD=5.25 seconds) and delivery rate (M=146.20 wpm, SD=1.00 wpm).



	Flesch Reading Ease Score	LIX	Gunning Fog	Flesch-Kincaid Grade Level	The Coleman-Liau Index	The SMOG Index	Automated Reability Index
Segment 1	80.7	29.0	8.0	5.0	8.5	6.2	4.1
Segment 2	58.0	42.0	13.8	9.7	11.9	10.1	9.7
Segment 3	55.0	41.0	13.5	9.8	12.3	9.8	9.6
Segment 4	79.2	26.0	8.5	5.2	8.2	6.2	3.9

Figure 5-1. Readability scores of the four source speech segments

Table 5-1. Details about the source speech segments

Segment number	Difficulty level	Word count	Proposition count	Idea density	Speech duration	Delivery rate
1	Easy	406	220	0.54	2’45”	147.63
2	Difficult	380	217	0.57	2’36”	146.15
Opera clip interval						
3	Difficult	405	230	0.57	2’47”	145.51
4	Easy	405	222	0.55	2’47”	145.51

It is worth mentioning that a 35-second opera clip from the original speech was kept between Segment 2 and Segment 3 for the purpose of ensuring the logical development of the speech and removing the possible spill-over effects that Segment 2 (difficult) could exert on Segment 3 (difficult). Moreover, this separation between the first two segments and the last two segments of interpreting allows the researcher to explore whether the effects of source speech difficulty on interpreters’ note-taking behaviour would be mediated by the presentation sequence. In other words, a difficulty-increase direction of presentation (from Segment 1 to Segment 2) and a difficulty-decrease direction of presentation (from Segment 3 to Segment 4) could exert different impacts on interpreters’ note-taking behaviour in interpreting.

In addition to the objective assessment methods of source speech difficulty, 6 university interpreting lecturers who were also part-time interpreters (10.5 years of teaching and interpreting experience on average) and 4 freelance interpreters (6 years of professional experience on average) were asked to rate the difficulty of interpreting the segments from eight aspects: words, syntactic structure, information density, coherence, logic, clarity, abstractness, and required background knowledge (Liu & Chiu, 2009, p. 248). The rating was conducted with a five-level Likert scale, from 1 being "very easy" to 5 being "very difficult". Results show that Segments 2 and 3 always received higher scores than Segments 1 and 4 (Table 5-2), indicating greater difficulty in interpreting the former segments than the latter ones. In addition, paired *t*-test results demonstrate that many of the observed differences between the "easy" and "difficult" segments reached a significant level (Table 5-3). At the same time, Segment 1 and Segment 4 received similar scores in all the assessed aspects of interpreting difficulty, as did Segment 2 and Segment 3. Paired *t*-test results present no significant differences between the two segments that are designed at the same difficulty level in these ratings. Taken together, the expert judgement results present that the difficulty levels of interpreting the four speech segments follow an easy-difficult-difficult-easy manner as framed in the research design.

Table 5-2. Experts' judgement of the difficulty level of interpreting the four source speech segments

	Segment 1		**Segment 2**		**Segment 3**		**Segment 4**	
	M	SD	M	SD	M	SD	M	SD
Word difficulty	2.55	0.90	3.65	0.88	3.35	0.75	2.40	0.70
Syntactic difficulty	2.40	0.84	3.15	0.75	3.45	0.50	2.50	0.53
Information density	2.90	0.74	4.10	0.74	4.15	0.88	2.60	0.70
Coherence	2.00	1.15	2.15	1.06	2.30	1.16	2.10	0.99
Logic	2.00	1.05	2.05	1.12	2.25	1.27	2.05	0.96
Clarity	1.90	0.88	2.80	0.92	2.45	1.01	1.90	0.88
Abstractness	2.30	1.16	3.25	0.98	3.00	1.25	2.10	0.74
Knowledge difficulty	2.40	0.97	4.00	0.94	3.95	0.60	2.10	0.88

Table 5-3. Paired sample *t*-test results for the expert scores in speech segments at different levels of interpreting difficulty

	Segment 1-Segment 2				**Segment 3- Segment 4**			
	From easy to difficult				*From difficult to easy*			
	MD	*t*	*df*	Sig.	MD	*t*	*df*	Sig.
Word difficulty	-1.10	6.128	9	<.05	0.95	5.019	9	<.05
Syntactic difficulty	-0.75	3.308	9	<.01	0.95	4.385	9	<.05
Information density	-1.20	9.000	9	<.01	1.55	4.841	9	<.01
Coherence	-0.15	0.410	9	>.05	0.20	0.802	9	>.05
Logic	-0.05	0.130	9	>.05	0.20	0.612	9	>.05
Clarity	-0.90	3.857	9	<.05	0.55	2.905	9	<.05
Abstractness	-0.95	5.019	9	>.05	0.90	2.862	9	<.05
Knowledge difficulty	-1.60	0.600	9	<.001	1.85	10.091	9	<.001

Note. MD=mean difference

Finally, a pilot study was conducted with 10 MA students (coded from S_1 to S_{10}) in Translation Studies from Durham University to confirm the consistency between the designed and the perceived difficulty levels of the four interpreting segments. The participants went through the same experimental procedure as in the formal study[13] (see Section 5.3). They were provided with a glossary list which was prepared by the researcher based on the ten raters' advice. After they became familiarized with the vocabulary, the eye-tracker and the digital pen, they interpreted the four speech segments in succession (with the 35-second opera clip played between Segments 2 and 3), and then they filled in a NASA-TLX questionnaire to report interpreting difficulty from four aspects: mental demand, effort, frustration and performance. Results show that, compared with Segments 1 and 4, Segments 2 and 3 entailed more mental demand, greater effort, higher frustration and poorer performance. Meanwhile, the two segments at the same difficulty level (Segments 1 and 4 as easy segments and Segments 2 and 3 as difficult segments) received similar scores in the four NASA-TLX items. The detailed results are presented in Table 5-4.

Table 5-4. The NASA-TLX scores collected in the pilot study

	Segment 1		**Segment 2**		**Segment 3**		**Segment 4**	
	M	SD	M	SD	M	SD	M	SD
Mental demand	3.70	1.06	5.00	1.05	5.20	0.79	4.50	0.85
Effort	3.90	1.10	5.30	0.82	5.40	0.84	4.60	0.84
Frustration	3.50	0.71	5.00	0.94	5.00	1.05	4.50	1.08
Performance	3.80	1.14	3.20	1.03	3.20	1.32	3.60	1.17

Paired sample *t*-tests were conducted to examine whether the observed differences between the easy and difficult speech segments in the NASA-TLX scores reached a significant level (Table 5-5). Overall, many significant results are observed in the *t*-tests, but no significant differences are found in the *Effort* and *Performance* scores. These insignificant results could be attributed to three reasons. Firstly, researchers have found that interpreters with higher linguistic competence are more sensitive to self-expectations than those with lower linguistic competence (Jiménez Ivars et al., 2014). Since only student interpreters were involved in the pilot study, differences in the *Performance* scores might be obscured because of their comparatively low linguistic competence. Secondly, according to the participants' feedback, a Likert scale with only five levels might prevent them from making accurate judgments. S5 and S10 suggested increasing the range of the scale to ten, and S2 and S4 advised allowing one decimal in the ratings. Thirdly, S2 and S3 also mentioned that the 35-second music clip between

[13] At the time of the pilot study, the Tobii Pro Fusion 250 eye-tracker was not yet available in the university, so the participants completed the experiment with a PC-based Tobii TX300 eye-tracker.

Segment 2 and Segment 3 could be extended thus further easing their fatigue caused by interpreting the first two segments. Based on such feedback from the participants in the pilot study, two adjustments were made to the experimental materials: (1) the Likert scale was adjusted to ten levels and one decimal was allowed during rating; and (2) the music clip was slowed down by approximately 28.57%, extending the length of the opera clip to from 35 seconds to 45 seconds. The rest of the experimental material remained unchanged.

Table 5-5. Paired sample *t*-tests for the NASA-TLX scores collected in the pilot study

	Segment 2-Segment 1				**Segment 3- Segment 4**			
	From easy to difficult				*From difficult to easy*			
	MD	*t*	*df*	Sig.	MD	*t*	*df*	Sig.
Mental demand	1.30	2.751	9	<.05	.70	1.91	9	<.05
Effort	1.400	3.221	9	<.01	.80	2.121	9	>.05
Frustration	1.50	4.025	9	<.01	.50	1.048	9	<.05
Performance	-.60	-1.236	9	>.05	-.40	-0.717	9	>.05

Note. MD=mean difference

5.3 Apparatus and procedures

Because of the COVID-19 pandemic, the Tobii Fusion Pro 250 eye-tracker in the University was not accessible at the time of the formal study. With the help of Tobii China, the same type of eye-tracker was provided for the formal study. To confirm that this eye-and-pen approach was applicable and suitable for the study design, a second pilot study was conducted with 10 students from the Master of Translation and Interpreting (MTI) programs. The students' eye movements were recorded with the eye-tracker throughout the interpreting task, and their pen movements were recorded using the Wacom CTL 672 digital pen and the equipped tablet (Figure 5-2). All of the devices were connected to a laptop with a 15.6-inch screen which could simultaneously show interpreters' writing on the tablet. The screen presentation was set with two areas: a speaker window for video-playing, and a blank area for note-taking. Audacity 2.4.4 software was used for voice recording throughout the experiment.

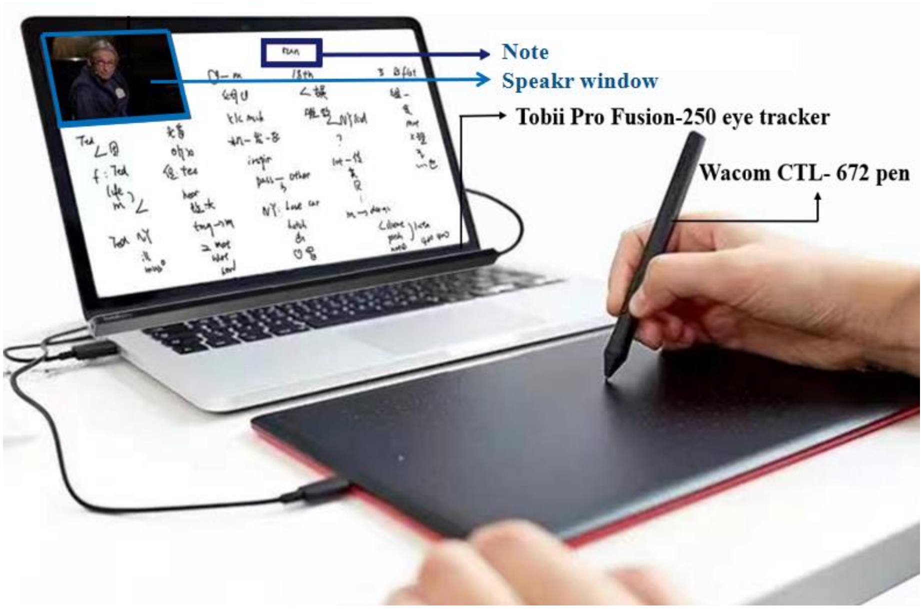


Figure 5-2. The experimental setup for the second pilot study and the formal study

During the retrospective interview after the interpreting task, all the students claimed that they were able to adapt to this experimental setting, which they attributed to two factors. Firstly, the time-free practice session before the main interpreting task, which included information copying, vocabulary learning and a warm-up CI exercise with the digital pen (Figure 5.3), greatly helped them to familiarize themselves with the experimental setting. On average, the participants could start the main interpreting session after 18 minutes 27 seconds' (*SD*=5 minutes 10 seconds) practice with the use of the digital pen and the eye-tracker. Secondly, at the time of the experiment when many areas of China were under lockdown, the students had over one semester's online training experience. They had gotten used to conducting interpreting through a computer screen, and this also greatly helped them to overcome the hand-eye coordination problems caused by the eye-tracker and the digital pen. They described that, when they were used to this experimental setting, it was easier than blind typing in translation. For these reasons, this experimental setting was kept in the formal study and the 10 students' data was considered valid for later processing.

Experimental procedures are illustrated in Figure 5-3, and instruction details are presented in Appendix 2. Firstly, to minimise any negative influences on data quality, the participants were required not to drink alcohol in the 24 hours before the experiment. On the experiment day, they were asked to first sign a consent form (see Appendix 3) of participation. Then, the participants were asked to complete an English (L2) listening span task (cf. Cai et al. 2015) which is adapted from Daneman and Carpenter (1980,

1983) to measure their WM capacity. During the test, the participants listened to 60 English sentences (8-12 words) which were divided into different sets (3-6 sentences). "Of these 60 sentences, 30 were either syntactically or semantically incorrect (e.g. 'Being an environmentalist, I like to newspaper plastic bags'), with the other 30 being grammatically and semantically acceptable (e.g. 'Kate should do well in school because she is a bright child.')" (Cai et al., 2015, p. 108). The participants had to judge whether the sentence made sense by pressing F (for True) and J (for False) on the keyboard while remembering the last word of the sentence. After finishing a set of sentences, they needed to orally recall all the last words of the sentences in the last set. Finally, the total number of correctly recalled words in the test (60 at maximum) is the participant's score. The presentation sequence of the sentences was randomly decided by the software, called E-prime 2.0. After the listening span task, the participants were allowed to have a rest without a time limit before proceeding to the experiment.

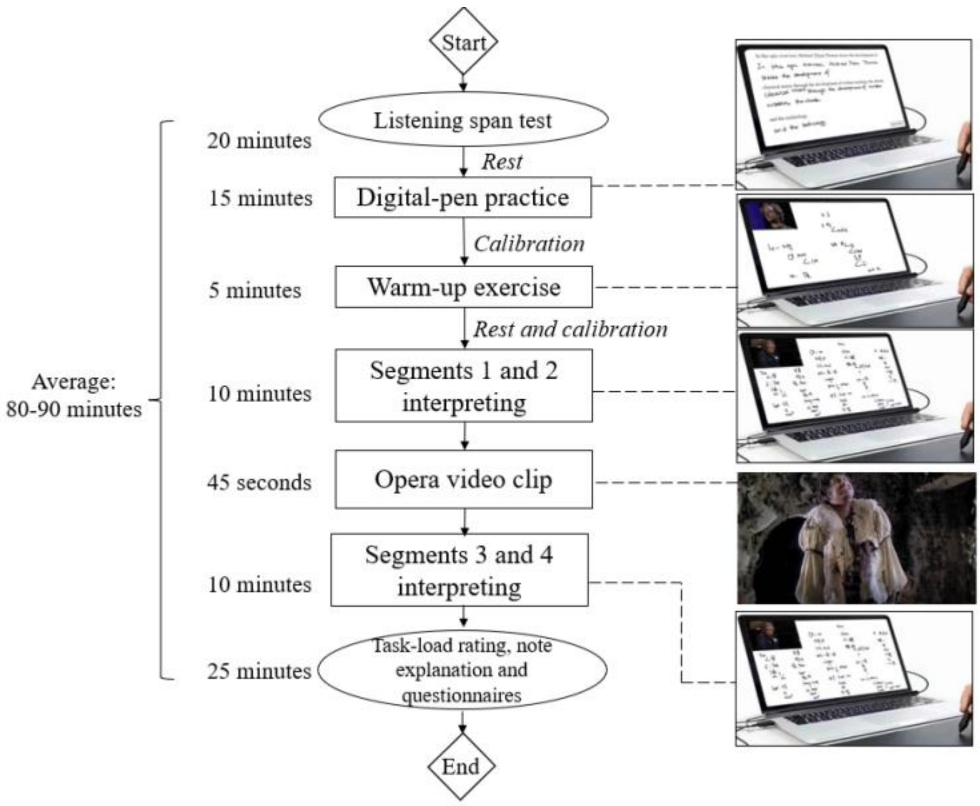


Figure 5-3. Experimental procedures of the formal study

In the main experiment, the participants first went through a familiarization session with the digital pen and the eye-tracker. They copied the gist of the speech and the background information of the speaker with the digital pen. They then familiarised themselves with the prepared vocabularies of the speech by copying, circling or creating symbols with the digital pen on the screen without time limits. After that, they had a nine-point eye calibration by sitting approximately 60cm away from the eye-tracker, and did a one-minute warm-up exercise of CI with the digital pen and the eye-tracker. After ensuring that their raw sample rates during the warm-up task exceeded 85%, the participants were asked to conduct a second calibration and proceeded to the main interpreting task. If the raw sample rate in the warm-up practice fell below 85%, technical advice would be provided and a second one-minute exercise would be required. Proceeding to the main interpreting task, the participants interpreted Segments 1 and 2, watched the 45-second opera clip and interpreted Segments 3 and 4 in succession. Eye-tracking stopped upon the completion of the interpreting task. Then, they scored the difficulty level of interpreting each segment with a NASA-TLX questionnaire (cf. Sun & Shreve, 2014) (see Appendix 4), evaluated the helpfulness of note-taking (from 0 to 10), and identified the difficulties they encountered during note-reading. After that, they were asked to explain their notes, including but not limited to the form (language, symbol or number), language (Chinese or English), and corresponding source and target speech units of the notes. Finally, they completed a questionnaire regarding their demographic information and interpreting background (see Appendix 5). The experiment lasted approximately 80 to 90 minutes. The participants were rewarded with a supermarket gift token upon task completion.

5.4 Assessment of data quality

Before data analysis, the quality of the eye-tracking data was assessed with Hvelplund's (2011) three criteria: MFD, GTS and GSF. The standards of the three criteria in the present study are calculated in the following. Data points that were lower than one SD of the sample's mean would be considered invalid:

1) Mean fixation duration (MFD): [total fixation duration ÷ total fixation count]

   The MFD of the participants can be directly exported from the eye-tracking software of Tobii Pro Lab. Based on the collected data, the threshold in the present study is 276 ms, which is comparatively higher than that of some translation tasks (around 200 ms in Cui & Zheng, 2020; Hvelplund, 2011; Sjørup, 2013). This can be caused by the high cognitive demand of VRI (e.g., Doherty et al., 2022), which underlines a different cognitive mechanism from translation tasks. Overall, seven participants did not meet this criterion.

2) Gaze time on screen (GTS): [total fixation duration ÷ total task time × 100%].
The GTS standard for the present study is 51.15%, which is slightly higher than 46.40% in translation tasks (Cui & Zheng, 2020). Ultimately, six participants were found not to meet this criterion.

3) Gaze sample to fixation percentage (GSF): [the total number of gaze samples ÷ the total number of gaze samples that formed part of a fixation]
After calculations, 76.58% was adopted as the GSF standard for the present study. Similar standards can be found in translation studies, such as 75.00% in Hvelplund (2011) and 74.67% in Cui and Zheng (2020). Altogether, five participants did not reach this standard of inclusion.

Data points of P04, P15, P33 and P59 were excluded from the analysis because they did not meet at least two of the three criteria. Table 5-6 presents the detailed information about the four excluded data points.

Table 5-6. Details about the data points that are excluded from the data analysis

	MFD	GTS	GFP
Threshold	276	51.15%	76.58%
P04	272	30.02%	83.96%
P15	223	30.96%	67.58%
P33	335	78.29%	90.83%
P59	279	37.88%	82.88%

Note. MFD is measured in milliseconds.

In addition, P42 and P53 were excluded from the analysis for reporting discomfort with using the digital pen. Overall, the percentage of invalid data was 10.91%. Eventually, this study analysed the data obtained from 20 professionals and 29 students (Table 5-7). The two groups' WM capacity was found to be similar ($t(47)=-1.612, p>.05, d=0.48$) through the L2 listening span task. All the students passed the Test for English Majors Band 8 with an average score of 71.00[14] ($SD=4.88$). It is worth mentioning that the gender skew (female-dominated) in sampling was not deliberately designed, but

[14] There are four bands in TEM8: "excellent" (score between 80 and 100), "good" (score between 70 and 79), "pass" (score between 60 and 69) and "failed" (score lower than 60).

this was not expected to exert decisive impacts on the conclusions (Hvelplund, 2011).

Table 5-7. Details about the participants who were included in the data analysis

	Professionals	**Students**
Number	20 (17 females and 3 males)	29 (25 females and 4 males)
Age	32.9 years (*SD*=4.30)	23.3 years (*SD*=2.18)
Interpreter work experience	7.3 years (*SD*=2.99)	--
English listening span	34.95 (*SD*=7.85)	31.83 (*SD*=6.02)

On the other hand, for the EPS data which are calculated based on the pen-recording measures, "extreme values are often observed in cases such as note additions at the end of a source speech segment" (S. Chen, 2020b, p. 126). Therefore, following S. Chen (2020b), extreme EPS values that were three SD higher or lower than the mean of the participant's EPS in each segment of interpreting were excluded. Overall, approximately 95% of the notes had EPS data.

As for the interpreting products, the reliability of the interpreting scores was checked for each aspect of the assessment (InfoCom, FluDel and TLQual) via a Pearson's correlation test with a 95% confidence interval. Results show that all the scores given by the two raters are significantly correlated (Table 5-8). Although the coefficients for TLQual are noticeably lower than those of the other two aspects, the scores were not deliberately "corrected" because such phenomena have also been observed in interpreting assessment studies (C. Han, 2019). Moreover, the correlation coefficients for the total interpreting scores are above 0.70, indicating strong associations (Cohen, 1992) between the two raters. Therefore, the ratings were finally accepted and included in the data analysis.

Table 5-8. Pearson's correlation test results of the interpreting scores provided by the two raters

	InfoCom	**FluDel**	**TLQual**	**Total score**
Segment 1	0.726***	0.736***	0.527***	0.769***
Segment 2	0.719***	0.752***	0.545***	0.731***
Segment 3	0.724***	0.846***	0.489***	0.736***
Segment 4	0.715***	0.747***	0.438***	0.725***

Note. * p<.05, ** p<.01, and *** p<0.001. This applies to all tables.

5.5 Statistical analysis

Before conducting statistical analysis, data normality was checked through the Shapiro-Wilk test with the two-tailed *p-value* at 0.05 setting as the bar. For between-subject comparisons, independent *t*-tests were conducted for normally distributed data and

Mann-Whitney tests were adopted for non-normally distributed data. For within-subject comparisons, paired *t*-tests were applied for normally distributed data and the Wilcoxon signed rank tests were used for non-normally distributed data.

In addition, Pearson's correlation tests were adopted to explore the relationships between interpreters' note-taking behaviour and CI quality. It is worth mentioning that there was a proportion of the data that was not normally distributed. However, conducting parametric (Pearson's) and non-parametric (Spearman's) correlation tests at the same time would make the results incomparable. According to Tabachnick and Fidell (2007), when the group size ratio is less than 4:1 and F_{max} is less than 10, non-normally distributed data can be tested with parametric tests. The ratio of group sample size is 29:20, and the F_{max} of all datasets was within 1.00-1.50. Therefore, Pearson's correlation tests were conducted to examine the relationships between interpreters' note-taking behaviour and CI quality. Additional non-parametric tests were also conducted to ensure that Pearson's and Spearman's tests returned the same results; Spearman's test results are only reported when the two correlation tests returned different results.

5.6 Summary

This chapter introduced the experimental setup and data processing procedure in detail. Firstly, 49 participants took part in the formal experiment, among which 29 were MTI students and 20 were professional interpreters. All participants had at least four months of experience in interpreting through a computer screen, and there was no significant difference between the two groups' general WM capacity. The participants were grouped according to whether they had over four years' professional interpreting experience, as the one of the research aims of the present study was to investigate the effects of interpreter work experience on interpreters' note-taking behaviour. Secondly, four speech segments excerpted from a TED talk were selected and manipulated to be the source speech of the present study. Both objective and subjective methods were adopted to assess the difficulty level of interpreting these segments. Overall, the four speech segments follow an easy-difficult-difficult-easy manner in interpreting difficulty. A 45-second opera clip adapted from the original TED talk was kept between Segments 2 and 3 for removing potential spill-over effects of interpreting difficulty. At the same time, with two segments presented in a difficulty-increase direction of interpreting and the other two in the difficulty-decrease direction of interpreting, this research design allows the present study to investigate whether the effects of source speech difficulty on interpreters' note-taking behaviour in VRI are mediated by the presentation sequence of the segments. Thirdly, two pilot studies were conducted to decide the difficulty level of the speech segments and the applicability of the eye-and-pen experimental design. Based on the participants' feedback, the source speech, experimental setting and experimental procedures were finalised for the formal study.

The research adopted a between-subject and within-subject design with two groups of interpreters conducting VRI tasks at two difficulty levels. During the experiment, the participants signed the consent form, tested their WM span, practiced with the digital pen, conducted a warm-up exercise and completed the main interpreting task. The participants' eye fixations on the screen and pen movements on the tablet were registered, respectively, through a Tobii Pro Fusion-250 remote eye-tracker and a Wacom CLT-672 digital pen. Following the interpreting task, the participants scored the difficulty level of interpreting the segments in a NASA-TLX questionnaire, explained their notes in detail and completed another questionnaire concerning their demographic and interpreting-specific information. The whole experiment was recorded through the Audacity software. The whole experiment lasted, on average, 80-90 minutes.

Following Hvelplund's (2011), data points collected from P04, P15, P33 and P59 were excluded from data analysis because they failed to meet (i.e., one SD below the mean) at least two of the three criteria, MFD, GTS and GSP, with the thresholds setting at 276 ms, 51.15% and 76.58%, respectively, in this study. Moreover, the data points of P42 and P53 were discarded as they reported difficulties in using the digital pen with ease. Overall, the percentage of invalid data was 10.91%.

Lastly, all the statistical analyses were conducted using the software IBM SPSS Statistics 26. pair-tests and Mann-Whitney tests were conducted for between-subject comparisons, with the former targeting normally distributed data and the latter for non-normally distributed data. Paired *t*-tests and the Wilcoxon signed rank tests were adopted for within-subject comparisons. The former centred on normally distributed data and the latter suited non-normally distributed data. Pearson's correlation tests were mainly adopted to explore the relationships between interpreters' note-taking behaviour and CI quality. To ensure the consistency and comparability of the results, for non-normally distributed data, only when the Pearson's and Spearman's correlation tests report different results are the Spearman's test results presented.

Chapter 6: Results

Before answering the five research questions proposed in this study, it is important to examine whether the designed difficulty levels of the four speech segments are in accordance with the perceived difficulty levels for the participants. Therefore, this chapter begins by reporting the NASA-TLX scores that the participants gave to the segments after interpreting (Section 6.1). It then reports on the effects of interpreter work experience and source speech difficulty on the process (Section 6.2) and product (Section 6.3) of note-taking. Section 6.4 continues to examine the process of note-reading by visualizing the participants' reading patterns and quantifying their reading effort. Section 6.5 then focuses on the relationship between the note-taking effort and the note-reading effort; and Section 6.6 reports on the associations between the participants' note-taking behaviour and interpreting performance. It is worth mentioning that the first two segments of interpreting are designed in an easy-to-difficult direction of interpreting, while the last two segments are arranged in the reverse direction. With such sequence designs, the present study explores whether the effects of source speech difficulty on interpreters' note-taking behaviour are mediated by the sequence of segments interpreted. Therefore, except for reporting on the participants' note-taking behaviour in the easy (Segments 1 and 4) and difficult (Segments 2 and 3) segments of interpreting, the collected data are also analysed from a sequence perspective, namely from easy to difficult (between Segments 1 and 2) and from difficult to easy (between Segments 3 and 4). Hence, for research questions that concern the effects of source speech difficulty, results will be reported from two perspectives, in terms of: the difficulty level of each segment; and the difficulty-level increase/decrease across segment sequence.

6.1 Perception of task difficulty

The consistency between the objective assessment and subjective perception of interpreting difficulty serves as an important foundation for the present study to explore the potential difficulty effects on the participants' note-taking behaviour. Therefore, this section reports on the participants' ratings in the adapted NASA-TLX questionnaire, to examine how the participates rated the difficulty levels of the interpreting segments. The specific questions and ratings of the adapted NASA-TLX questionnaire (cf. Sun & Shreve, 2014) are illustrated in Table 6-1.

Table 6-1. Details about the adapted NASA-TLX questionnaire

Aspect	Question in the questionnaire	Scale explanation
Mental demand	How mentally demanding was the task?	0-10 (The higher the score, the more demanding the task was)
Effort	How hard did you have to work to accomplish your level of performance?	0-10 (The higher the score, the more effort was devoted to completing the task)
Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you?	0-10 (The higher the score, the more frustrated the task performer was)
Performance	How successful were you in accomplishing what you were asked to do?	0-10 (The higher the score, the more satisfied the task performer was about his/her performance)

The two groups of participants' NASA-TLX scores are presented according to the easy and difficult segments of the source speech in Table 6-2. After performing the normality check for the score differences in the two task conditions (Table 6-3), paired-sample *t*-tests were performed to examine whether these differences reach a significant level. The results show that both the student interpreters and the professional interpreters reported significantly greater mental demands and higher frustration levels in interpreting the difficult segments than in interpreting the easy ones (Table 6-4). However, only the professionals expended significantly more *Effort* on dealing with the increased difficulty of interpreting. The students, on the other hand, reported a similar amount of *Effort* in interpreting the easy and difficult segments. The same tendency is observed in *Performance*, where significant differences were only observed in the professional group but not in the student group.

Table 6-2. NASA-TLX ratings in easy and difficult segments of interpreting

	Students (N=29)				Professionals (N=20)			
	Easy		Difficult		Easy		Difficult	
	M	SD	M	SD	M	SD	M	SD
Mental demand	6.03	1.17	6.59	1.24	5.59	1.94	6.93	1.46
Effort	6.03	1.40	6.35	1.36	6.21	1.65	7.14	1.51
Frustration	5.41	1.76	5.95	1.73	4.78	2.11	6.18	2.16
Performance	4.95	1.52	4.70	1.64	5.99	1.71	5.40	1.60

Table 6-3. Shapiro-Wilk normal distribution test results for within-group comparisons in the NASA-TLX scores (easy vs. difficult)

	Students			Professionals		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Mental demand	.932	29	>.05	.964	20	>.05
Effort	.929	29	>.05	.929	20	>.05
Frustration	.981	29	>.05	.972	20	>.05
Performance	.937	29	>.05	.935	20	>.05

Table 6-4. Paired-sample *t*-test results for within-group comparisons in the NASA-TLX

scores (easy vs. difficult)

	Students			Professionals		
	t	*df*	Sig.	*t*	*df*	Sig.
Mental demand	-2.962	28	<.01	-4.437	19	<.001
Effort	-1.866	28	>.05	-3.305	19	<.01
Frustration	-3.103	28	<.05	-4.968	19	<.001
Performance	1.274	28	>.05	2.393	19	<.05

Between-group comparisons are conducted in each task condition to explore whether there were significant differences between the two groups of participants in task difficulty perception. Since the test of homogeneity of variances (Table 6-5) and the Shapiro-Wilk normal distribution test (Tables 6-6 and 6-7) returned some values lower than .05, non-parametric Mann-Whitney tests are performed to investigate the group differences. As presented in Table 6-8, the two groups of participants reported similar levels of *Mental demand* and *Frustration* in the two task conditions. No significant differences are found between the two groups of participants in the *Effort* of completing the interpreting of the easy segments. However, the professionals reported a significantly larger amount of *Effort* than the students in the difficult segments of interpreting, presenting both difficulty and experience effects on task difficulty perception. Meanwhile, in the *Performance* scores, the two groups present significant differences in the easy segments, with the professionals rating their performance significantly higher than the students did, but in the difficult segments, the two groups rated their interpreting performance at a similar level.

Table 6-5. Test of homogeneity of variances for between-group comparisons in the NASA-TLX scores (easy vs. difficult)

	Easy				Difficult			
	Statistic	*df1*	*df2*	Sig.	Statistic	*df1*	*df2*	Sig.
Mental demand	9.565	1	47	<.01	.330	1	47	>.05
Effort	9.358	1	47	>.05	.041	1	47	>.05
Frustration	.009	1	47	>.05	.135	1	47	>.05
Performance	.213	1	47	>.05	.204	1	47	>.05

Table 6-6. Shapiro-Wilk normal distribution test results for between-group comparisons in the NASA-TLX scores in easy segments of interpreting

	Students			Professionals		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Mental demand	.924	29	<.05	.969	25	>.05
Effort	.968	29	>.05	.880	25	<.05
Frustration	.935	29	>.05	.979	25	>.05
Performance	.972	29	>.05	.979	25	>.05

Table 6-7. Shapiro-Wilk normal distribution test results for between-group

comparisons in the NASA-TLX scores in difficult segments of interpreting

	Students			Professionals		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Mental demand	.904	29	<.05	.902	25	<.05
Effort	.949	29	>.05	.934	25	>.05
Frustration	.930	29	>.05	.924	25	>.05
Performance	.925	29	<.05	.976	25	>.05

Table 6-8. Mann-Whitney test results for between-group comparisons in the NASA-TLX scores in each task condition (easy vs. difficult)

	Easy			Difficult		
	U	*Z*	Sig.	*U*	*Z*	Sig.
Mental demand	235.00	-1.130	>.05	219.50	-1.453	>.05
Effort	285.00	-.103	>.05	161.00	-2.651	<.01
Frustration	209.50	-1.645	>.05	250.00	-.822	>.05
Performance	186.00	-2.125	<.05	220.00	-1.143	>.05

To explore whether the participants' perception of interpreting difficulty was affected by the presentation sequence of the speech segments, the participants' NASA-TLX ratings are additionally analysed in a segment-based manner. The students' NASA-TLX ratings in each segment of speech are displayed in Table 6-9 and Figure 6-1. Regardless of the presentation sequence of the segments, the students perceived more *Mental demand* and *Frustration* in interpreting the difficult segments than in interpreting the easy ones. In the meantime, they reported a similar amount of *Effort* and a similar level of *Performance* in completing all the four interpreting segments.

Table 6-9. The students' (N=29) NASA-TLX scores in each segment of interpreting

	Segment 1		Segment 2		Segment 3		Segment 4	
	Difficulty increase				*Difficulty decrease*			
	M	SD	M	SD	M	SD	M	SD
Mental demand	5.95	1.23	6.62	1.24	6.55	1.26	6.12	1.31
Effort	6.07	1.48	6.47	1.40	6.22	1.33	6.00	1.34
Frustration	5.64	1.80	6.17	1.76	5.72	1.70	5.19	1.72
Performance	4.81	1.49	4.50	1.61	4.90	1.68	5.09	1.56

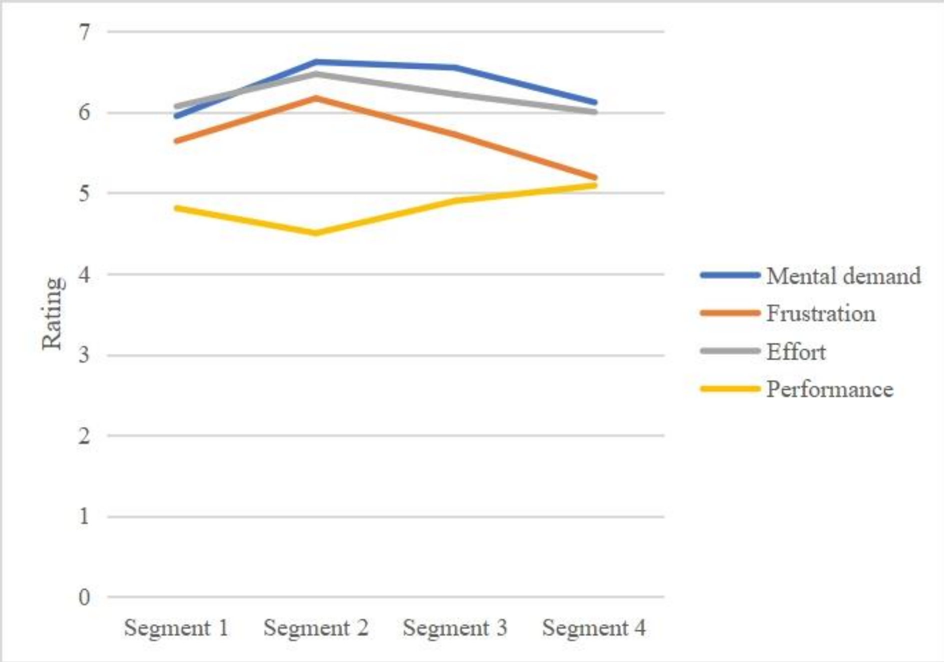


Figure 6-1. The students' NASA-TLX scores in each segment of interpreting

After checking the normality distribution of the mean difference in the NASA-TLX scores of the two pairs of speech segments (Table 6-10), Wilcoxon Signed Rank tests were conducted to investigate the significance level of the observed differences in the students' NASA-TLX ratings (Table 6-11). The results show that significant differences only existed in the difficulty-increase direction of interpreting (Segment 1 vs. Segment 2), but not in the difficulty-decrease direction of interpreting (Segment 3 vs. Segment 4). In other words, the students perceived significantly more difficulty in interpreting Segment 2 than in interpreting Segment 1, but they regarded Segment 3 and Segment 4 as similarly demanding and frustrating. Meanwhile, the reported *Effort* and *Performance* scores remain steady across the four segments.

Table 6-10. Shapiro-Wilk normal distribution test results for within-student comparisons in the NASA-TLX scores (easy-to-difficult vs. difficult-to-easy)

	Segment 1-Segment 2			**Segment 3-Segment 4**		
	Difficulty increase			*Difficulty decrease*		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Mental demand	.934	29	>.05	.943	29	>.05
Effort	.950	29	>.05	.913	29	<.05
Frustration	.924	29	<.05	.835	29	<.05
Performance	.946	29	>.05	.914	29	<.05

Table 6-11. Wilcoxon Signed Rank tests for within-group comparisons in the students' NASA-TLX scores (easy-to-difficult vs. difficult-to-easy)

	Segment 1-Segment 2		Segment 3-Segment 4	
	Difficulty increase		*Difficulty decrease*	
	Z	Sig.	*Z*	Sig.
Mental demand	-2.524	<.05	-1.534	>.05
Effort	-1.654	>.05	-1.173	>.05
Frustration	-2.013	<.05	-1.904	>.05
Performance	-1.157	>.05	-.681	>.05

The professionals' NASA-TLX ratings in each segment of interpreting are presented in Table 6-12 and visualized in Figure 6-2. As the Figure shows, the professionals reported greater difficulty in interpreting the difficult segments of speech than in interpreting the easy ones, regardless of the sequence of difficulty levels of segments. At the same time, the score difference is more obvious in the difficulty-increase direction (Segment 1 vs. Segment 2) than in the difficulty-decrease direction (Segment 3 vs. Segment 4) direction of interpreting.

Table 6-12. The professionals' (N=20) NASA-TLX scores in each segment of interpreting

	Segment 1		Segment 2		Segment 3		Segment 4	
	Difficulty increase				*Difficulty decrease*			
	M	SD	M	SD	M	SD	M	SD
Mental demand	5.53	1.73	7.28	1.33	6.58	1.53	5.65	2.15
Frustration	4.55	2.19	6.68	1.93	5.68	2.30	5.00	2.06
Effort	6.15	1.60	7.58	1.46	6.70	1.46	6.28	1.74
Performance	5.83	1.66	5.05	1.79	5.70	1.34	6.15	1.78

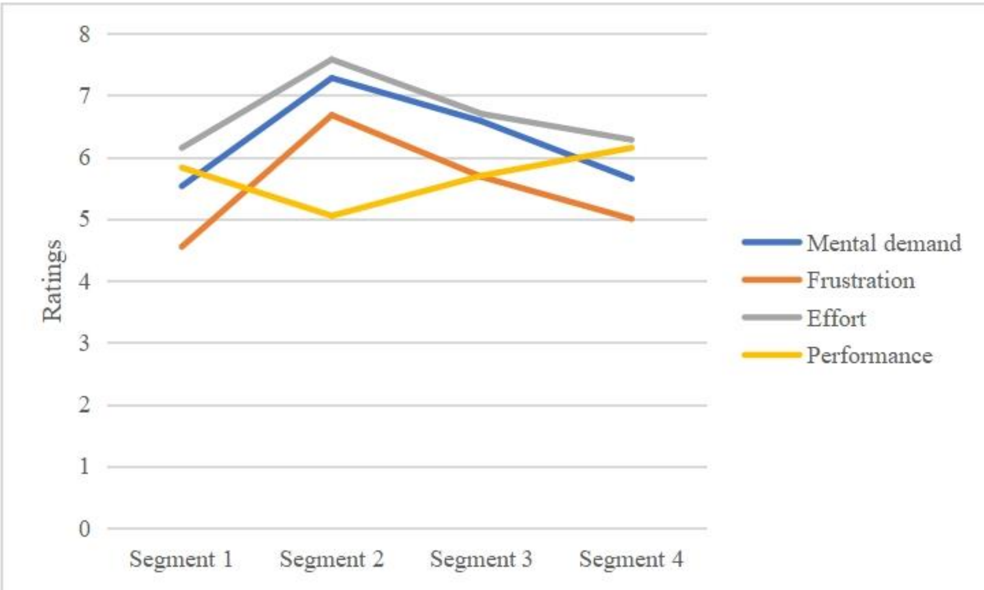


Figure 6-2. The professionals' NASA-TLX scores in each segment of interpreting

Wilcoxon Signed Rank tests (Table 6-13) are adopted to examine the significance level of the observed differences in the professionals' NASA-TLX scores after the normality check of the mean differences in the data (Table 6-14). According to the test results, all the score differences in the difficulty-increase direction of interpreting are statistically significant. By contrast, in the difficulty-decrease direction of interpreting, only the difference in *Mental demand* is marginally significant with a *p*-value at .049.

Table 6-13. Wilcoxon Signed Rank tests for within-professional comparisons in different directions of interpreting difficulty sequence (easy-to-difficult vs. difficult-to-easy)

	Segment 1-Segment 2		**Segment 3-Segment 4**	
	Difficulty increase		*Difficulty decrease*	
	Z	Sig.	*Z*	Sig.
Mental demand	-3.246	<.001	-1.969	<.05 (.049)
Effort	-3.137	<.01	-1.134	>.05
Frustration	-3.614	<.001	-1.407	>.05
Performance	-2.269	<.05	-1.345	>.05

Table 6-14. Shapiro-Wilk normal distribution test results for within-group comparisons in the professionals' NASA-TLX scores (easy-to-difficult vs. difficult-to-easy)

	Segment 1-Segment 2			**Segment 3-Segment 4**		
	Difficulty increase			*Difficulty decrease*		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Mental demand	.965	20	>.05	.916	20	>.05
Effort	.939	20	>.05	.979	20	>.05
Frustration	.965	20	>.05	.940	20	>.05
Performance	.943	20	>.05	.889	20	<.05

Group differences in the NASA-TLX scores in each segment of interpreting are examined through Mann-Whitney tests (Tables 6-15 and 6-16), as a part of the data was found to be not normally distributed[15]. Overall, the two groups of participants have similar NASA-TLX ratings, with only three (*Performance* in Segment 1 and Segment 4 as well as *Effort* in Segment 2) significant differences being observed. A further look at the data (Table 6-17) reveals three clear tendencies. Firstly, throughout the four segments of interpreting, the professionals reported greater *Effort* and better *Performance* than the students. Secondly, most of the time, the students experienced a higher level of *Frustration* than the professionals during interpreting. Thirdly, compared to the student group, the professional group also shows a wider gap in the perception of interpreting difficulty by assessing the easy segments more easily (with lower *Mental demand* scores) and the difficult segments with more difficulty (with

[15] Altogether, 32 Shapiro-Wilk normal distribution tests (2 participant groups*4 segments of interpreting*4 NASA-TLX scores) are performed.

higher *Mental demand* scores) than the students.

Table 6-15. Mann-Whitney tests for between-group comparisons in the NASA-TLX scores (easy segments)

	Segment 1			Segment 4		
	U	*Z*	Sig.	*U*	*Z*	Sig.
Mental demand	224.50	-1.163	>.05	246.50	-.899	>.05
Effort	289.00	-.210	>.05	275.00	-.311	>.05
Frustration	202.00	-1.813	>.05	266.50	-.483	>.05
Performance	185.50	-2.161	<.05	189.50	-2.078	<.05

Table 6-16. Mann-Whitney tests for between-group comparisons in the NASA-TLX scores (difficult segments)

	Segment 2			Segment 3		
	U	*Z*	Sig.	*U*	*Z*	Sig.
Mental demand	207.00	-1.749	>.05	288.00	-.042	>.05
Effort	162.50	-2.652	<.01	232.50	-1.205	>.05
Frustration	252.00	-.786	>.05	282.00	-.165	>.05
Performance	233.50	-1.170	>.05	220.00	-1.450	>.05

Table 6-17. Summarized results of the comparison between the two groups of participants' NASA-TLX scores in each segment of interpreting

	Segment 1 (easy)	**Segment 2 (difficult)**	**Segment 3 (difficult)**	**Segment 4 (easy)**
Mental demand	Stu	Pro	Pro	Stu
Effort	Pro	Pro	Pro	Pro
Frustration	Stu	Pro	Stu	Stu
Performance	Pro	Pro	Pro	Pro

Notes.
1. Stu=Student, Pro=Professional. This applies to all tables.
2. In this table, "Stu" (Student) means that the student group has a higher score than the professional group in this NASA-TLX index, and "Pro" (Professional) means that the professional group has a higher score than the student group in this NASA-TLX index.

In summary, both groups of participants perceived more difficulty in interpreting Segments 2 and 3 than in interpreting Segments 1 and 4, which is in accordance with the design of the present study. However, it is worth mentioning that, most of the time, significant differences in difficulty perception are only observed in the difficulty-increase direction of interpreting (from Segment 1 to Segment 2), but not in the difficulty-decrease direction of interpreting (from Segment 3 to Segment 4). Another discovery presented in this section concerns the experience and difficulty effects on the *Effort* of interpreting. For the student interpreters, despite the changes in the perceived difficulty of interpreting, they always reported a similar level of *Effort* across the four segments of interpreting. By contrast, the professional interpreters increased their *Effort* to interpret Segment 2 (difficult) compared to Segment 1 (easy), but such a difference in the *Effort* of interpreting is not observed between Segment 3 (difficult) and Segment

4 (easy). Taken together, these results suggest that the participants' perception of interpreting difficulty was affected by the sequence of the speech segments in terms of different difficulty levels; and such effects were more obvious in the student group than in the professional group. Therefore, for the rest of the data analysis, except for investigating the two groups' differences in different task conditions (easy vs. difficult), it is also important to examine how the two groups behaved differently in terms of the different directions in the sequence of difficulty levels of interpreting segments (easy-to-difficult vs. difficult-to-easy).

6.2 The process of note-taking

This section responds to RQ1 by focusing on the effects of interpreter work experience and source speech difficulty on the process of note-taking. It firstly looks into the participants' allocation of overt visual attention and cognitive resources on the screen, with the purpose of examining how the two groups of participants balanced video-watching and note-taking during the input phase of VRI. Then it presents the results regarding the experience and difficulty effects on the effort of note-taking, which was assessed from four aspects: overt visual attention, cognitive effort, physical effort of note-writing, and time to make note decisions. The specific research question and corresponding hypotheses, which were formulated in Chapter 1, are summarized as follows:

RQ 1: What are the impacts of interpreter work experience and source speech difficulty on interpreters' note-taking process?

> H1: Between easy and difficult speech segments, as the cognitive demand for listening comprehension increases in the difficult segments, this could limit the availability of cognitive resources for note-taking. Thus, the interpreters are expected to expend less effort on note-taking in the difficult speech segments than in the easy ones. At the same time, compared with the students, the professionals who are assumed to be less reliant on the notes during interpreting would expend less effort on note-taking, regardless of source speech difficulty.

> H2: For the speech segments that are presented in an easy-to-difficult order, the interpreters are expected to reduce their note-taking effort in the latter segment;

while for the two segments arranged in a difficult-to-easy order, they would maintain their note-taking effort at a similar level because of the potential distorted perception of task difficulty. As student interpreters could be more sensitive to the change in source speech difficulty, such sequence effects of task difficulty on the interpreters' note-taking effort are expected to be more obvious in the student group than in the professional one.

These hypotheses are tested with heat maps, gaze plots and eye-tracking measures. Specifically, Section 6.3.1 answers RQ1 from the perceptive of cognitive resource allocation on the screen, with the tested eye-tracking measures including TFD and TFC for the amount of overt visual attention, MFD for cognitive effort and VC for the shift of overt visual attention. Section 6.3.2 answers RQ1 by measuring the effort of note-taking from visual, cognitive, temporal and physical aspects, with eye-tracking and pen-recording indicators including TFD and TFC for the overt visual attention that is paid to the notes, MFD for the cognitive effort of note-taking, CC for the physical effort of note-writing, and EPS for the time to make note decisions.

6.2.1 Cognitive resource allocation

This section presents the results concerning the distribution of the participants' eye fixations on the screen. It firstly identifies the major areas that the participants looked at during the input phase of VRI, through heat maps and gaze plots. Then, it examines how the participants allocated their limited overt visual attention and cognitive resources to these different areas on the screen.

6.2.1.1 Visualization tools

Heat maps are created based on the distribution and duration of fixations in a given area. Basically, the warmer the colour is, the more overt visual attention is paid to the area. Through observing each participant's heat maps generated based on the process of note-taking in each segment of interpreting, two areas that attracted most of the participants' attention during the input phase of VRI are easily identified: the speaker window where the source speech was played, and the virtual notepad where the notes were taken. Figure 6-3 and Figure 6-4, respectively, exemplify a student interpreter's and a professional interpreter's heat map during the process of note-taking. The figures show that the participants' fixations basically matched with the area of the notes. At the same time, the participants looked frequently at the speaker window for visual information collection.

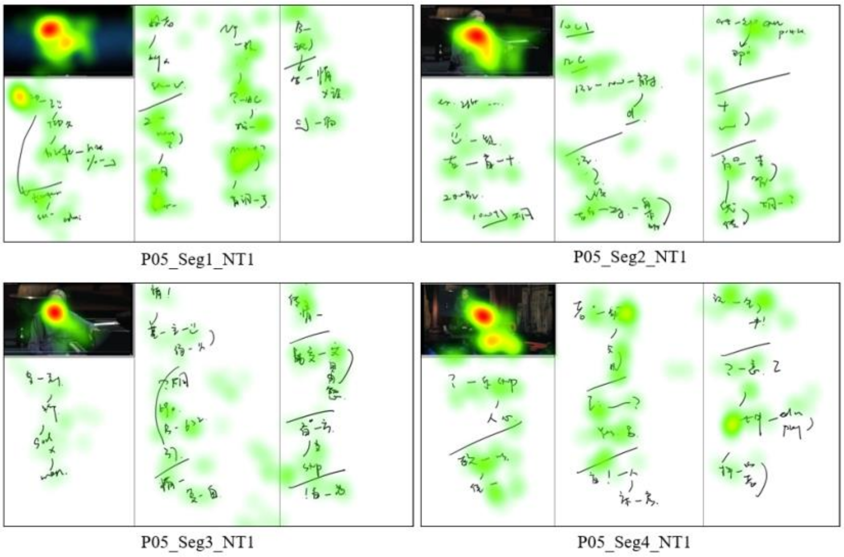


Figure 6-3. An example of the heat maps of the note-taking process of student interpreters (P05)

Note. Each heat map is named based on the number of the participant (P), the number of the speech segment (Seg), and the page number of the virtual "notepad" during note-taking (NT).

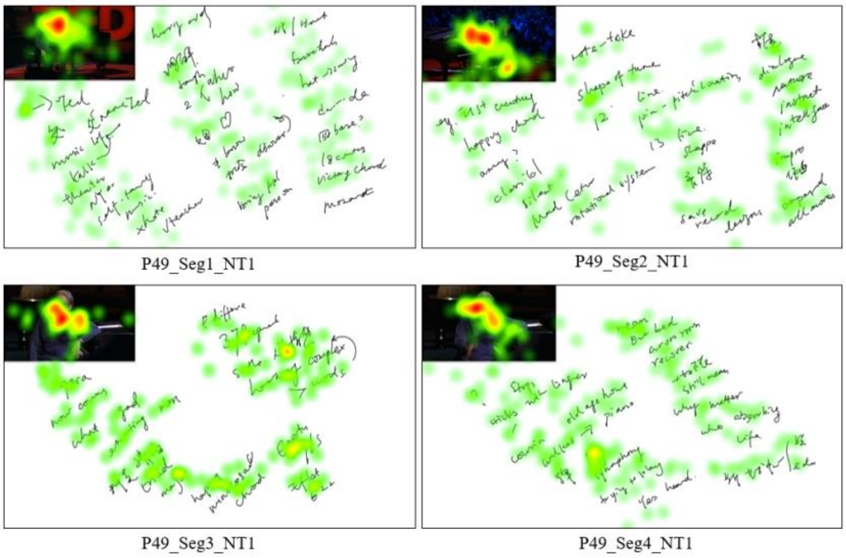


Figure 6-4. An example of the heat maps of the note-taking process of professional interpreters (P49)

Gaze plots that display the sequence of fixations on the screen can further illustrate how the participants shifted between the speaker window and the notes during the input phase of VRI. After observing all the gaze plots of the participants in each segment of interpreting, three general patterns can be summarized. Firstly, as shown in Figure 6-5, regardless of interpreter work experience and source speech difficulty, the participants frequently turned to the speaker window for visual information collection (as illustrated by the constant fixation shifts between the video and the notes). Secondly, the participants usually organized the individual notes into groups by leaving space or drawing lines among the different note groups. Regressions to notes that belong to the same note groups were frequently observed during the process of note-taking. This viewing pattern can be well illustrated with the help of the event log. As shown in Figure 6-6, the event log lists the participant number, the sequence of fixations, the temporal details about each fixation (start time, stop time and duration), the name of the AOI (named by the appearance and content of the note), and the corresponding source speech unit. For instance, for the notes presented in the blue rectangle in Figure 6-7, P49 first processed the notes in sequence till the 56^{th} fixation (as indicated by the red rectangle) where she started to look back to previous notes (as indicated by the blue rectangle). It was after the 63rd fixation that P49 continued to process the next note (which is a full word, “Ted”). Thirdly, although most fixations are found at the exact places where notes were taken, there are also fixations on the space among the notes, which could have resulted from the fact that the participants tried to establish the logical relations among the individual notes and form meaning chunks based on their understanding of the source speech. These observational findings suggest that, when analysing the participants’ allocation of overt visual attention and cognitive resources on the screen during the input phase of VRI, three AOIs can be drawn for exporting the eye-tracking data: the speaker window, the noted area (including the space between the notes), and the whole screen.

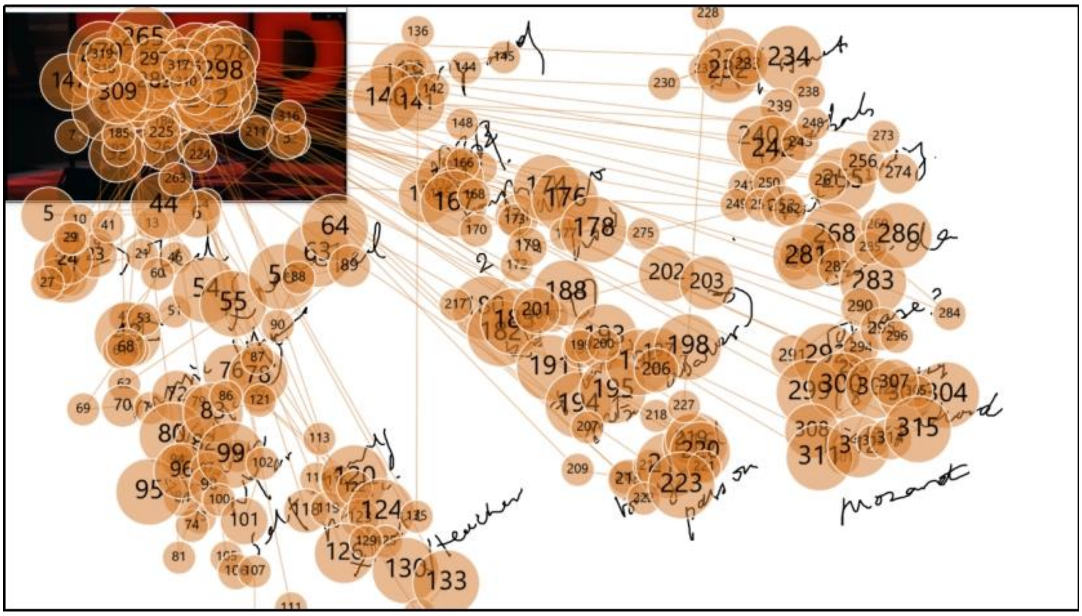


Figure 6-5. The gaze plot of P49 (professional) during the process of note-taking in Segment 1 (easy)

Participan	Fixation sequence	Start time	Stop time	Fixation Duration	AOI name (coded notes)	Identified	Source speech unit
P49	46	19396	19562	167	E%fw_en*ted^1$1ted	Ted	Ted(1st)
P49	47	19587	19754	167	E%fw_ch*笑^1$1chuckled	笑	chuckled
P49	48	19771	20362	592	E%fw_ch*笑^1$1chuckled	笑	chuckled
P49	49	20404	20529	125	Speaker window		
P49	50	20621	20804	183	Speaker window		
P49	51	21079	21212	133	Blank space		
P49	52	21237	21404	167	E%fw_ch*笑^1$1chuckled	笑	chuckled
P49	53	21421	21662	242	E%fw_ch*笑^1$1chuckled	笑	chuckled
P49	54	21679	22179	500	Blank space		
P49	55	22196	22987	792	E%ab_ch*父^1$1father's	父	father's
P49	56	23004	23854	850	E%fw_en*name^1$1name	name	name
P49	57	23896	24029	133	Speaker window		
P49	58	24046	24287	242	Speaker window		
P49	59	24296	24396	100	Speaker window		
P49	60	24612	24771	158	E%fw_en*ted^1$1ted	Ted	Ted(1st)
P49	61	24796	25071	275	E%fw_ch*笑^1$1chuckled	笑	chuckled
P49	62	25087	25221	133	Blank space		
P49	63	25262	25979	717	E%fw_en*name^1$1name	name	name
P49	64	25996	26787	792	E%fw_en*ted^1$2ted	Ted	Ted(2nd)
P49	65	26837	26946	108	Speaker window		
P49	66	26962	27146	183	Speaker window		
P49	67	27171	27412	242	Speaker window		

Figure 6-6. The event log of P49 during the process of note-taking in Segment 1

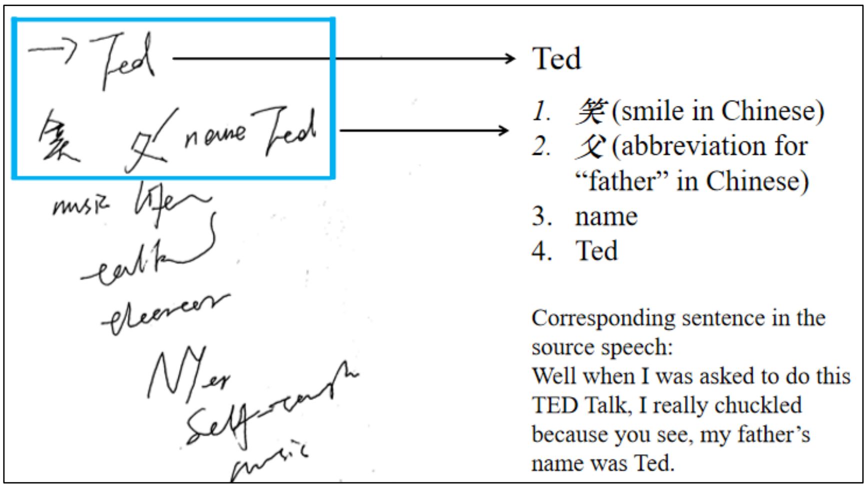


Figure 6-7. A part of the notes created by P49 in Segment 1

6.2.1.2 Eye-tracking measures

To examine how the participants distributed their limited overt visual attention and cognitive resources to video-watching and note-taking during the input phase of VRI, AOIs are respectively drawn on the whole screen, the speaker window and the noted areas (Figure 6-8). Two types of data are presented in this section: 1) the original values of the fixation counts and fixation durations on the three AOIs; and 2) the proportions of the fixations on the speaker window and the noted areas as compared to that on the whole screen.

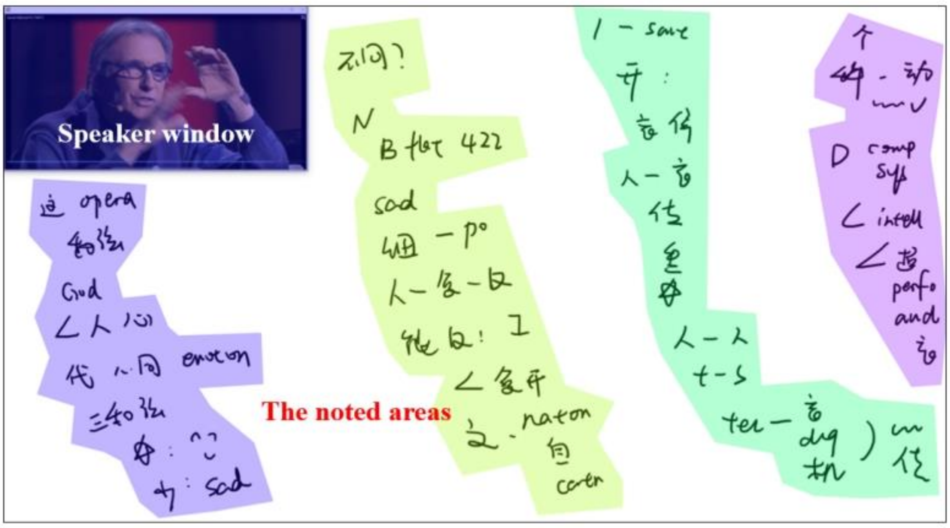


Figure 6-8. AOIs on the screen: the speaker window and the noted areas

Table 6-18 presents the distribution of the students' eye fixations on the screen in easy and difficult segments of interpreting. After checking the normal distribution of the mean difference of the students' data (Table 6-19), a series of Wilcoxon Signed Rank tests (Table 6-20) are performed to examine the effects of source speech difficulty on the students' distribution of overt visual attention and cognitive effort to the three AOIs. The results show significant differences in the overt visual attention that they paid to the whole screen and the noted area (as illustrated by TFD and TFC), but not in the cognitive effort that they expended on processing these AOIs (see MFD data). Specifically, compared with the easy segments, the TFD and TFC on the screen and the noted area are significantly lower in the difficult ones. However, when comparing the proportions to which the noted area occupied the student interpreters' overt visual attention across the easy and difficult segments of interpreting (Figure 6-9), no significant difference ($t(28)=.012$, $p>.05$) is found between the two. In other words, the students decreased the overt visual attention that they paid to the noted area in the difficult segments because of the overall decrease in their attention to the whole screen. As for the speaker window, the students paid a similar amount of overt visual attention and cognitive effort to watching the video across the two different task conditions (Table 6-18 and Table 6-20).

Table 6-18. The students' (N=29) fixation distribution on the screen during the input stage of CI (easy vs. difficult)

		Easy		**Difficult**	
		M	SD	M	SD
Screen	TFD	156143.10	16016.98	141152.55	20847.55
	TFC	366.59	55.09	341.40	59.33
	MFD	441.38	102.21	426.28	102.94
Noted area	TFD	100346.08	29898.68	91173.69	29445.08
	TFC	224.67	66.24	209.62	70.78
	MFD	461.37	116.07	447.26	110.54
Speaker	TFD	41868.97	28697.11	38895.66	23029.40
	TFC	98.97	52.33	95.41	43.65
	MFD	401.16	150.35	400.67	150.26
	VC	41.45	13.26	39.60	11.16

Note. The temporal measures are presented in milliseconds.

Table 6-19. Shapiro-Wilk normal distribution test results for within-group comparisons in the students' fixation distribution on the screen (easy vs. difficult)

		Statistic	***df***	**Sig.**
Screen	TFD	.793	29	<.001
	TFC	.860	29	<.01
	MFD	.927	29	<.05
Noted area	TFD	.988	29	>.05
	TFC	.959	29	>.05
	MFD	.881	29	<.01
Speaker	TFD	.947	29	>.05

TFC	.945	29	>.05
MFD	.939	29	>.05
VC	.957	29	>.05

Table 6-20. Wilcoxon Signed Rank tests for within-group comparisons in the students' fixation distribution on the screen (easy vs difficult)

		Z	*Sig.*
Screen	TFD	-5.447	<.001
	TFC	-3.607	<.001
	MFD	-1.715	>.05
Noted area	TFD	-2.995	<.01
	TFC	-2.437	<.05
	MFD	-1.719	>.05
Speaker	TFD	-.832	>.05
	TFC	-.378	>.05
	MFD	-.184	>.05
	VC	-.865	>.05

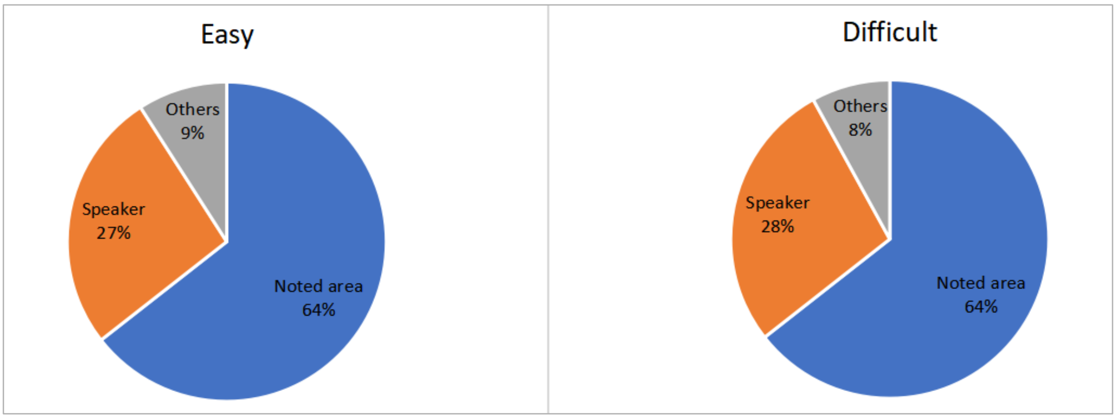


Figure 6-9. The proportions of the student interpreters' fixations on the noted area and the speaker window in easy (left) and difficult (right) segments of interpreting

The professionals' fixations to the three AOIs during the input phase of VRI are shown in Table 6-21. Overall, the professional interpreters present a similar gazing pattern as the students. Compared with the difficult segments, the professionals paid significantly more overt visual attention to the screen in the easy segments (see Table 6-22 for normal distribution test results and Table 6-23 for significant test results). One different pattern that is witnessed in the professionals compared to the students is that the former group also shows a significant difference in the MFD on the screen, which is an indicator of cognitive effort. Since there are no significant differences in the MFDs on the noted area and the speaker window (Table 6-23), the significantly decreased MFD on the screen in the difficult segments could be attributed to the participants' fixations on the blank space on the screen. The above examinations are also conducted in percentages (Figure 6-10). The results show that the professional interpreters distributed similar amounts of overt visual attention to the noted area (TFD: Z=-.672, p>.05) and the speaker window (TFD: Z=-.112, p>.05) between the two task conditions. This again indicates that the observed changes in the TFD and TFC on the noted area were caused by the overall increase or decrease of the TFD and TFC on the whole screen.

Table 6-21. The professionals' (N=20) fixation distribution on the screen during the input stage of CI

		Easy		**Difficult**	
		M	SD	M	SD
Screen	TFD	163991.10	12283.80	150305.39	13844.45
	TFC	377.58	61.15	357.93	63.38
	MFD	448.95	96.61	436.07	101.04
Noted area	TFD	117274.64	26366.47	109982.21	23145.80
	TFC	245.60	47.79	238.08	53.58
	MFD	486.28	116.82	474.93	114.98
Speaker	TFD	34426.83	19692.84	32211.56	17698.11
	TFC	93.60	47.33	92.53	42.09
	MFD	362.80	110.79	347.64	103.52
	VC	38.60	15.60	40.75	17.17

Table 6-22. Shapiro-Wilk normal distribution test results for within-group comparisons in the professionals' fixation distribution on the screen (easy vs difficult)

		Statistic	***df***	**Sig.**
Screen	TFD	.589	20	<.001
	TFC	.908	20	>.05
	MFD	.879	20	<.05
Noted area	TFD	.962	20	>.05
	TFC	.958	20	>.05
	MFD	.957	20	>.05
Speaker	TFD	.960	20	>.05
	TFC	.933	20	>.05
	MFD	.961	20	>.05
	VC	.968	20	>.05

Table 6-23. Wilcoxon Signed Rank tests for within-group comparisons in the professionals' fixation distribution on the screen (easy vs difficult)

		Z	**Sig.**
Screen	TFD	-3.920	<.001
	TFC	-2.838	<.01
	MFD	-1.979	<.05
Noted area	TFD	-1.979	<.05
	TFC	-1.307	>.05
	MFD	-1.792	>.05
Speaker	TFD	-.485	>.05
	TFC	-.299	>.05
	MFD	-.933	>.05
	VC	-1.102	>.05

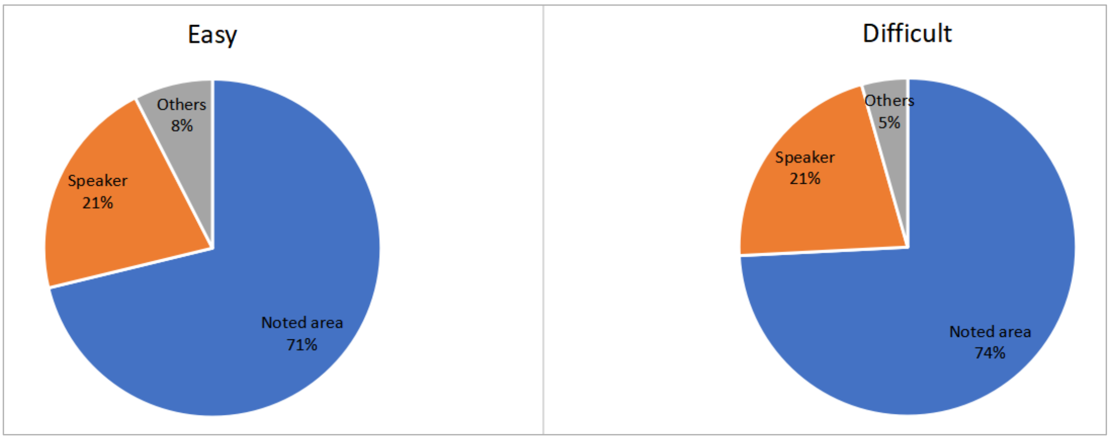


Figure 6-10. The proportions of the professionals' fixations on the noted area and the speaker window in easy (left) and difficult (right) segments of interpreting

Group comparisons are also conducted for each task condition, same results are obtained by using the original values of the eye-tracking measures and the percentages of the eye fixations on different areas on the screen, and the details are presented in Table 6-24. Although most of the group differences did not reach a significant level, it is found that, regardless of source speech difficulty, the professional group always shows higher values in TFD, TFC and MFD with AOIs on the screen and noted area than the student group. In turn, the student group always shows longer TFD, more TFC, longer MFD, and more VC in the speaker window than the professional group (except for VC in the difficult segments). In the meantime, when examining the average overt visual attention that the participants paid to the noted area in percentages across the four speech segments, a significant difference is observed in the TFD data (professionals: M=72.71% SD=14.97%, students: M=64.43%, SD=16.68%; Z=-2.205, p<.05). Overall, both groups of participants paid most of their overt visual attention and cognitive resources to processing the note area, and frequently consulted the speaker window to collect visual information for listening comprehension. This prioritization of note-taking in both groups was not affected by source speech difficulty. However, compared with the students, the professionals distributed more visual attention to the noted area during the input stage of VRI.

Table 6-24. Group comparisons of the distribution of overt visual attention to the screen, the noted area and the speaker window (Wilcoxon Signed Rank tests)

	Easy			**Difficult**		
	Student	Professional	Sig.	Student	Professional	Sig.

Screen	TFD		+	>.05		+	>.05
	TFC		+	>.05		+	>.05
	MFD		+	>.05		+	>.05
Noted area	TFD		+	<.05		+	>.05
	TFC		+	>.05		+	>.05
	MFD		+	>.05		+	>.05
Speaker	TFD	+		>.05	+		>.05
	TFC	+		>.05	+		>.05
	MFD	+		>.05	+		>.05
	VC	+		>.05		+	>.05

Notes. "+" means higher values in this group. This applies to all tables in Section 6.3.

The participants' fixation distribution on the screen is also examined from the perspective of difficulty-level sequence, namely, for easy-to-difficult and difficult-to-easy sequences. Summarized results for the student group and the professional group can be found in Table 6-25 and Table 6-29, respectively. For the student interpreters, in the difficulty-increase direction of interpreting, all the tested measures present higher values in the easy segment (Segment 1) than in the difficult one (Segment 2), with the differences in the screen and the noted area reaching a significant level. In the difficulty-decrease direction of interpreting, the students also distributed more overt visual attention to the screen and the noted area in the easy segment (Segment 4) than in the difficult one (Segment 3). However, the values for the indicators of cognitive effort (MFD) often present the opposite results. Again, no significant differences are observed in the measures concerning the speaker window. The specific values for each indicator in each segment of interpreting are presented Table 6-26 and Table 6-27.

Table 6-25. Comparisons of the students' (N=29) distribution of overt visual attention and cognitive resources on the whole screen, the noted area and the speaker window during the note-taking process

		Segment 1	**Segment 2**	**Segment 3**	**Segment 4**
		Difficulty increase		*Difficulty decrease*	
Screen	TFD	+			+
	TFC	+			+
	MFD	+		+	
Noted area	TFD	+			+
	TFC	+		+	
	MFD	+		+	
Speaker	TFD	+			+
	TFC	+			+
	MFD	+		+	
	VC	+			+

Table 6-26. The student interpreters' (N=29) overt visual attention and cognitive

resources on the screen during the input phase of VRI in Segments 1 and 2 (Wilcoxon Signed Rank tests)

		Segment 1		Segment 2		
		Difficulty increase				
		M	SD	M	SD	Sig.
Screen	TFD	163609.81	15440.05	136576.49	23308.41	<.001
	TFC	368.57	58.45	338.37	65.01	<.01
	MFD	456.80	96.21	413.38	88.83	<.01
Noted area	TFD	103712.52	28134.84	84896.23	30127.01	<.001
	TFC	216.38	65.72	197.07	69.69	<.05
	MFD	495.98	113.78	437.78	101.35	<.01
Speaker	TFD	45060.87	26517.66	40382.85	23061.75	>.05
	TFC	109.21	51.53	102.83	47.49	>.05
	MFD	399.47	127.07	397.15	137.99	>.05
	VC	44.59	14.42	41.14	12.27	>.05

Table 6-27. The student interpreters' (N=29) overt visual attention and cognitive resources on the screen during the input phase of VRI in Segments 3 and 4 (Wilcoxon Signed Rank tests)

		Segment 3		Segment 4		
		Difficulty decrease				
		M	SD	M	SD	Sig.
Screen	TFD	145972.49	23037.70	148712.73	21910.942	<.05
	TFC	345.69	61.00	366.00	65.52	<.05
	MFD	438.13	123.51	424.09	121.77	>.05
Noted area	TFD	97451.14	27981.86	96979.65	31697.69	>.05
	TFC	222.17	70.83	232.97	60.86	>.05
	MFD	456.75	120.06	426.76	109.53	<.05
Speaker	TFD	37408.48	25837.03	38677.07	31421.02	>.05
	TFC	88.00	43.91	89.72	53.65	>.05
	MFD	404.19	178.29	402.84	176.51	>.05
	VC	38.07	12.69	38.31	11.68	>.05

In addition, since the total time that the participants fixated on the screen differed from segment to segment, the comparisons concerning the overt visual attention they allocated to the noted area and the speaker window are also compared in percentages. According to Wilcoxon Signed Rank test results (Table 6-28), no significant differences are found in these regards across the two directions of difficulty-level sequence.

Table 6-28. Comparisons of the student interpreters' overt visual attention to the noted area and speaker window during the input phase of VRI (easy-to-difficult vs. difficult-to-easy)

	Segment 1-Segment 2		Segment 3-Segment 4	
	Difficulty increase		*Difficulty decrease*	
	Z	Sig.	*Z*	Sig.
TFD_noted area%	-.768	>.05	-.545	>.05
TFD_speaker%	-1.546	>.05	-.134	>.05

The professional interpreters present the same tendencies as the student interpreters in that nearly all the measures in Segment 1 show larger values than those in Segment 2. When it comes to Segments 3 and 4, again, higher values are more frequently observed in the easy segment of interpreting than in the difficult one. Specific data about the professional group's fixations on the screen, the noted area and the speaker window can be found in Table 6-30 and Table 6-31.

Table 6-29. The professional interpreters' (N=20) distribution of overt visual attention and cognitive resources on the screen during the input phase of VRI

		Segment 1	**Segment 2**	**Segment 3**	**Segment 4**
		Difficulty increase		*Difficulty decrease*	
Screen	TFD	+			+
	TFC	+			+
	MFD	+			+
Noted area	TFD	+		+	
	TFC	+		+	
	MFD	+			+
Speaker	TFD	+			+
	TFC		+		+
	MFD	+			+
	VC	+		+	

Table 6-30. The professional interpreters' (N=20) overt visual attention and cognitive resources on the screen during the input phase of VRI in Segments 1 and 2 (Wilcoxon Signed Rank tests)

		Segment 1		**Segment 2**		
		Difficulty increase				
		M	SD	M	SD	Sig,
Screen	TFD	167864.06	13448.87	146574.66	11861.69	<.001
	TFC	387.45	62.99	358.30	60.61	<.01
	MFD	448.17	101.18	426.06	113.06	<.05
Noted area	TFD	117523.37	27658.15	100533.29	24731.04	<.01
	TFC	244.15	51.80	218.85	53.20	<.05
	MFD	491.33	127.64	471.68	124.56	>.05
Speaker	TFD	38138.48	22239.19	35808.93	20350.68	>.05
	TFC	105.55	56.08	106.75	51.71	>.05
	MFD	362.16	111.45	333.16	93.00	>.05
	VC	43.30	16.61	41.60	16.42	>.05

Table 6-31. The professional interpreters' (N=20) overt visual attention and cognitive resources on the screen during the input phase of VRI in Segments 3 and 4 (Wilcoxon Signed Rank tests)

		Segment 3		**Segment 4**		
		Difficulty decrease				
		M	SD	M	SD	Sig,
Screen	TFD	154232.20	21304.77	160823.90	12323.31	>.05
	TFC	355.53	74022.05	366.63	65.51	>.05
	MFD	449.60	97.88	453.50	97.41	>.05
Noted area	TFD	119431.13	17322.42	117025.92	25726.77	>.05

	TFC	257.30	47.80	247.05	44.72	>.05
	MFD	478.17	107.71	481.22	108.01	>.05
Speaker	TFD	28614.20	16426.60	31048.45	20907.14	>.05
	TFC	78.30	35.79	82.21	42.39	>.05
	MFD	362.12	124.28	364.49	129.86	>.05
	VC	39.90	18.60	33.90	15.81	<.05

When examining the proportions of overt visual attention that the professional interpreters paid to the noted area and the speaker window in the different directions of difficulty-level sequence, no significant differences are found across the segment sequences (Table 6-32).

Table 6-32. Within-group comparisons of the professionals' overt visual attention to the noted area and speaker window (easy-to-difficult vs. difficult-to-easy) (Wilcoxon Signed Rank tests)

	Segment 1-Segment 2		**Segment 3-Segment 4**	
	Difficulty increase		*Difficulty decrease*	
	Z	Sig.	*Z*	Sig.
TFD_noted area%	-1.408	>.05	-.411	>.05
TFD_speaker%	-.443	>.05	-.681	>.05

Between-group comparisons are also conducted in the percentages of fixation durations (Table 6-33). Although no significant group difference is detected, the professional group generally allocated more overt visual attention and cognitive effort to view the screen and take their notes than the student group. By contrast, the student group is frequently found to have had longer and more fixations on the speaker window than the professional group. In other words, compared with the students, the professionals attached more importance to note-taking, whereas in comparison with the professionals, the students paid more attention to video-watching.

Table 6-33. Group comparisons in each segment of interpreting in terms of the allocation of cognitive resource allocations on the screen during the input stage of VRI (Wilcoxon Signed Rank tests)

		Segment 1	**Segment 2**	**Segment 3**	**Segment 4**
Screen	TFD	Pro	Pro	Pro	Pro*
	TFC	Pro	Pro	Pro	Pro
	MFD	Stu	Pro	Pro	Pro
Noted area	TFD	Pro	Pro	Pro**	Pro*
	TFC	Pro	Pro	Pro	Pro
	MFD	Stu	Pro*	Pro	Pro
Speaker	TFD	Stu	Stu	Stu	Stu
	TFC	Stu	Pro	Stu	Stu
	MFD	Stu	Stu*	Stu	Stu
	VC	Stu	Pro	Pro	Stu

Notes. * p<.05, ** p<.01, and *** p<.001. This applies to all tables.

6.2.2 Note-taking effort

For RQ1, the hypotheses formulated in Chapter 1 can also be examined from the perspective of note-taking effort. As identified in Chapter 3, Section 3.2.1, the effort of note-taking can be investigated from four aspects: overt visual attention, cognitive effort, physical effort, and time to make note decisions. Therefore, the eye-tracking and pen-recording measures concerning these four aspects of note-taking process are examined in this section to explore the effects of interpreter work experience and source speech difficulty on interpreters' note-taking effort in VRI.

Measures for investigating these hypotheses include TFD (overt visual attention), MFD (cognitive effort), CC (physical effort) and EPS (time to make note decisions). It is worth mentioning that the TFD and CC data are presented in two formats: TFD_sum and CC_sum indicating the total of TFD and CC values obtained from all the notes in one segment of interpreting; and TFD_mean and CC_mean representing the average of all collected values in one segment. This is because participants who wrote more notes usually had a larger area of notes than those who noted less, resulting in bigger sums of TFD and CC. TFD_sum and CC_sum are used to indicate the total overt visual attention and physical effort of note-taking. By contrast, TFD_mean, CC_mean, MFD, VC and EPS, which were not affected by the area of notes because of their 'mean' attribute, are adopted to indicate the average effort of taking a note from visual, cognitive, physical and temporal aspects.

In addition, it should be noted that the eye-tracking data reported in this section are exported based on different AOIs from those in Section 6.2.1. In Section 6.2.1, the eye-tracking data were obtained by drawing AOIs on the whole noted area with spaces among the notes being included. In this section, instead, the eye-tracking data are exported by drawing AOIs on the individual notes (Figure 6-11), where note-writing definitely happened. In other words, compared with the data reported in Section 6.2.1, this section focuses on the effort of note-taking where handwriting control was indispensable, something that could lead to dramatic increases in the cognitive effort of note-taking and decide the quality of note-taking (see Chapter 3, Section 3.2.2).

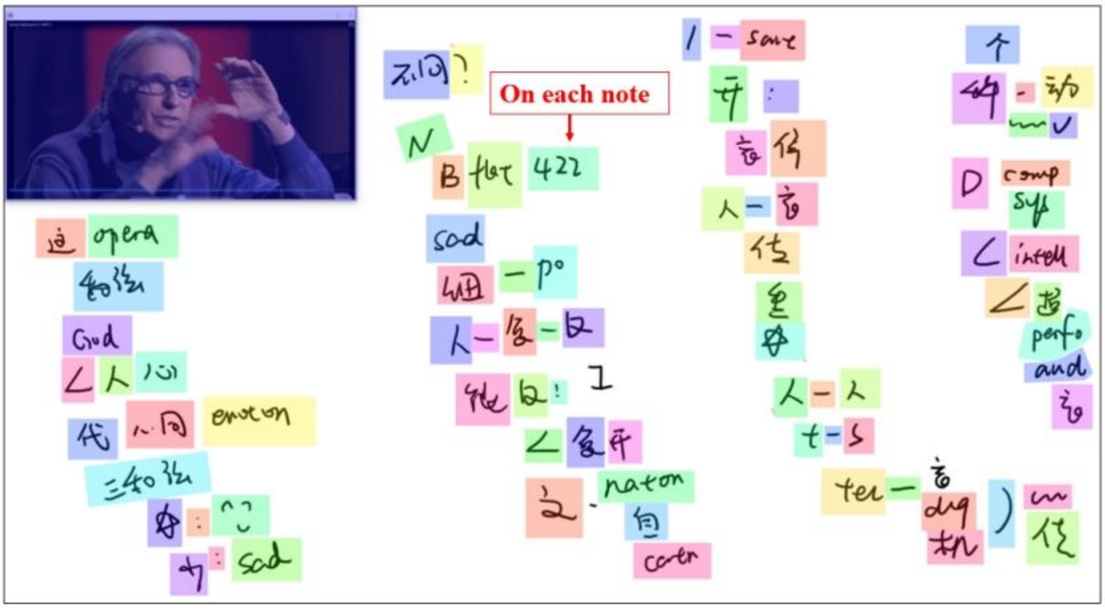


Figure 6-11. An example of AOIs drawn on the individual notes

As shown in Table 6-34, the two groups of participants present the same tendency in distributing their effort of note-taking to the easy and difficult segments of interpreting. All the participants present higher values in the sum of TFD and CC in the easy segments than in the difficult ones. In addition, both groups show longer TFD_mean, more CC_mean, shorter MFD and longer EPS in the difficult segments than in the easy ones. After normality tests (Table 6-35), non-parametric Wilcoxon Signed Rank tests are conducted for examining the students' data, and paired-sample *t*-tests are performed for testing the professionals' data. Basically, no significant differences are found in the data of the student group (Table 6-36). As for the professionals, no significant differences are observed in the indicators of cognitive effort, but significant differences are found in those of the overt visual attention, physical effort and time to make note decisions (Table 6-36).

Table 6-34. The students' and professionals' note-taking effort in easy and difficult segments of interpreting

	Students (N=29)				**Professionals (N=20)**			
	Easy		Difficult		Easy		Difficult	
	M	SD	M	SD	M	SD	M	SD
TFD_sum	74001.42	25269.08	69257.36	25051.37	89709.29	20515.09	82075.88	18605.26
TFD_mean	900.38	356.61	902.55	489.70	993.82	395.96	996.22	363.99
CC_sum	215.31	73.55	206.09	67.30	270.25	71.47	257.86	63.91
CC_mean	2.49	.60	2.53	.65	2.83	.82	3.04	.87
MFD	517.07	124.04	502.38	126.87	543.65	141.29	529.26	133.07
EPS	3993.50	901.30	4880.11	932.17	4168.75	956.62	4826.55	1168.84

Table 6-35. Shapiro-Wilk normal distribution test results for within-group comparisons in the note-taking effort (easy vs. difficult)

	Student			**Professional**		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
TFD_sum	.931	29	>.05	.982	20	>.05
TFD_mean	.632	29	<.05	.956	20	>.05
CC_sum	.990	29	>.05	.925	20	>.05
CC_mean	.932	29	>.05	.910	20	>.05
MFD	.970	29	>.05	.928	20	>.05
EPS	.985	29	>.05	.981	20	>.05

Table 6-36. Summarized results of within-group comparisons in the note-taking effort (easy vs. difficult)

	Student			**Professional**		
	Easy	Difficult	Sig.	Easy	Difficult	Sig.
TFD_sum	+		>.05	+		<.05
TFD_mean		+	<.05		+	>.05
CC_sum	+		>.05 (.56)	+		>.05
CC_mean		+	>.05		+	<.001
MFD	+		>.05	+		>.05
EPS		+	<.01		+	=.05

In terms of the group differences in the note-taking effort, almost all the values are higher in the professional group than in the student one. These between-group differences are also examined through Mann-Whitney tests (Table 6-37) after data normality checks[16]. The results show that the professionals distributed significantly more physical effort (CC) to note-writing, regardless of source speech difficulty. Except for that, no other measures show significant group differences.

Table 6-37. Between-group comparisons of the note-taking effort in easy and difficult segments of interpreting

	Easy			**Difficult**		
	Student	Professional	Sig.	Student	Professional	Sig.
TFD_sum		+	>.05		+	>.05
TFD_mean		+	>.05		+	>.05
CC_sum		+	<.05		+	<.01
CC_mean		+	<.05		+	<.05
MFD		+	>.05		+	>.05
EPS		+	>.05	+		>.05

Note. "+" means the value is higher in this group.

In addition, the sequence effects of source speech difficulty on the participants' note-taking effort are also investigated. As shown in Table 6-38 and Table 6-40, both groups of participants show significantly higher values in the sum of TFD and CC in Segment 1 (easy) than Segment 2 (difficult), indicating that the participants distributed

[16] In total, 2 participant groups*2 task conditions*6 indicators=24 normality tests are conducted.

significantly more effort to take notes for the first segment than for the second. However, such significant differences disappear when they came to Segment 3 (difficult) and Segment 4 (easy) (see Table 6-39 and Table 6-41), indicating sequence effects of source speech difficulty level on the total overt visual attention and physical effort that the participants devoted to note-taking during VRI.

Table 6-38. Wilcoxon Signed Rank test results of the students' (N=29) note-taking effort in Segment 1 and Segment 2 (easy-to-difficult)

	Segment 1		**Segment 2**		
	M	SD	M	SD	Sig.
TFD_sum	76698.53	24769.61	64227.87	23173.31	<.001
TFD_mean	979.08	401.86	886.70	434.87	<.01
CC_sum	206.34	65.94	191.97	67.48	<.05
CC_mean	2.52	.66	2.52	.75	>.05
MFD	556.46	133.23	488.83	128.48	<.01
EPS	3217.63	1577.74	4908.51	1120.93	<.001

Table 6-39. Wilcoxon Signed Rank test results of the students' (N=29) note-taking effort in Segment 3 and Segment 4 (difficult-to-easy)

	Segment 3		**Segment 4**		
	M	SD	M	SD	Sig.
TFD_sum	74286.85	30628.30	71304.32	28493.13	>.05
TFD_mean	918.40	574.64	821.00	354.27	>.05
CC_sum	220.21	70.04	224.28	84.42	>.05
CC_mean	2.54	.58	2.47	.59	>.05
MFD	515.93	148.32	477.68	134.26	<.05
EPS	4851.71	1047.68	4771.36	975.47	>.05

Table 6-40. Wilcoxon Signed Rank test results of the professionals' (N=20) note-taking effort in Segment 1 and Segment 2 (easy-to-difficult)

	Segment 1		**Segment 2**		
	M	SD	M	SD	Sig.
TFD_sum	90658.03	22172.15	76171.78	20955.29	<.01
TFD_mean	1045.39	434.56	962.33	395.94	>.05
CC_sum	260.35	62.78	243.50	63.99	<.01
CC_mean	2.85	.82	2.97	.79	>.05
MFD	556.77	159.11	531.72	149.68	>.05
EPS	3648.80	1754.32	5201.88	1752.89	<.05

Table 6-41. Wilcoxon Signed Rank test results of the professionals' (N=20) note-taking effort in Segment 3 and Segment 4 (difficult-to-easy)

	Segment 3		**Segment 4**		
	M	SD	M	SD	Sig.
TFD_sum	87979.99	19346.80	88760.56	24056.73	>.05
TFD_mean	1030.11	357.21	942.25	377.33	<.05
CC_sum	272.25	69.78	280.15	85.07	>.05
CC_mean	3.10	1.00	2.82	.84	<.01
MFD	526.80	121.43	530.53	132.73	>.05

EPS	4451.22	1135.02	4688.70	1054.23	>.05

In terms of the average effort of taking one note, the students are found to have significantly longer TFD (overt visual attention), longer MFD (cognitive effort) and shorter EPS (time to make note decisions) in Segment 1 than in Segment 2. Between Segment 3 and Segment 4, almost all the values are higher in the former segment than in the latter, with only the difference in the MFD reaching a significant level. Overall, compared with the ending segment in each pair of interpreting segments, the students distributed more effort to taking notes for the beginning segment. By comparison, between Segments 1 and 2, the professional interpreters show only one significant difference in the EPS data. Between Segments 3 and 4, they paid significantly more overt visual attention and used significantly more physical effort to complete the production of each note in the former than in the latter segment. In both directions of the sequence of source speech difficulty levels, no significant difference is found in the cognitive effort (MFD) of note-taking.

Between-group comparisons are conducted through Mann-Whitney tests in a segment-based manner. The results show that, in almost all the tested measures, the professional group presents higher values than the student group (Table 6-42). None of the group differences concerning MFD and EPS are not significant, but those related to TFD and CC show many significant results. Specifically, the professionals distributed significantly more overt visual attention and physical effort to completing the overall note-taking task and the production of each note than the students. This result suggests that professional-novice differences in the process of note-taking do not lie in the cognitive effort of note-taking but in the visual and physical input during note production.

Table 6-42. Mann-Whitney test results for between-group comparisons in the note-taking effort in each segment of interpreting[17]

	Segment 1	Segment 2	Segment 3	Segment 4
TFD_sum	Pro**	Stu	Pro*	Pro*
TFD_mean	Pro***	Pro**	Pro**	Pro*
CC_sum	Pro	Pro	Pro	Pro
CC_mean	Pro**	Pro*	Pro**	Pro
MFD	Pro	Pro	Pro	Pro
EPS	Pro	Pro	Stu	Stu

[17] Altogether, 2 groups*4 segments*6 indicators=48 data normality tests were conducted. To ensure the consistency of the results, only non-parametric tests are conducted and presented in the table.

6.2.3 Summary of RQ1

To answer the first research question, What are the impacts of interpreter work experience and source speech difficulty on interpreters' note-taking process?, Section 6.2 has examined the data collected from the process of note-taking from two perspectives: the allocation of cognitive resources on the screen, and the effort of note-taking in various aspects. Specifically, eye-tracking data concerning three AOIs, the whole screen, the noted area and the speaker window, were analysed based on different participant types and task conditions. The effort of note-taking was examined from four aspects: overt visual attention (TFD and TFC), cognitive effort (MFD), physical effort (CC) and time to make note decisions (EPS). Summarized results are presented in Table 6-43 and Table 6-44.

Table 6-43. Summarized results for RQ1 from a cognitive resource allocation perspective

Independent variables		Data type	Result
Source speech difficulty	Easy vs difficult	Original values	Easy>Difficult
		Proportions	Easy≈Difficult
	Difficulty increase vs difficulty decrease	Original values	Segment 1>Segment 2 Segment 3≈Segment 4
		Proportions	Segment 1≈Segment 2 Segment 3≈Segment 4
Interpreter work experience	Professional vs. student	Original values	Professional>Student
		Proportions	Professional>Student

Note. Results are summarized based on the eye-tracking data collected from the noted area.

In accordance with H1, in the difficult segments of interpreting, both groups of participants decreased the total overt visual attention they paid to the noted area. However, the significant differences disappeared when the comparisons were conducted based on the proportions of the fixations on the noted area, indicating that the way the participants distributed their overt visual attention to different areas on the screen remained unchanged across the two task conditions. In addition, regardless of source speech difficulty, the professional allocated a greater amount of overt visual attention and cognitive effort to process the noted area than the students. The students, on the contrary, always distributed more overt visual attention and cognitive resources to watch the speaker window than the professionals. These results suggest different patterns of attention distribution between the two groups of participants, with the professionals giving higher priority to note-taking and the students placing video-watching in a more important position. Finally, the results showed that neither source speech difficulty nor interpreter work experience was influential on the cognitive effort that the participants allocated to the noted area.

As for the sequence effects of source speech difficulty on the participants' cognitive resource allocation on the screen, both groups of participants showed many more significant differences in the two segments in the difficulty-level increase direction (Segment 1 and Segment 2), than in the difficulty-level decrease direction (Segment 3 and Segment 4). In other words, when they interpreted a difficult segment before an easy one, they would not change the effort they devote to note-taking and video-viewing. The significant differences in the overt visual attention to the first two segments disappeared when the comparisons were conducted in percentages. This again indicates that the observed decrease in the overt visual attention paid to the noted area in Segment 2 was caused by the decreased attention to the whole screen.

The hypotheses for RQ1 were also examined by measuring the effort of note-taking in different interpreter groups and task conditions. Summarized results are presented in Table 6-44. At first, the professionals always distributed more effort to complete note-taking than the students, but it was only in the measures of overt visual attention and physical effort of note-writing that there the group differences at a significant level. Then, a series of similarities were observed between the student interpreters and the professional interpreters in three respects: firstly, both groups of participants distributed more overt visual attention and physical effort to completing the overall note-taking task in the easy segments than in the difficult ones; secondly, when it comes to the average effort of taking one note, this consumed the participants more overt visual attention, physical effort and time for note-planning while taking notes for the difficult segments of interpreting than for the easy ones; and thirdly, the cognitive effort of note-taking was found to be higher in the easy segments of interpreting, although the differences were not significant.

In terms of the sequence effects of source speech difficulty level on the participants' note-taking effort, in the difficulty-level increase direction, the total and average effort of note-taking was generally greater in Segment 1 than in Segment 2. However, when it comes to Segment 3 than in Segment 4 in the difficulty-level decrease direction, only few significant differences are witnessed in the measures of the note-taking effort. Overall, the differences between Segments 1 and 2 were more obvious than those between Segments 3 and 4, suggesting a sequence effect of source speech difficulty level on the effort of note-taking. Thus, H2 was only partially supported by the results.

Table 6-44. Summarized results for RQ1 from a note-taking effort perspective

	Total note-taking effort	**Average note-taking effort**
Experience effects	Professional>student	Professional>student
Difficulty effects	Easy>difficult	Difficult>easy
Sequence effects	Both groups: Segment1>Segment2 Segment3≈Segment4	Students: Segment1>Segment2; Segment3≈Segment4
		Professionals: Segment1>Segment2; Segment3≈Segment4

6.3 The product of note-taking

The second research question of the present study targets the effects of interpreter work experience and source speech difficulty on the product of note-taking, which can be assessed from two aspects: the descriptive features of note choices, and the adoption of note-taking strategies. As introduced in Chapter 4, Section 4.3.1, the descriptive features of notes include three parts: note quantity, note form (language notes vs. symbols, full words vs. abbreviations) and note language (the SL vs. the TL). Furthermore, there are three note-taking strategies in the present study: ellipsis, replacement and restructuring. As mentioned in Chapter 1, RQ2 and the relevant hypotheses can be summarized as follows:

RQ2: What are the impacts of interpreter work experience and source speech difficulty on interpreters' note-taking product?

> H3: As source speech difficulty increases and interpreter work experience decreases, the participants would deprioritize note-taking to make more processing capacity available for listening comprehension. As a result, note quantity would decrease and the interpreters would opt for note forms, note languages and note-taking strategies that require less note-taking effort.

> H4: In the easy-to-difficult direction of interpreting segment sequence, the interpreters would opt for more effort-saving note choices in the difficult segment than in the easy one. In the reverse interpreting direction, the interpreters could maintain their note preferences. The students are expected to show more obvious changes in their choices of notes in the easy-to-difficult direction of interpreting difficulty level as compared with the professionals.

Since the participants' note quantities varied, the use of note forms, note languages and note-taking strategies are analysed in percentages. Moreover, as there were notes that the participants could not recognize during the experiment, the sum of the percentages is not always 100%. In addition, numbers are not regarded as being an important research focus in the present study, thus results concerning numbers are not be discussed in detail.

6.3.1 Choices of notes

The descriptive patterns of the students' (Table 6-45) and the professionals' (Table 6-46) notes present many similarities in easy and difficult segments of interpreting. Paired-sample *t*-tests are conducted to compare the participants' note choices in the two task conditions after normal distribution checks (Table 6-47). Overall, an increase in source speech difficulty led to a significant decrease in the note quantity, but didn't trigger significant differences in the participants' preferences for the different note forms and note languages. As summarized in Table 6-48, in both task conditions, the student and professional interpreters preferred language notes over symbols, and English notes over Chinese notes. Despite this, for both groups of participants, the percentages of language notes and full words are significantly higher in the easy segments than in the difficult ones. For the student interpreters, the proportion of English notes is significantly greater in the easy segments than in the difficult ones, whereas the proportion of abbreviation shows the opposite tendency. These results indicate that source speech difficulty did not lead to changes in note preferences, but that it did affect the frequency with which the participants used these preferred note forms and note languages. One clear difference between the two groups of participants is their preferences for full words and abbreviations. Specifically, compared with the students, the professionals took more of their notes in full words and less of their notes in abbreviations. This difference was significant for the use of full words in both task conditions (easy segments: Z=-2.136, p<.05, difficult segments: Z=-3.133, p<.01) and for abbreviations in the difficult segments (Z=-2.482, p<.05).

Table 6-45. The distribution of note quantities, note forms and note languages in the student interpreters' (N=29) notes

		Easy		**Difficult**		
Aspect	Category	M	SD	M	SD	Sig.
Note quantity	Number of notes	88.50	28.63	84.24	28.08	<.05
Note form	Language	54.16	10.89	51.74	11.33	<.01
	Symbol	36.79	10.14	34.01	11.41	<.05

	Number	2.29	.97	5.72	1.70	<.05
	Full word	24.54	12.96	18.81	11.37	<.001
	Abbreviation	29.62	7.63	32.93	9.36	<.001
Note language	Chinese	20.63	13.59	22.53	13.09	>.05
	English	33.53	10.18	29.21	10.56	<.01

Table 6-46. Distribution of note quantities, note forms and note languages in the professional interpreters' (N=20) notes

		Easy		**Difficult**		
Aspect	Category	M	SD	M	SD	Sig.
Note quantity	Number of notes	99.18	28.44	88.05	23.65	<.01
Note form	Language	61.60	14.06	57.85	12.20	<.01
	Symbol	30.70	12.63	29.08	11.19	>.05
	Number	1.82	.71	6.10	6.10	<.001
	Full word	35.35	19.52	31.86	17.27	<.01
	Abbreviation	26.25	11.99	26.30	10.32	>.05
Note language	Chinese	23.61	11.48	22.33	10.86	>.05
	English	37.99	17.78	35.52	17.83	>.05

Table 6-47. Shapiro-Wilk normal distribution test results for within-group comparisons in easy and difficult task conditions

		Student			**Professional**		
Aspect	Category	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Note quantity	Number of notes	.989	29	>.05	.930	20	>.05
Note form	Language	.927	29	>.05	.944	20	>.05
	Symbol	.957	29	>.05	.976	20	>.05
	Number	.940	29	>.05	.951	20	>.05
	Full word	.984	29	>.05	.968	20	>.05
	Abbreviation	.968	29	>.05	.934	20	>.05
Note language	Chinese	.971	29	>.05	.982	20	>.05
	English	.975	29	>.05	.944	20	>.05

Table 6-48. The descriptive patterns of the two groups of participants' notes in easy and difficult task conditions

		Note form	**Note language**
Student	Language>symbol	Abbreviation>full words	English>Chinese
Professional	Language>symbol	Full word>abbreviation	English>Chinese

Sequence effects of source speech difficulty levels are not frequently observed in the descriptive features of the participants' notes. Firstly, in terms of note quantity, both student interpreters and professional interpreters noted more in the easy segment than in the difficult one in each sequence of difficulty levels of the interpreting segments (Figure 6-12). It was in Segment 2 in the difficulty-level increase direction and Segment 4 in the difficulty-level decrease direction that the participants noted, respectively, the least and the most. As for group differences, the professional group always presented greater note quantities than the student group, but these differences don't reach a significant level. At the same time, compared with the difficult segments, the two groups' difference in note quantity is more obvious in the easy segments of interpreting.

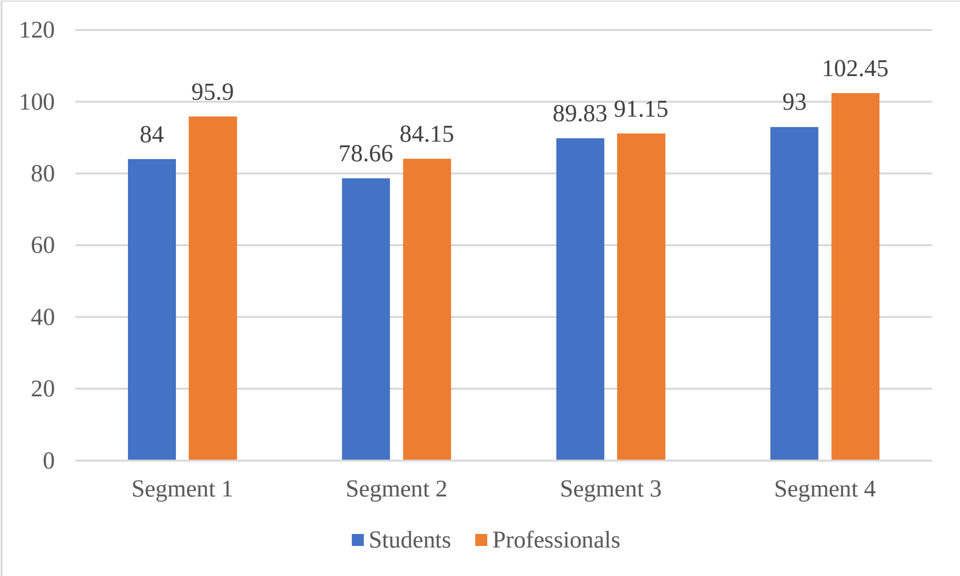


Figure 6-12. The students' and professionals' note quantities in each segment of interpreting

Secondly, the sequence of the interpreting segments did not affect the participants' preference for language notes and symbols, full words and abbreviations, or Chinese and English notes (see Figure 6-13 and Figure 6-14). However, compared with the first two segments of the interpreting, the second two segments of interpreting show more obvious gaps in the participants' preferences for different note forms and note languages, which indirectly suggests a sequence effect of source speech difficulty level on the participants' note-taking product.

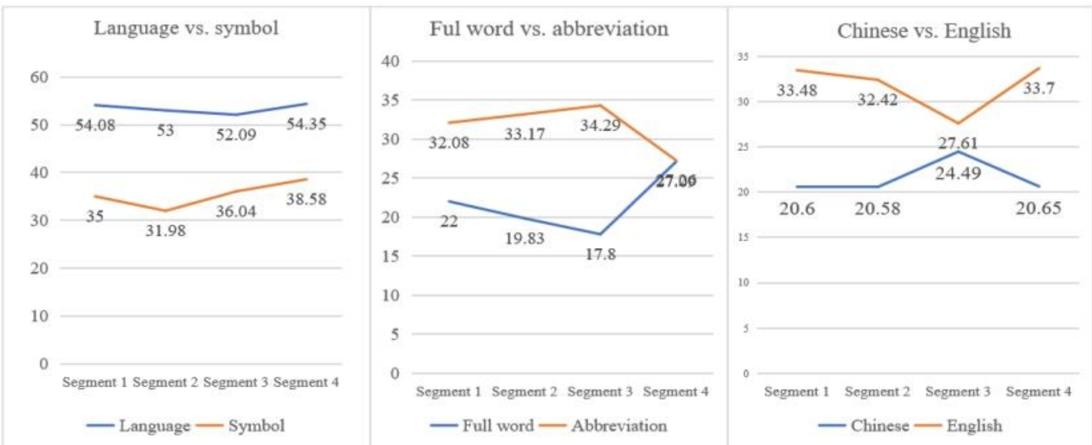


Figure 6-13. The proportions of different note categories in the students' (N=29) notes

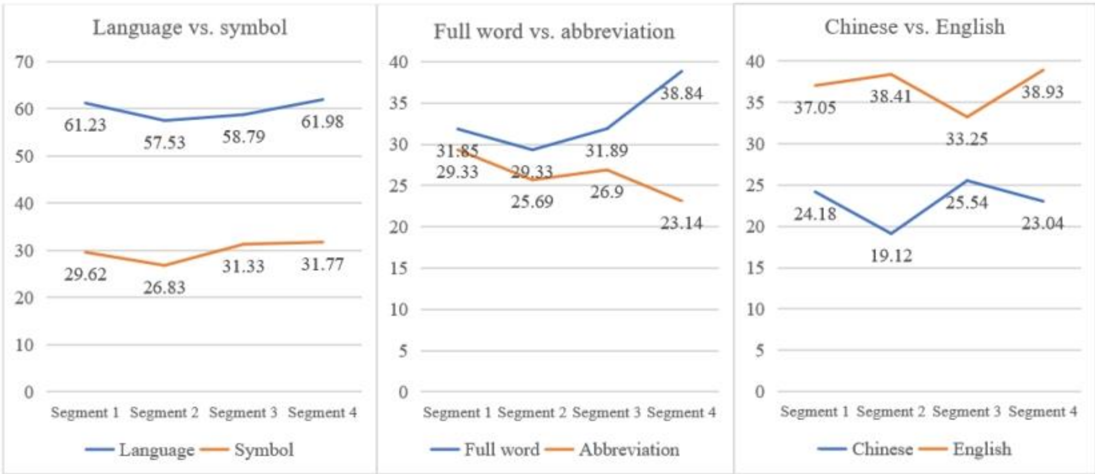


Figure 6-14. The proportions of different note categories in the professionals' (N=20) notes

Thirdly, the sequence of difficulty levels of the source speech segments did not affect the two groups' difference in the adoption of full words and abbreviations. In all the segments, the professionals used significantly more full words to produce notes than the students (Table 6-49). In turn, in every segment of interpreting, the students show a higher proportion of notes in abbreviations than the professionals. Except for full words and abbreviations, no other significant differences are observed between the two groups' use of different note forms and note languages[18].

Table 6-49. Group comparisons in terms of the use of full words and abbreviations in notes (Mann-Whitney tests)

	Full word	**Abbreviation**
Segment 1	Professional*	Student
Segment 2	Professional**	Student*
Segment 3	Professional**	Student*
Segment 4	Professional*	Student*

Note. "Professional" represents that the professional group noted more in the specific category of notes than the student group, and vice versa.

6.3.2 Adoption of note-taking strategies

In terms of note-taking strategies, both the students' (see Figure 6-15 and Table 6-50) and the professionals' (see Figure 6-16 and Table 6-51) notes were dominated by the ellipsis strategy. Although no significant differences are observed in the adoption of these note-taking strategies across different task conditions or segments of interpreting,

[18] In total, 2 groups*4 segments*8 note categories=64 data normality tests are conducted.

it is possible to determine that the participants decreased the use of the replacement and restructuring strategies in the difficult segments of interpreting (Segments 2 and 3 compared with Segments 1 and 4, as well as Segment 2 in comparison with Segment 1, and Segment 3 compared to Segment 4).

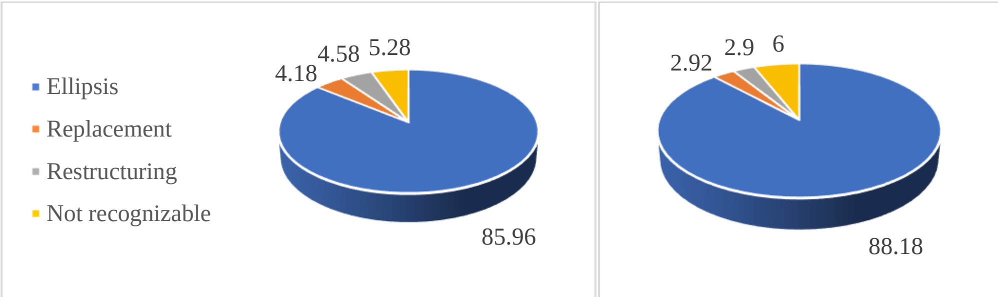


Figure 6-15. The students' (N=29) adoption of different note-taking strategies in easy (left) and difficult (right) segments of interpreting

Table 6-50. The students' adoption (N=29) of note-taking strategies in each segment of interpreting (Wilcoxon Signed Rank tests)

Strategy	Sub-category	Segment 1		Segment 2			Segment 3		Segment 4		
		M	SD	M	SD	Sig.	M	SD	M	SD	Sig.
Ellipsis		86.81	6.99	90.32	5.79	<.05	86.05	6.17	85.12	7.46	>.05
Non-ellipsis	Replacement	4.33	2.84	2.44	2.00	<.01	3.40	2.32	4.04	3.71	>.05
	Restructuring	3.91	4.95	2.25	2.35	>.05	3.56	3.26	5.24	4.80	>.05

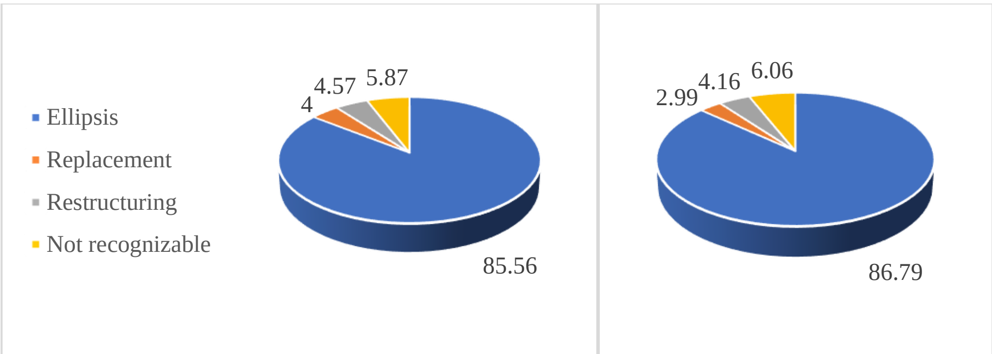


Figure 6-16. The professionals' (N=20) adoption of different note-taking strategies in easy (left) and difficult (right) segments of interpreting

Table 6-51. The professionals' (N=20) adoption of note-taking strategies in each segment of interpreting (Wilcoxon Signed Rank tests)

Strategy	Sub-category	Segment 1		Segment 2			Segment 3		Segment 4		
		M	SD	M	SD	Sig.	M	SD	M	SD	Sig.
Ellipsis		86.51	7.82	87.92	7.27	>.05	85.66	7.17	84.62	7.39	>.05
Non-ellipsis	Replacement	3.80	2.63	2.68	1.65	>.05	3.31	2.01	4.21	3.24	>.05
	Restructuring	4.21	3.93	2.89	3.60	>.05	5.43	3.78	4.93	2.85	>.05

6.3.3 Summary of RQ2

In summary, in response to RQ2, What are the impacts of interpreter work experience and source speech difficulty on interpreters' note-taking product?, this section examined four aspects of the product of note-taking, namely, note quantity, note form, note language and note-taking strategy. The most sensitive metric in responding to the two independent variables of interpreter work experience and source speech difficulty is found to be note quantity. From a between-group perspective, the professionals always produced larger quantities of notes than the students; while from a within-group perspective, the participants always presented greater numbers of notes in the easy segments of interpreting than in the difficult ones. Another prominent difference between the two groups of participants is their preference for full words and abbreviations during note-taking: whereas the professionals favoured full words during note production, the students opted to take more notes in abbreviations. Table 6-52 summarizes the results.

Table 6-52. Summarized results for RQ2

Aspect	Student	Professional
Note quantity	Easy>difficult	Easy>difficult
Note form	Language>symbol Abbreviation>full word	Language>symbol Full word>abbreviation
Note language	English>Chinese	English>Chinese
Note-taking strategy	Ellipsis>non-ellipsis	Ellipsis>non-ellipsis

Overall, H3 is partially supported by the result that the participants decreased the number of notes in the difficult segments of interpreting compared to in the easy ones, but it is also rejected by the finding that the professionals always took more notes than the students across the different segments of interpreting. In terms of H4, both groups preferred language notes and English (L2) notes regardless of source speech difficulty, which could be more effortful in production than symbols and Chinese (L1) notes. In addition, the professionals show a clear preference for full words, which are conventionally regarded as time-consuming note choices, in all the segments of interpreting. From this perspective, H4 is rejected by the findings. However, here the result is interpreted with a conventional understanding of the difference in the effort of taking notes across varied forms and languages. In Section 6.4, the effort of note-taking in various note forms and note languages is measured with eye-tracking and pen-recording measures. Whether H4 is supported also requires further evidence about the production and reception effort of these note forms and note languages, presented in

Section 6.5.

6.4 The process of note-reading

The third research question in the present study concerns how the participants read back their notes during the output phase of VRI. The participants' reading patterns are examined in Section 6.4.1 with regard to heat maps, gaze plots and event logs. Their cognitive effort of note-reading is explored in Section 6.4.2 by examining a series of fixation- and visit-related eye-tracking measures. Altogether, these two subsections address RQ3 and the corresponding hypotheses, as presented below:

RQ3: How do interpreters read back their notes? Is note-reading a cognitively demanding task?

> H5: As interpreters usually take notes in groups based on the meaning units in the source speech (e.g., Albl-Mikasa, 2008; S. Chen, 2022), they are expected to read notes in groups during target speech production.
>
> H6: Note-reading could be very cognitively demanding considering that the notes are highly condensed and varied in forms and languages.

6.4.1 Visualization tools

Firstly, after a visual check of each participant's heat maps of note-reading in each segment of interpreting, it is clearly evident that the participants' fixations on the screen basically match the layout of their notes. One difference that could be sometimes identified between the professional interpreters (Figure 6-17) and the student interpreters (Figure 6-18) is that around one-third of the students show fixations in the speaker window, whereas only few of the professionals did so. The students explained in the retrospective interview that this was because they were instructed by their teachers to keep eye contact with the interlocutors in the interpreting class. Still, for most of the time, they followed the development of the notes to conduct note-reading and target speech production.

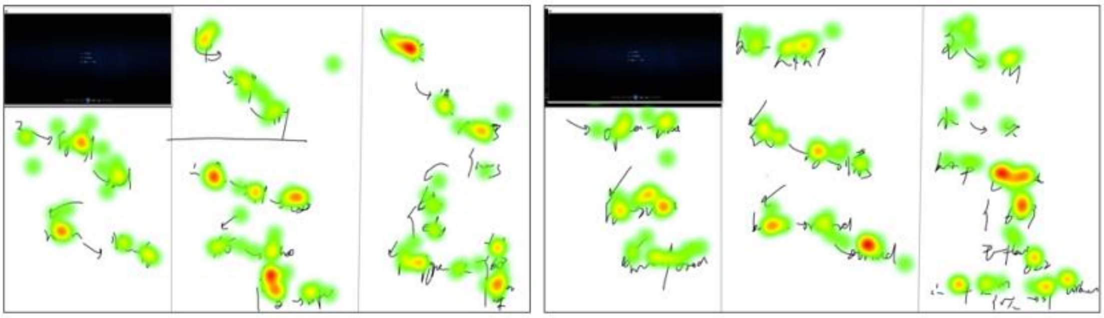


Figure 6-17. The heat maps of P52 (professional) during the process of note-reading in easy (left) and difficult (right) segments of interpreting

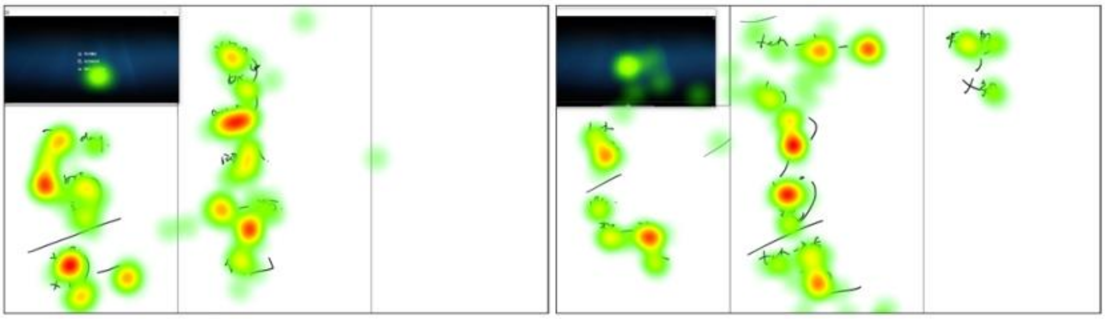


Figure 6-18. The heat maps of P05 (student) during the process of note-reading in easy (left) and difficult (right) segments of interpreting

Secondly, a closer look at the gaze plot of the participants reveals that note-reading is a non-linear process. Figure 6-19 presents a part of the gaze plot of P26. It can clearly be identified that the 50th, 51st and 53rd fixations were located at the notes that were placed before the notes that received the 25th and 29th fixations. This non-linearity of reading can be well illustrated with the event log exported from the Tobii Pro Lab, which clearly lists the AOIs that received fixations in sequence. As shown in Figure 6-20, from the 34th fixation to the 44th fixation (indicated by the first blue rectangle in the Figure), the participant constantly shifted between the notes of "musician" (*乐*) and "note" (the abbreviation of *调*). These two notes belonged to the same note group (as indicated by the red rectangle in Figure 6-19), which corresponded to the sentence that "*He didn't read a note*" in the source speech). Then, she moved forward slightly to read the notes of "hearing" (*听*) and "impaired" (x) (the 45th and 46th fixations in Figure 6-20), which belonged to the next note group ("*And he was profoundly hearing-impaired*"). However, through checking the audio recording of the participant's interpretation, it is found that, at this time, she did not finish the interpretation of the previous note group. Therefore, after such a glimpse of the two notes in the next note group, the participant went back to the former group of notes again to finish the interpretation (from the 47th fixation to the 61st fixation in Figure 6-20). Such group-based and group-crossed reading patterns are frequently observed in both groups of participants' interpreting process and both conditions of task difficulty, suggesting a common gazing pattern during the process of note-reading.

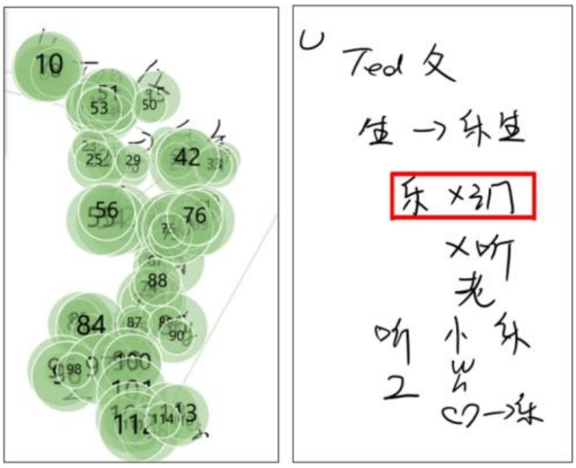


Figure 6-19. An example of gaze plots on notes (P26_student)

Participant	Fixation sequence	Start time	Stop time	Fixation duration	AOI name (coded notes)	Identifeid note	Source speech unit
P26	34	14162	14462	300	E%ab_ch*乐^1$1musician	乐(2nd)	musician
P26	35	14487	14979	492	E%ab_ch*调^1$1note	调	note
P26	36	15004	15421	417	E%sym*x^1$1didn't	乐(2nd)	musician
P26	37	15446	15829	383	E%ab_ch*调^1$1note	调	note
P26	38	15854	16004	150	E%ab_ch*调^1$1note	调	note
P26	39	16187	16471	283	E%ab_ch*调^1$1note	调	note
P26	40	16512	16712	200	E%ab_ch*乐^1$1musician	乐(2nd)	musician
P26	41	16729	16962	233	E%ab_ch*乐^1$1musician	乐(2nd)	musician
P26	42	16996	17437	442	Blank space		
P26	43	17471	17796	325	E%sym*x^1$1didn't	x	didn't
P26	44	17954	18129	175	E%ab_ch*调^1$1note	调	note
P26	45	18154	18646	492	E%ab_ch*听^1$1hearing	听	hearing
P26	46	18671	18946	275	RS%sym*x^1$1impaired	x	impaired
[illegible]	47	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
P26	48	19179	19387	208	E%ab_ch*乐^1$1musician	乐(2nd)	musician
P26	49	19421	19612	192	Blank space		
P26	50	19637	19754	117	Blank space		
P26	51	19929	20312	383	RS%fw_en*Ted^1$2ted	Ted	Ted
P26	52	20354	20929	575	E%ab_ch*乐^1$1musician	乐(2nd)	musician
P26	53	20971	21287	317	Blank space		
P26	54	21321	22162	842	E%ab_ch*乐^1$1musician	乐(2nd)	musician
P26	55	22321	23104	783	E%ab_ch*乐^1$1musician	乐(2nd)	musician
P26	56	23237	23637	400	E%ab_ch*乐^1$1musician	乐(2nd)	musician
P26	57	23671	23962	292	E%ab_ch*调^1$1note	调	note
P26	58	23979	24721	742	E%ab_ch*调^1$1note	调	note
P26	59	24879	25146	267	E%ab_ch*调^1$1note	调	note
P26	60	25154	25687	533	E%ab_ch*调^1$1note	调	note
P26	61	25896	26162	267	E%ab_ch*调^1$1note	调	note

2022/6/7

Figure 6-20. The event log of P26's note-reading process

Overall, the heat maps, gaze plots and evet logs in the present study present two major note-reading patterns. Firstly, interpreters would read notes that belonged to the same note groups back and forth while they were interpreting the current meaning chunk. Secondly, when they were going to finish the interpretation of the present note group, they would move forward to the next one for a glance and then come back to the present notes to finish interpretation.

6.4.2 Eye-tracking measures

Note-taking is a highly-individualized activity. One interpreter's notes can differ considerably from those of another. From this perspective, the participants were actually presented with different reading materials during the process of note-reading, i.e., the notes created by themselves. Without the same stimulus, there was no way to compare the cognitive effort of note-reading in different participant groups and task conditions. Therefore, instead of examining the effects of interpreter work experience and source speech difficulty on the note-reading effort, the present study only presents the MFD data (an indicator of cognitive effort) collected from the process of note-reading and conducts within-subject comparisons.

Altogether, there are two types of MFD data: one set obtained through the fixations on the noted area (see Figure 6-8), and the other gained through the fixations on the individual notes (see Figure 6-11). Both types of MFD data can indicate the cognitive effort of note-reading. However, compared with the former which provides a general picture of the note-reading effort by including the fixations on the blank space among the notes, the latter specifically targets the cognitive effort of note-processing during note-reading by having AOIs drawn on the individual notes. Figure 6-21 and Figure 6-22, respectively, display the student interpreters' and the professional interpreters' MFD data, in easy and difficult task conditions. After checking the normality of the differences in the two sets of data (Table 6-53), paired-sample t-tests are conducted to examine whether the differences between the two types of MFD reached a significant level. The results show that, in both task conditions, the students show significant differences in their cognitive effort of 'reading' the noted area and the cognitive effort of reading the individual notes. In other words, as long as the specific notes were involved in the reading process, the students' cognitive load would increase significantly. By contrast, the professionals present no such significant differences in the two types of MFD data in either task condition, suggesting that reading the notes per se did not increase their cognitive load significantly during the reformulation phase of CI.

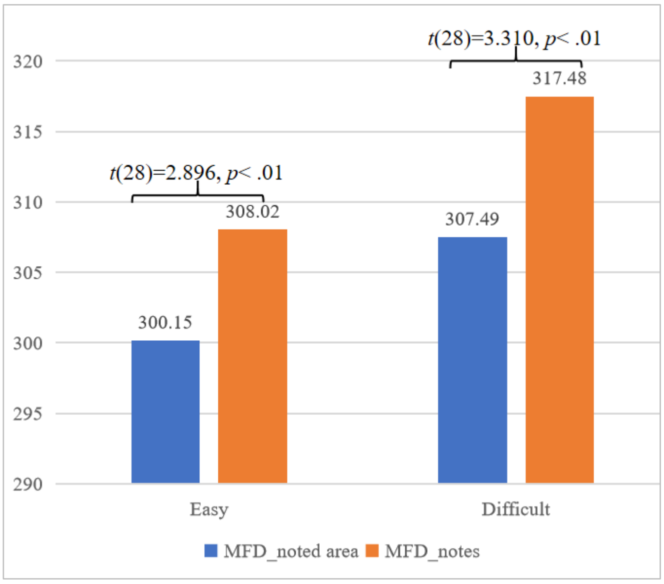


Figure 6-21. The students' MFD data during the process of note-reading

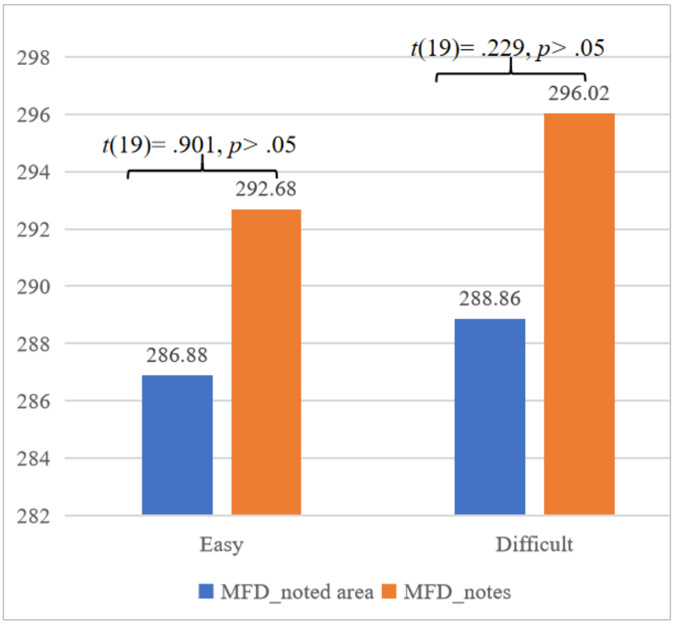


Figure 6-22. The professionals' MFD data during the process of note-reading

Table 6-53. Shapiro-Wilk normal distribution test results for within-group comparisons in the MFD data (easy vs. difficult)

		Statistic	***df***	**Sig.**
Student (N=29)	Easy	.953	29	>.05
	Difficult	.988	29	>.05
Professional (N=20)	Easy	.966	20	>.05
	Difficult	.951	20	>.05

It is worth mentioning that the reported MFD data are obtained from approximately 70% of the collected notes, because around 30% of the notes were not fixated at all during the process of note-reading in the present study (see Table 6-54), which is a comparatively higher percentage than in other studies. This could be caused by the fact that some of the fixations fell on the speaker window. It could also have resulted from the participants' strategic choice of reducing the cognitive load of note-reading by jumping some notes during note-reading. A further look at the heat maps and fixation event logs can visualize this strategic move of the participants.

Table 6-54. Skip rates during the process of note-reading

	Student (N=29)		**Professional (N=20)**	
	M	SD	M	SD
Segment 1	31.72%	13.13%	27.22%	12.31%
Segment 2	28.69%	13.75%	27.77%	13.71%
Segment 3	31.07%	12.28%	30.39%	20.42%
Segment 4	31.95%	12.03%	27.68%	15.56%

The heat maps show that both the students (see Figure 6-23) and the professionals (Figure 6-24) had some fixations on the blank space on the screen. Similarly, in their eye-tracking event logs (see grey rows in Figure 6-25 and Figure 6-26), a certain proportion of the fixations are found to be located on the blank space on the screen. It seems that, in the reformulation phase of CI, instead of staring at the notes, the participants sometimes would look away from the notes and deliver interpretations with eyes fixating on nothing. As presented in the previous section, regardless of interpreter work experience and source speech difficulty, the MFD values obtained from the individual notes are always higher than those obtained from the noted areas, suggesting the high cognitive load of note-reading. Therefore, the comparatively higher skip rate of note-reading observed in the present study could be attributed to the participants' strategic move of reducing the interference caused by note-reading during the process of target speech production. In this way, the participants could try to maintain their cognitive load of interpreting within an acceptable level and ensure the delivery of the target speech. Further discussions of this can be found in Chapter 7, Section 7.2.2.

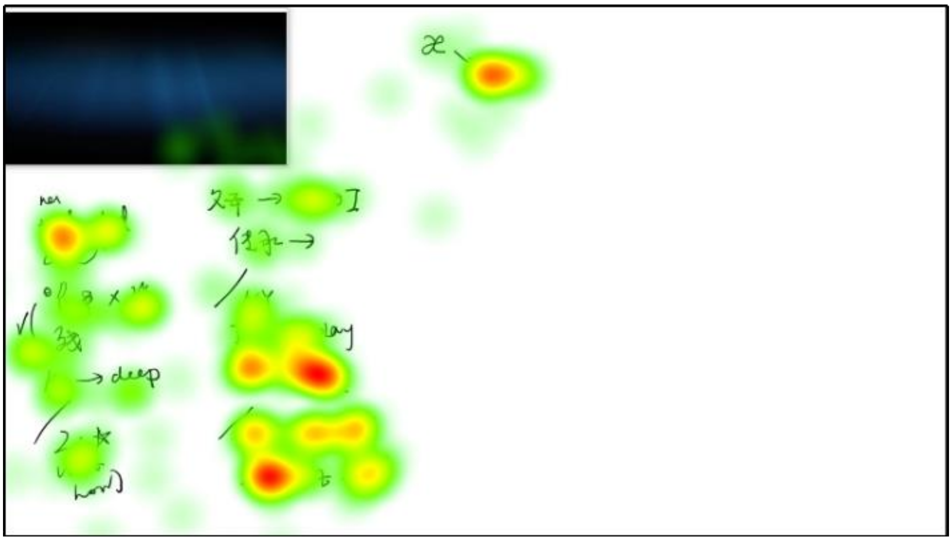


Figure 6-23. An example of the heat map of P01 (student) in Segment 1 of interpreting

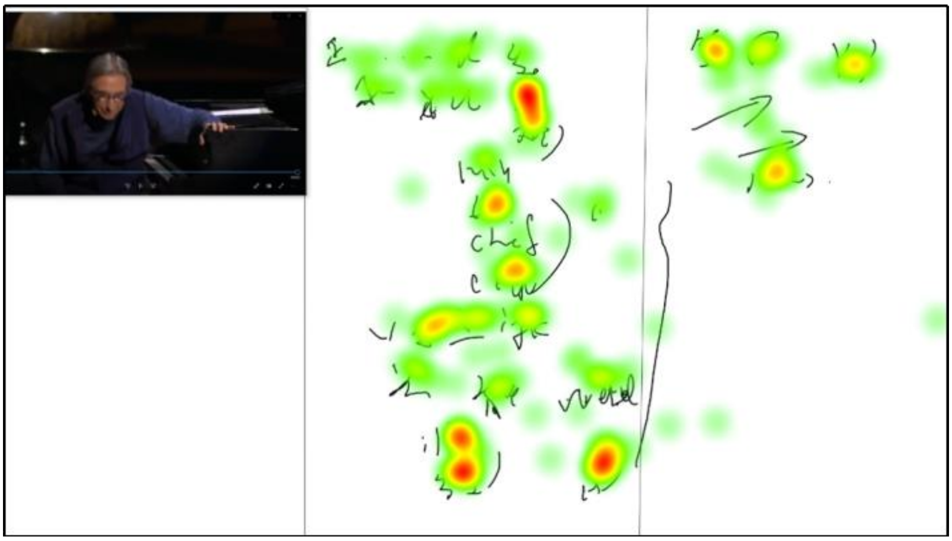


Figure 6-24. An example of the heat map of P54 (professional) in Segment 1 of interpreting

Participant	Fixation sequence	Start time	Stop time	Fixation duration	AOI name (coded notes)
P01	45	16529	17429	900	RS%sym*pen+○^2$1illustrator
P01	46	17446	17821	375	RS%ab_ch*扌 ^1$1ted
P01	47	18162	18354	192	Blank space
P01	48	18379	18954	575	E%ab_ch*谱^1$1note
P01	49	18979	19329	350	RS%ab_ch*扌 ^1$1ted
P01	50	19337	19796	458	Blank space
P01	51	19821	20929	1108	E%ab_ch*谱^1$1note
P01	52	20946	21596	650	Blank space
P01	53	21604	21837	233	E%ab_ch*谱^1$1note
P01	54	21871	22029	158	Blank space
P01	55	22037	22804	767	RS%ab_en*
P01	56	22829	23296	467	RP%ab_ch*残^1$1impaired
P01	57	23304	23446	142	Blank space
P01	58	23454	23612	158	RS%ab_en*
P01	59	23646	24162	517	E%ab_ch*谱^1$1note
P01	60	24196	24462	267	RP%ab_ch*残^1$1impaired
P01	61	24479	24746	267	RS%ab_en*
P01	62	24796	25054	258	Blank space
P01	63	25087	25304	217	Blank space
P01	64	25321	25554	233	Blank space
P01	65	26796	27112	317	RS%sym*pen+○^2$1illustrator
P01	66	27129	27587	458	RS%ab_en*
P01	67	27612	27829	217	RP%ab_ch*残^1$1impaired
P01	68	27854	27996	142	RS%ab_en*
P01	69	28029	28496	467	E%ab_ch*谱^1$1note

Figure 6-25. A screenshot of the event log of P01 (student) in Segment 1 of interpreting

Participant	Fixation sequence	Start time	Stop time	Fixation duration	AOI name (coded notes)
P54	65	28412	28737	325	Blank space
P54	66	28937	29421	483	Blank space
P54	67	29454	29679	225	Blank space
P54	68	29704	30329	625	E%fw_en*head^1$1heard
P54	69	30354	30554	200	Blank space
P54	70	30587	31137	550	E%ab_en*sym^1$2symphony
P54	71	31179	31479	300	Blank space
P54	72	31521	31771	250	E%ab_en*Bee^1$2beethoven
P54	73	31796	31946	150	E%ab_en*sym^1$2symphony
P54	74	32096	32246	150	Speaker window
P54	75	32262	32746	483	Speaker window
P54	76	32829	32912	83	Blank space
P54	77	33071	33154	83	Blank space
P54	78	33179	33512	333	E%fw_en*music^1$20music
P54	79	33529	33871	342	RS%ab_ch*意^1$2mean
P54	80	33887	34137	250	Blank space
P54	81	34154	34296	142	RS%ab_ch*意^1$2mean
P54	82	34329	34746	417	Blank space
P54	83	34779	34946	167	Blank space
P54	84	34962	35179	217	RS%ab_ch*意^1$2mean
P54	85	35204	35362	158	Blank space
P54	86	35396	35737	342	RS%fw_ch*老人^2$5man
P54	87	35771	35996	225	E%fw_ch*找^1$1recover
P54	88	36021	36162	142	RS%ab_ch*意^1$2mean
P54	89	36187	36471	283	Blank space

Figure 6-26. A screenshot of the event log of P54 (professional) in Segment 4 of interpreting

6.4.3 Summary of RQ3

There were two major reading patterns during the process of note reception. Firstly, the participants read back and forth in the same note group to generate a holistic understanding of the individual notes. Secondly, they would also read across the note groups to prepare for the next part of their interpretation and ensure a fluent delivery of the target speech. With these observations, H5 is corroborated. Moreover, the observed long MFD on the notes and high skip rate during note-reading indicate that note reception is a cognitively-demanding process, providing direct evidence to support H6.

6.5 Associations between the note-taking effort and the note-reading effort

This section reports how the effort of note production and the effort of note reception associate with each other. It firstly examines the relationship between the average effort of taking and reading one note through correlation tests (Section 6.5.1), and then explores whether there were trade-offs in the note-taking and note-reading effort across different note forms and note languages (Section 6.5.2). Altogether, these two subsections respond to RQ4 and the corresponding hypotheses that are summarized below:

RQ4: Is there a trade-off between the note-taking effort and the note-reading effort?

> H7: A trade-off is expected to be found between the overall effort of note-taking and that of note-reading.
>
> H8: This trade-off between the two types of noting effort is also expected to be observed in the comparisons of different note forms (language notes vs. symbols; full words vs. abbreviations) and note languages (Chinese vs. English).

The effort of note-taking includes the devotion of overt visual attention to each note (TFD), the physical effort of note-writing (CC), the cognitive effort of note-taking (MFD) and the temporal management of note-planning (EPS). The note-reading effort mainly includes the distribution of overt visual attention (TFD on each note) and the cognitive effort of note- reading (MFD), the latter which could be further divided into that of incomplete processing (RVC), note recognition (FFD), the integration of note meanings at a lexical level (FPD) and the integration of note meanings at a syntactical and textual level (SPD).

It is worth mentioning that Pearson's correlation tests are conducted to explore the relationships between the participants' note-taking effort and note-reading effort. As a proportion of the data were not normally distributed[19], additional non-parametric Spearman's correlation tests are also conducted. To ensure the consistency of the reported data, when the two correlation tests return the same results, only Pearson's results are reported. When the two correlations tests show different results, the selected correlation test and the corresponding result are specified.

6.5.1 Correlations between the effort of the two noting stages

Table 6-55 and Table 6-56, respectively, present the students' and professionals' correlation test results, between the effort of note-taking and the effort of note-reading in easy and difficult segments of interpreting. A general tendency for these results is that all the significant correlations are positive. In other words, an increase in the note-taking effort is associated with an increase, rather than a decrease, in the note-reading effort. Moreover, many coefficients in the results that achieve significance are above 0.50, suggesting a strong level of the positive correlations between these two types of noting effort. Therefore, H7 is rejected by these results.

Table 6-55. The correlations between the effort of note-taking and the effort of note-reading in the student group (N=29)

		NT effort							
		Easy				*Difficult*			
		TFD	MFD	EPS	CC	TFD	MFD	EPS	CC
NR effort	TFD	.814**	.340	.354#	.222#	.767**	.389*	.236	.236
	MFD	.506**	.655**	.280	.302	.374*#	.598**	.026	.154
	RVC	.529**	.093	.309	.208	.685**	.184	.265	.108
	FFD	.440**	.646**	.307	.160	.287	.445*	.019	.159
	FPD	.531**	.251	.474**	.318	.298	.233	-.071	.255#
	SPD	.813**	.334	.481**	.233#	.777**	.380*	.267	.190

Notes.
1. NT=note-taking, NR=note-reading.
2. "#" means that the result is obtained through a Spearman correlation test rather than a Pearson correlation test.

Table 6-56. The correlations between the effort of note-taking and the effort of note-reading in the professional group (N=20)

		NT effort							
		Easy				*Difficult*			
		TFD	MFD	EPS	CC	TFD	MFD	EPS	CC
NR effort	TFD	.678**	.371	.351	.474*	.695**	.502**	-.066	.415

[19] In total, 2 interpreter groups*2 task conditions* (4 indicators of the process of note-taking + 6 indicators of the process of note-reading)=40 data normality tests are conducted.

MFD	.330	.484*	.004	.245	.405	.595**	.068	.228
RVC	.668**	.194	.425	.461*	.594**	.288	.019	.364
FFD	.249	.555*	-.058	.002	.395	.686**	.066	.014
FPD	.312	.224	.119	.098	.403	.451*	-.005	.166
SPD	.705**	.376	.373	.515*	.558#	.478*	.001	.438

Notes.

1. NT=note-taking, NR=note-reading.
2. "#" means that the result is obtained through a Spearman correlation test rather than a Pearson correlation test.

Both similarities and differences are observed in the two groups of participants' correlation results (Table 6-57). Firstly, the eye-tracking measures which indicate the cognitive aspects of the two noting efforts show many positive correlations with each other in both groups of participants. All of these positive correlations indicate that what troubled the participants during the process of note-taking still bothered them during the process of note-reading. According to the specific difficulties they encountered during note-taking, their note-reading effort could change correspondingly at different stages (early and late) and in different aspects (visual and cognitive). For instance, an increase in the overt visual attention (TFD) in the note-taking stage is associated with an increase in overt visual attention (TFD), the cognitive effort of incomplete processing (RVC) and late-stage processing during note-reading (SPD). A greater amount of cognitive effort of note-taking (MFD) is also positively correlated with a larger amount of cognitive effort of note-reading (MFD and FFD). For both groups, when the source speech was difficult, the cognitive effort of note-taking (MFD) is associated with the overt visual attention (TFD) and cognitive effort of late-stage processing in note-reading (SPD). Secondly, for the pen-recording measures, only the students present significant correlations between the time to make note decisions (EPS) and the effort of note-reading in the easy segments of interpreting. Such correlations are especially noticeable in the cognitive effort of integrating the meaning of notes at both lexical and syntactical levels (FPD and SPD). This implies that, the longer the students needed to plan for a note during the process of note-taking, the more effort they needed to integrate the meaning of the individual notes into a coherent whole during the process of note-reading. Thirdly, for the physical effort of note-writing (CC), the professional group present significant positive correlations with the overt visual attention during note-reading (TFD), the cognitive effort of incomplete processing (RVC) and that of late-stage processing (SPD) in the easy segments. This result echoes with what was found in Section 6.3.1, that the largest note quantity was achieved by the professionals in the easy task condition with frequent use of the ellipsis strategy. The inadequate comprehension of the source speech during the input phase of CI could thus entail reading difficulties during the output phase of CI. Finally, a further look at the coefficients reveals that: for the students, the coefficients are generally larger in the easy segments of interpreting than in the difficult ones; while for the professionals, it is in the difficult segments that the cognitive effort of note-taking presents a series of positive correlations with the cognitive effort of note-reading. These results indicate that the connection between the note-taking effort and note-reading effort could vary in different interpreter groups and task conditions.

Table 6-57. Summarized results of the significant positive correlations between the effort of note-taking and the effort of note-reading in the two groups of participants

		NT effort			
		Overt visual attention (TFD)	Overall cognitive effort (MFD)	Temporal management (EPS)	Physical effort (CC)
NR effort	Overt visual attention (TFD)	All applied	Both_D		Pro_E
	Overall cognitive effort (MFD)	Stu_E&D	All applied		
	Incomplete processing (VC)	All applied			Pro_E
	Note recognition (FFD)	Stu_E	All applied		
	Note integration (lexical-level) (FPD)	Stu_E	Both_D	Stu_E	
	Note integration (sentence-level) (SPD)	All applied	Both_D	Stu_E	Pro_E

Notes.
1. NT=note-taking, NR=note-reading.
2. “All applied” represents that significant positive correlations are observed in both interpreter groups and task conditions.
3. Stu=student group, Pro=professional group, Both=both groups, E=easy segments of interpreting, and D=difficult segments of interpreting.

6.5.2 Comparisons of the note-taking and note-reading effort across note categories

Following the Levels of Processing Hypothesis (Craik & Lockhart, 1972), a deeper level of language processing could lead to an easier recall of the processed information. If note activities conform to this hypothesis, then the note forms, note languages and note-taking strategies that demand in-depth language processing during the process of note-taking should be easy to recall during the process of note-reading. To test this hypothesis, the note-taking and note-reading effort between different note forms (language vs symbol; full word vs abbreviation) and note languages (Chinese vs

English) are calculated and compared[20]. For data that did not meet the assumptions of parametric paired-sample *t*-tests, additional non-parametric Wilcoxon Signed Rank tests are conducted to see whether the two tests return the same results. For ensuring consistency of the results, only the paired-sample *t*-test results are presented in this section to ensure the conformity of the reported data.

The effort of taking and reading notes in language and symbol are presented in Tables 6-58 to 6-61, with the first two tables showing the students' data and the last two providing the professionals' data. It is found that, regardless of interpreter work experience and source speech difficulty, the effort of taking and reading language notes is always significantly greater than that of symbols (except for the students' EPS in the difficult segments of interpreting). Even so, the participants still preferred language notes over symbols during note-taking (see Section 6.3). Given these findings, if the use of language notes doesn't enhance interpretation quality, then the great effort that is expended on taking and reading the language notes would appear to not avail. The results concerning the correlation between the use of language notes and the interpreting performance are presented in Section 6.6.

Table 6-58. Effort of taking notes in language and symbol (the student group, N=29)

		Easy				Difficult			
Category		TFD	CC	MFD	EPS	TFD	CC	MFD	EPS
Language	M	1486.78	3.25	564.64	2418.22	1386.70	3.25	518.41	4963.11
	SD	428.60	.72	169.55	929.58	426.32	.82	132.71	962.19
Symbol	M	835.59	1.82	465.86	1284.94	826.19	1.74	469.09	5175.76
	SD	220.93	.28	117.01	937.30	254.38	.31	109.650	2946.89
Sig.		<.001	<.001	<.001	<.001	<.001	<.001	<.01	>.05

Table 6-59. Effort of reading notes in language and symbol (the student group, N=29)

	Category		TFD	MFD	RVC	FFD	FPD	SPD
Easy	Language	M	1424.84	318.15	2.27	305.55	407.01	1017.85
		SD	544.86	73.33	.03	51.25	87.91	497.80
	Symbol	M	731.40	285.03	1.17	281.99	324.77	406.65
		SD	286.00	60.73	.78	65.52	90.31	247.96
	Sig.		<.001	<.001	<.001	<.01	<.01	<.001
Difficult	Language	M	1505.79	328.20	2.31	312.43	411.67	1094.13
		SD	571.45	75.06	1.18	65.65	85.70	542.69
	Symbol	M	846.74	298.71	1.32	275.88	343.63	503.13
		SD	366.22	98.27	1.02	85.98	163.94	347.25
	Sig.		<.001	<.01	<.001	<.01	<.01	<.001

[20] Altogether, 2 interpreter groups*2 task conditions*8 note categories* (4 indicators of the process of note-taking +6 indicators of the process of note-reading)=320 data normality tests are conducted.

Table 6-60. Effort of taking notes in language and symbol (the professional group, N=20)

		Easy				Difficult			
Category		TFD	CC	MFD	EPS	TFD	CC	MFD	EPS
Language	M	1507.86	3.50	604.53	2545.39	1499.96	3.85	585.40	2607.45
	SD	392.47	.86	174.36	919.01	356.00	1.02	171.23	875.39
Symbol	M	759.70	1.97	425.70	1285.83	804.70	1.97	472.83	1759.50
	SD	251.98	.36	109.21	708.43	240.74	.48	162.83	890.83
Sig.		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.01

Table 6-61. Effort of reading notes in language and symbol (the professional group, N=20)

	Category		TFD	MFD	RVC	FFD	FPD	SPD
Easy	Language	M	1409.69	295.63	2.26	280.96	393.25	1016.46
		SD	595.44	62.45	.78	50.84	104.19	528.96
	Symbol	M	648.09	247.85	1.12	243.65	269.73	378.41
		SD	248.36	46.27	.64	53.38	63.89	215.76
	Sig.		<.001	<.001	<.001	<.01	<.001	<.001
Difficult	Language	M	1505.88	305.29	2.38	297.40	396.42	1109.48
		SD	619.98	75.31	.67	59.98	107.38	552.96
	Symbol	M	692.44	265.48	1.13	258.25	289.47	402.99
		SD	267.35	63.82	.53	63.21	85.28	242.77
	Sig.		<.001	<.01	<.001	<.01	<.001	<.001

Language notes can be further divided into full words and abbreviations according to the form of the notes, or Chinese and English notes based on the language of the notes. For the note-taking effort between full words and abbreviations, both groups present longer TFD, more CC, shorter MFD and shorter EPS in the former compared to the latter (as shown by the data presented in Table 6-62 and Table 6-63). These results indicate that, noting in full words demanded more overt visual attention and physical effort of handwriting than noting in abbreviations, and taking notes in abbreviations required more cognitive effort than taking notes in full words. In addition, when the source speech was difficult, the participants needed significantly longer time to compose abbreviations than full words. In terms of the note-reading effort, it is always significantly more demanding to read the notes in full words than read notes in abbreviations. This tendency exists in both participant groups and task conditions (Table 6-64 and Table 6-65). Looking back to the participants' use of these two note forms during note-taking, the students preferred abbreviations over full words, whereas the professionals showed the opposite preference. It appears that the students opted for the note forms that require less overt visual attention and physical effort in handwriting during note-taking, even though they had to deal with a great cognitive demand and longer note-planning time during note production. The professionals, on the other hand, counter-intuitively adopted note forms that demand much note-taking and note-reading effort. However, if the use of full words could enhance their interpreting performance, then this could explain why the professionals preferred full words over abbreviations.

Table 6-62. Effort of taking notes in full word and abbreviation (the student group, N=29)

		Easy				Difficult			
Category		TFD	CC	MFD	EPS	TFD	CC	MFD	EPS
Full word	M	1620.24	3.27	550.46	2325.26	1468.81	3.37	525.31	2284.28
	SD	464.94	.76	174.19	1042.70	512.32	.87	151.59	1036.42
Abbreviation	M	1417.57	3.19	588.58	2420.70	1286.20	3.20	550.01	2500.18
	SD	408.90	.66	181.86	940.37	312.84	.70	170.32	925.03
Sig.		<.05	>.05	<.05	>.05	<.01	>.05	>.05	<.05

Table 6-63. Effort of taking notes in full word and abbreviation (the professional group, N=20)

		Easy				Difficult			
Category		TFD	CC	MFD	EPS	TFD	CC	MFD	EPS
Full word	M	1588.38	3.71	571.74	2453.33	1576.05	4.11	577.14	2534.49
	SD	419.43	.82	159.48	854.06	398.70	1.07	276.73	854.49
Abbreviation	M	1417.85	3.30	647.94	2632.84	1373.85	3.50	596.21	2690.48
	SD	429.00	.98	229.18	1043.53	319.47	.88	156.30	938.07
Sig.		<.01	<.01	<.05	>.05	<.01	<.01	>.05	<.05

Table 6-64. Effort of reading notes in full word and abbreviation (the student group, N=29)

	Category		TFD	MFD	RVC	FFD	FPD	SPD
Easy	Full word	M	1468.47	315.33	2.42	306.33	402.40	1066.06
		SD	539.47	70.51	1.18	55.84	86.69	500.59
	Abbreviation	M	731.40	285.03	1.17	281.99	324.77	406.65
		SD	286.00	60.73	.78	65.52	90.31	247.96
	Sig.		<.001	<.01	<.001	<.05	<.001	<.001
Difficult	Full word	M	1592.56	319.84	2.47	301.21	407.73	1184.81
		SD	622.69	96.78	1.34	82.55	122.94	562.14
	Abbreviation	M	846.74	298.71	1.32	275.88	343.63	503.13
		SD	366.22	98.27	1.02	85.98	163.94	347.25
	Sig.		<.001	<.05	<.001	<.05	<.05	<.001

Table 6-65. Effort of reading notes in full word and abbreviation (the professional group, N=20)

	Category		TFD	MFD	RVC	FFD	FPD	SPD
Easy	Full word	M	1416.72	286.28	2.35	272.30	378.56	1038.17
		SD	601.95	56.02	.80	52.11	105.11	553.10
	Abbreviation	M	648.09	247.85	1.12	243.65	269.73	378.41
		SD	248.36	46.27	.64	53.38	63.89	215.76
	Sig.		<.001	<.01	<.001	<.05	<.001	<.001
Difficult	Full word	M	1619.99	300.99	2.54	286.25	404.57	1215.44
		SD	731.02	72.09	.82	46.93	123.16	681.25

	Abbreviation	M	692.44	265.48	1.13	258.25	289.47	402.99
		SD	267.35	63.82	.53	63.21	85.28	242.77
	Sig.		<.001	<.01	<.001	<.01	<.001	<.001

On Chinese and English notes, the student group (see Table 6-66) present no significant differences in the overt visual attention (TFD) they paid to the notes during note-taking and the time they spent on note-planning (EPS). Compared with noting in English, noting in Chinese entailed significantly more physical effort of note-writing (CC) in the easy segments of interpreting and cognitive effort of note-taking (MFD) in both task conditions. In terms of the effort of note-reading (Table 6-67), only two close-to-significant differences are observed in the TFD and SPD data in the easy segments of interpreting. Except for that, reading Chinese notes was similarly as effortful as reading English notes for the students.

Table 6-66. Effort of taking notes in Chinese and English (the student group, N=29)

		Easy				**Difficult**			
Category		TFD	CC	MFD	EPS	TFD	CC	MFD	EPS
Chinese	M	1470.66	3.45	625.28	2369.28	1343.34	3.29	593.31	2486.19
	SD	482.53	.88	214.80	1016.84	529.05	.92	214.13	1032.63
English	M	1500.72	3.10	536.76	2403.04	1357.44	3.15	509.49	2395.57
	SD	427.14	.68	164.27	921.92	395.23	.76	145.95	912.66
Sig.		>.05	<.05	<.01	>.05	>.05	>.05	<.01	>.05

Table 6-67. Effort of reading notes in Chinese and English (the student group, N=29)

	Category		**TFD**	**MFD**	**RVC**	**FFD**	**FPD**	**SPD**
Easy	Chinese	M	1330.76	328.37	2.12	313.66	391.45	939.36
		SD	538.56	92.17	1.01	79.96	113.69	490.02
	English	M	1436.88	341.82	2.30	303.04	417.99	1018.89
		SD	530.93	72.81	1.11	53.47	102.71	479.59
	Sig.		>.05 (.057)	>.05	>.05	>.05	>.05	>.05 (.056)
Difficult	Chinese	M	1434.62	340.85	2.28	332.49	412.73	1021.90
		SD	619.94	85.96	1.41	92.08	114.47	613.35
	English	M	1494.32	324.76	2.25	313.03	419.33	1074.99
		SD	546.74	87.05	.92	82.02	95.66	492.02
	Sig.		>.05	>.05	>.05	>.05	>.05	>.05

As for the professionals, in the easy segments of interpreting, taking Chinese notes entailed significantly more overt visual attention (TFD), more cognitive effort (MFD) and longer note-planning (EPS) than taking English notes; while in the difficult segments, only one significant difference is observed in the MFD data (Table 6-68). When it comes to note-reading, it was only in the TFD and SPD data that the professionals show significant differences in the effort of reading these two types of notes in the easy segments of interpreting (see Table 6-69). Except for that, no other significant results are found.

Table 6-68. Effort of taking notes in Chinese and English (the professional group, N=20)

		Easy				Difficult			
Category		TFD	CC	MFD	EPS	TFD	CC	MFD	EPS
Chinese	M	1600.73	3.32	676.07	2812.56	1343.34	3.29	593.31	2486.19
	SD	453.88	.94	232.59	1223.88	529.05	.92	214.13	1032.63
English	M	1462.20	3.56	545.90	2445.07	1357.44	3.15	509.49	2395.57
	SD	421.66	.83	142.83	909.70	395.23	.76	145.95	912.66
Sig.		<.05	>.05	<.001	<.05	>.05	>.05	<.01	>.05

Table 6-69. Effort of reading notes in Chinese and English (the professional group, N=20)

	Category		**TFD**	**MFD**	**RVC**	**FFD**	**FPD**	**SPD**
Easy	Chinese	M	1335.32	297.74	2.15	279.39	398.19	937.15
		SD	600.37	68.31	.85	52.48	153.91	500.68
	English	M	1470.69	295.35	2.31	279.59	400.96	1069.75
		SD	633.79	62.78	.80	51.17	100.72	564.99
	Sig.		<.05	>.05	>.05	>.05	>.05	<.01
Difficult	Chinese	M	1468.28	303.22	2.42	302.52	388.27	1080.05
		SD	764.24	77.50	1.06	79.94	115.17	691.32
	English	M	1525.13	305.58	2.39	292.30	399.84	1125.32
		SD	598.71	75.87	.69	61.17	115.24	527.96
	Sig.		>.05	>.05	>.05	>.05	>.05	>.05

Overall, the two groups present greater differences in the note-taking effort rather than in the note-reading effort between Chinese and English notes. Chinese notes generally required more note-taking effort than English notes in both task conditions, but it was only in the easy segments that Chinese notes required less note-reading effort than English notes. This result means that the trade-off between Chinese and English notes in the note-taking and note-reading effort only exists in the easy task condition, which partially supports the present study's H8.

6.5.3 Summary of RQ4

Overall, a series of moderate-to-strong significant positive correlations are observed between the effort of note-taking and the effort of note-reading, suggesting that an increase in the effort of note production is associated with an increase in the effort of note reception. Such positive correlations are found to be especially frequent for the students in the easy segments of interpreting and noticeably stronger for the professionals in the difficult segments of interpreting. Taking a closer look at the significant correlations, SPD, which points to the cognitive effort of note integration at a syntactical and textual level, is often found to have the strongest correlations. This suggests that, when the note-reading effort increased along with the note-taking effort, the extra cognitive effort could be devoted to integrating the meaning of notes at a textual level. With these results, H7 is rejected.

H8 is only partially supported by the results, as summarized in Table 6-70. On the one hand, it is observed that, between language notes and symbol notes, the former always required a larger amount of note-taking and note-reading effort from the participants than the latter, throwing doubt on H8. On the other hand, between full words and abbreviations, it is found that, during note-taking, abbreviations demanded more cognitive effort than full words. When it comes to note-reading, the situation is the opposite. Such a phenomenon is also observed in the comparisons of Chinese notes and English notes, but it only exists in the easy segments of interpreting rather than in the difficult ones. Based on these results, H8 is supported under specific conditions.

Table 6-70. Summarised results of the note-taking and note-reading effort of different note categories

		Language vs. symbol	**Full word vs. abbreviation**	**Chinese vs. English**
Note-taking effort	Visual	LG>SYM	FW>AB	CN>EN
	Physical	LG>SYM	FW>AB	CN>EN
	Cognitive	LG>SYM	AB>FW	CN>EN
	Temporal	LG>SYM	AB>FW	CN>EN
Note-taking preference		LG>SYM	Pro: FW>AB Stu: AB>FW	EN>CN
Note-reading effort	Visual	LG>SYM	FW>AB	EN>CN (E)
	Overall	LG>SYM	FW>AB	EN≈CN
	Incomplete	LG>SYM	FW>AB	EN≈CN
	Early-stage	LG>SYM	FW>AB	EN≈CN
	Late-stage	LG>SYM	FW>AB	EN>CN (E)

Note. LG=language notes, SYM=symbol, FW=full word, AB=abbreviation, CN=Chinese, EN=English, and E=easy segments.

6.6 Associations between note-taking behaviour and interpretation quality

This section firstly looks into the participants' interpreting scores in different task conditions (Section 6.6.1), and then examines how these scores correlate with the indicators of the process of note-taking (Section 6.6.2), the product of note-taking (Section 6.6.3) and the process of note-reading (Section 6.6.4). It sets out to address RQ5 and the corresponding hypotheses as summarized below:

RQ5: What is the relationship between interpreters' note-taking behaviour and interpretation quality?

H9: Increased effort in note-taking would be associated with better interpreting performance because of a potentially deeper level of SL processing.

H10: A greater quantity of notes, a larger proportion of notes in symbols (compared with language notes), and a higher percentage of notes in TL (compared with SL) could lead to better interpretation quality.

H11: Increased note-reading effort would be correlated with poorer interpretation quality, as the processing capacity left for target speech production could be limited.

All the tested correlations between the participants' note-taking behaviour and interpreting performance are illustrated in Figure 6-27. Again, Shapiro-Wilk test results[21] showed that a proportion of the data was not normally distributed. Therefore, except for Pearson's correlation tests, additional non-parametric tests are also conducted to ensure that the two test results point to the same tendency. To ensure the consistency of the reported data, when the two correlation tests return the same results (which is frequently observed in data analysis), only Pearson's results are reported. When the two correlations tests show different results, the appropriate correlation test and the corresponding result are specified.

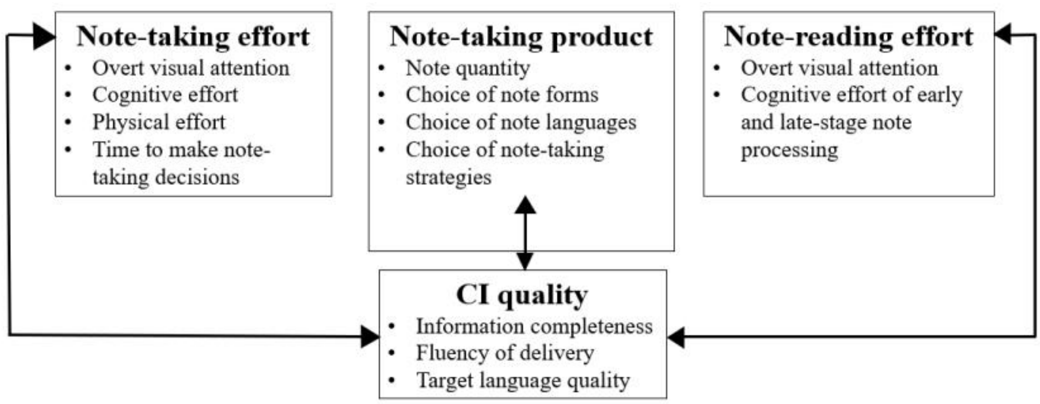


Figure 6-27. Tested correlations between note-taking behaviour and interpreting performance

[21] In total, 2 interpreter groups*2 task conditions* (6 indicators of the process of note-taking+11 indicators of the product of note-taking+6 indicators of the process of note-reading)=92 data normality tests are conducted.

6.6.1 Interpreting scores

After data normality tests (Tables 6-71 to 6-73), within-group and between-group comparisons are conducted to explore the interpreting scores in different interpreter groups and task conditions. Wilcoxon Signed Rank test results show that the students performed similarly across the easy and difficult segments of interpreting, whereas the professionals performed significantly better in the easy segments than in the difficult ones (Table 6-74). Moreover, Mann-Whitney tests show that the professionals had significantly better performance than the students in the easy segments in every aspect, but none of such significant differences are observed in the difficult ones (Table 6-75). In other words, when the speech segments were difficult, the professionals achieved a similar level of interpretation quality as the students.

Table 6-71. Shapiro-Wilk normal distribution test results for within-group comparisons of the interpreting scores (easy vs. difficult)

	Student (N=29)			**Professional (N=20)**		
Category	Statistic	*df*	Sig.	Statistic	*df*	Sig.
InfoCom	.944	29	<.05	.971	20	>.05
FluDel	.971	29	>.05	.912	20	<.05
TLQual	.955	29	<.05	.927	20	<.05
Total	.954	29	<.05	.978	20	>.05

Table 6-72. Shapiro-Wilk normal distribution test results for between-group comparisons of the interpreting scores in easy segments

	Student (N=29)			**Professional (N=20)**		
Category	Statistic	*df*	Sig.	Statistic	*df*	Sig.
InfoCom	.972	29	>.05	.960	20	>.05
FluDel	.966	29	>.05	.977	20	>.05
TLQual	.938	29	<.05	.944	20	<.05
Total	.990	29	>.05	.972	20	>.05

Table 6-73. Shapiro-Wilk normal distribution test results for between-group comparisons of the interpreting scores in difficult segments

	Student (N=29)			**Professional (N=20)**		
Category	Statistic	*df*	Sig.	Statistic	*df*	Sig.
InfoCom	.919	29	<.05	.950	20	>.05
FluDel	.960	29	>.05	.963	20	>.05
TLQual	.945	29	<.05	.930	20	<.05
Total	.945	29	<.05	.930	20	<.05

Table 6-74. The students and professionals' interpreting scores (easy vs. difficult)

	Students (N=29)					**Professionals (N=20)**				
	Easy		Difficult			Easy		Difficult		
	M	SD	M	SD	Sig.	M	SD	M	SD	Sig.
InfoCom	3.45	1.04	3.22	1.82	>.05	4.98	1.32	3.63	1.33	<.001

FluDel	4.92	1.06	4.71	1.08	>.05	5.63	1.24	5.03	1.32	<.01
TLQual	4.70	.82	4.55	.92	>.05	5.48	.96	4.85	1.08	<.01
Total score	16.52	3.64	15.71	4.01	>.05	21.06	4.65	17.13	4.70	<.001

Table 6-75. Between-group comparisons of interpreting scores (Mann-Whitney tests)

	Easy			Difficult		
	Student	Professional	Sig.	Student	Professional	Sig.
InfoCom		+	<.001		+	>.05
FluDel		+	<.01		+	>.05
TLQual		+	<.001		+	>.05
Total score		+	<.001		+	>.05

In addition, the two groups of participants' interpreting scores are examined on a segment basis to explore the sequence effects of source speech difficulty on interpreting performance. Wilcoxon Signed Rank tests (Tables 6-76 and 6-77) show that both student interpreters and professional interpreters performed significantly worse in Segment 2 than in Segment 1. When it comes to Segment 3 and Segment 4, only the professionals present significantly better results in the easy segment (Segment 4) than in the difficult one (Segment 3). The students, instead, performed similarly across the two segments of interpreting.

Table 6-76. The students' (N=29) interpreting scores in each segment of interpreting (Wilcoxon Signed Ranks)

	Segment 1		**Segment 2**			**Segment 3**		**Segment 4**		
	Difficulty-increase					*Difficulty-decrease*				
	M	SD	M	SD	Sig.	M	SD	M	SD	Sig.
InfoCom	3.59	.98	2.71	.85	<.001	3.74	1.25	3.31	1.10	>.05
FluDel	5.05	1.16	4.52	1.06	<.01	4.90	1.08	4.79	.95	>.05
TLQual	4.84	.88	4.19	.92	<.001	4.91	.77	4.56	.74	>.05
Total score	17.07	3.72	14.12	3.36	<.001	17.29	4.03	15.97	3.53	>.05

Table 6-77. The professionals' (N=20) interpreting scores in each segment of interpreting (Wilcoxon Signed Rank tests)

	Segment 1		**Segment 2**			**Segment 3**		**Segment 4**		
	Difficulty-increase					*Difficulty-decrease*				
	M	SD	M	SD	Sig.	M	SD	M	SD	Sig.
InfoCom	5.18	1.35	3.85	1.51	<.001	3.40	1.11	4.78	1.29	<.01
FluDel	5.78	1.36	5.35	1.35	<.01	4.70	1.24	5.48	1.13	<.05
TLQual	5.55	1.00	4.90	1.15	<.01	4.80	.86	5.40	.94	<.05
Total score	21.68	4.90	17.95	5.29	<.001	16.30	3.99	20.43	4.43	<.01

A further look at the tendencies of the interpreting scores (Figure 6-28 and Figure 6-29) reveals some potential sequence effects of source speech difficulty on the two groups of participants' interpreting performance. For instance, the students performed the worst in Segment 2 (difficult) in the difficulty-increase direction of interpreting, whereas the professionals performed the worst in Segment 3 (difficult) in the difficulty-decrease direction of interpreting. For the student group, when the interpreting task started with a difficult speech segment, they performed even worse in the easy part (Segment 4) than in the difficult one (Segment 3). By comparison, no such sequence effects are observed among the professionals. It appears that there were effects of source speech difficulty on both groups' of participants' interpreting performance, but the students were more obviously affected by the sequence of interpreting than the professionals.

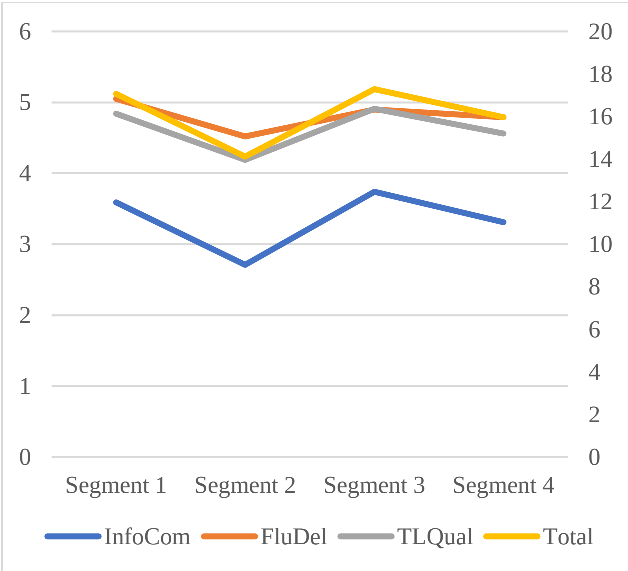


Figure 6-28. The students' (N=29) interpreting scores in each segment of interpreting

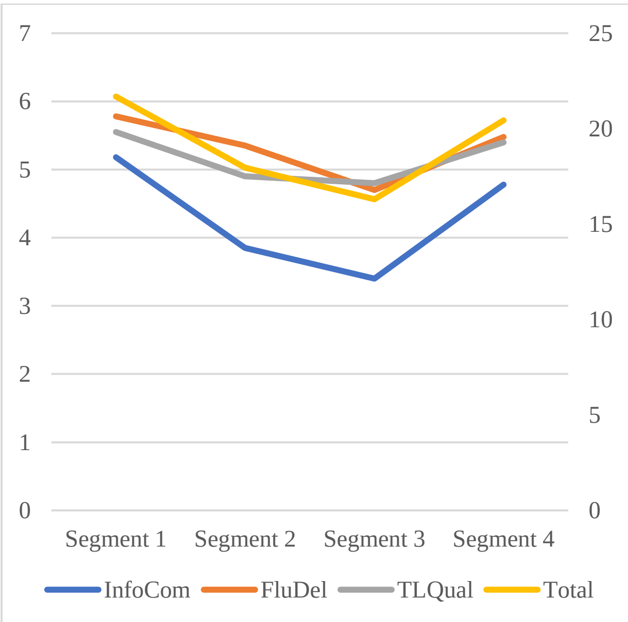


Figure 6-29. The professionals' (N=20) interpreting scores in each segment of

interpreting

6.6.2 Note-taking process and interpreting performance

As introduced in the previous sections, the process of note-taking is investigated with six eye-tracking and pen-recording measures: TFD_sum, TFD_average, CC_sum, CC_average, MFD, CC and EPS. Among these, TFD_sum and CC_sum are closely related to the number of notes. In general, the more notes were taken, the greater the TFD and CC on the notes. Therefore, these two measures point to the total potential cognitive and physical effort of note-taking. For the rest of the measures which were not affected by the number of notes, they are adopted to indicate the average effort of taking one note.

For the student group, in the easy segments of interpreting, only the indicators of the physical effort of note-writing present significant correlations with the interpreting performance (Table 6-78). Specifically, the sum of CC during note-writing is positively correlated with the fluency of delivery, whereas the average CC of note-writing shows negative correlations with the TL quality and the total score of interpreting. In the difficult segments of interpreting (Table 6-79), the average CC of note-writing again presents negative correlations with the InfoCom and FluDel scores. Except for that, the average overt visual attention (TFD_mean) that the students paid to each note is positively correlated with the TL quality.

Table 6-78. The correlations between the process of note-taking and the interpreting scores in the student group's (N=29) easy segments of interpreting

		InfoCom	**FluDel**	**TLQual**	**Total**
Overt visual attention	TFD_sum	.186	.204	.196	.210
	TFD_mean	-.127	-.140	-.087	-.133
Physical effort	CC_sum	.159	.322*	.161	.220
	CC_mean	-.258	-.199	-.261*#	-.260*
Cognitive effort	MFD	-.216	-.093	.485	-.182
Time to make note decisions	EPS	-.159	-.160	.231	-.188

Note. "#" means that the result is obtained through a Spearman correlation test rather than a Pearson correlation test.

Table 6-79. The correlations between the process of note-taking and the interpreting scores in the student group's (N=29) difficult segments of interpreting

		InfoCom	**FluDel**	**TLQual**	**Total**
Overt visual attention	TFD_sum	.292	.245	.240	.292
	TFD_mean	.139	.224	.297*	.210
Physical effort	CC_sum	.069	-.003	-.053	.028
	CC_mean	-.264*	-.285*#	-.141	-.248
Cognitive effort	MFD	.220	.239	.311	.265
Time to make note decisions	EPS	.097	.197	.235	.163

Note. "#" means that the result is obtained through a Spearman correlation test rather than a Pearson correlation test.

In terms of the professional group, regardless of source speech difficulty, there is no significant correlation between interpretation quality and the measures pointing to the total effort of note-taking (TFD_sum and CC_sum) (Table 6-80 and Table 6-81). By contrast, the average TFD on each note is negatively correlated with the InfoCom, FluDel and total scores in the easy segments, as well as the InfoCom score in the difficult ones. The average CC during note-writing also present negative correlations with the InfoCom and total scores in the easy segments, but no significant correlations are observed in the difficult ones. For the professionals' MFD on the notes, significant negative correlations are found with all the interpreting scores in both task conditions, suggesting a negative relationship between the cognitive effort of note-taking and the performance of interpreters. As for the EPS of note-planning, this presents a significant negative correlation with the FluDel score in the easy segments, but not in the difficult ones. Overall, compared with the difficult segments of interpreting, the easy segments of interpreting show more significant associations between the process of note-taking and the professionals' interpretation quality.

Table 6-80. The correlations between the process of note-taking and the interpreting scores in the professional group's (N=20) easy segments of interpreting

		InfoCom	**FluDel**	**TLQual**	**Total**
Overt visual attention	TFD_sum	-.100	-.222	-.046	-.126
	TFD_mean	-.352*	-.391*	-.298	-.366*
Physical effort	CC_sum	.062	.101	.145	.092
	CC_mean	-.328*	-.312	-.212	-.313*
Cognitive effort	MFD	-.357*	-.543**	-.357*	-.422**
Time to make note decisions	EPS	-.379	-.328*	-.259	-.353

Table 6-81. The correlations between the process of note-taking and the interpreting scores in the professional group's (N=20) difficult segments of interpreting

		InfoCom	**FluDel**	**TLQual**	**Total**
Overt visual attention	TFD_sum	-.206	-.197	-.159	-.206
	TFD_mean	-.325*	-.243	-.180	-.291
Physical effort	CC_sum	.079	-.006	.027	.049
	CC_mean	-.099	-.101	-.006	-.086
Cognitive effort	MFD	-.320*	-.345*	-.293*	-.341*
Time to make note decisions	EPS	-.056	.103	-.107	-.026

A common feature of the students' and the professionals' correlation test results is that, for the significant results, most of the coefficients range from .20 to .40, indicating a weak level of association between the process of note-taking and interpretation quality.

Overall, compared with the students, the professionals show many more significant correlations between the process of note-taking and interpreting performance. As summarized in Table 6-82, the student group presents a few positive and negative correlations between the process of note-taking and the interpretation quality, whereas the professional group only shows negative correlations between the two. These results indicate that, for the students, the increase in the certain aspects of the note-taking effort might have benefitted their interpreting performance, especially when the source speech was difficult. However, for the professionals, more note-taking effort was usually associated with worse interpretation quality, and this negative relationship is especially obvious in the easy segments of interpreting. Moreover, many of the students' correlations results concern the FluDel and the TLQual scores but not the InfoCom scores, suggesting that the note-taking effort was more associated with the delivery of the target speech but not the accuracy of interpreting. By comparison, many of those of the professionals are related to their InfoCom and FluDel scores rather than the TLQual scores, indicating that the professionals' TL quality is not associated with their note-taking effort. One thing in common between the two groups of participants is that the average physical effort of note-writing is negatively correlated with their interpretation quality, examining the potential harm that a heavy physical load of handwriting could exert on the interpreting performance.

Table 6-82. Significant associations observed between the process of note-taking and interpretation quality

		Students		**Professionals**	
Aspects		Easy	Difficult	Easy	Difficult
Overt visual attention	Sum				
	Mean		TLQual+	InfoCom, FluDel & Total-	InfoCom-
Physical effort	Sum	FluDel+			
	Mean	TLQual &Total-	InfoCom & FluDel-	InfoCom & Total-	
Cognitive effort				All aspects-	All aspects-
Time to make note decisions				FluDel-	

Note. "+" represents a positive relationship and "-" means a negative relationship.

6.6.3 Note-taking product and interpreting performance

The associations between interpretation quality and the product of note-taking are assessed from four aspects: note quantity, note form, note language and note-taking strategy. Since note quantity varied across different interpreter groups and task conditions, the choices of note forms, note languages and note-taking strategies are investigated in percentages.

As shown in Table 6-83, the students' note quantity shows significant weak positive correlations with every aspect of interpretation quality in the easy segments of interpreting, but none of these significant correlations are found in the difficult ones (Table 6-84). In terms of note forms, no significant correlations are witnessed in the easy segments, but various correlations are observed in the difficult ones. For instance, the percentage of symbols is positively and weakly correlated with InfoCom scores, and the percentage of full words is negatively correlated with the InfoCom and total scores at a weak level. As for note language, the use of Chinese notes doesn't associate with interpretation quality significantly. Only the use of English notes is negatively correlated with the interpreting scores in the difficult segments of interpreting. Lastly, in terms of the adoption of note-taking strategies, the use of replacement strategy during note-taking is positively correlated with the students' FluDel and total scores in the easy segments of interpreting at a weak level.

Table 6-83. The correlations between the product of note-taking and the interpreting scores in the student group's (N=29) easy segments of interpreting

		InfoCom	**FluDel**	**TLQual**	**Total**
Note quantity	Number of notes	.267*	.410**	.287**	.337**
Note form	Language notes	-.056	-.115	-.170	-.104
	Symbol	.097	.121	.193	.134
	Full word	-.116	-.129	-.179	-.144
	Abbreviation	.114	.063	.074	.100
Note language	Chinese	.046	-.024	-.014	.016
	English	-.0118	-.086	-.0155	-.0128
Note-taking strategy	Ellipsis	-.024	-.084	.062	-.025
	Replacement	.095	.356**#	-.003	.278*#
	Restructuring	.164	.062	.115	.138

Note. "#" means that the result is obtained through a Spearman correlation test rather than a Pearson correlation test.

Table 6-84. The correlations between the process of note-taking and the interpreting scores in the student group's (N=29) difficult segments of interpreting

		InfoCom	**FluDel**	**TLQual**	**Total**
Note quantity	Number of notes	.231	.136	.036	.181
Note form	Language notes	-.243	-.167	-.132	-.218
	Symbol	.280*	.139	.095	.224
	Full word	-.319*	-.189	-.122	-.266*
	Abbreviation	.116	.043	.001	.080
Note language	Chinese	.156	.189	.138	.174
	English	-.413**	-.384**	-.289**	-.412**
Note-taking strategy	Ellipsis	-.0112	-.008	-.110	-.094
	Replacement	-.008	.055	.029	.016
	Restructuring	.238	.153	.139	.213

Same as the students, in the easy segments of interpreting, the professionals present positive correlations between note quantity and interpretation quality in all aspects (Table 6-85). Except for that, no other significant correlations are observed between the product of note-taking and the interpreting scores in either easy or difficult segments of interpreting (Tables 6-85 and 86). In other words, for the professional group, the choice of the note forms, note languages and note-taking strategies is not associated with their interpreting performance at all.

Table 6-85. The correlations between the process of note-taking and the interpreting scores in the professional group's (N=20) easy segments of interpreting

		InfoCom	**FluDel**	**TLQual**	**Total**
Note quantity	Number of notes	.426**	.390*	.369*	.423**
Note form	Language notes	-.164	-.304	-.132	-.202
	Symbol	.173	.273	.132	.199
	Full word	-.307	-.258	-.199	-.284
	Abbreviation	.367	.136	.216	.319
Note language	Chinese	.171	.063	.117	.138
	English	-.242	-.0281	-.179	-.249
Note-taking strategy	Ellipsis	-.092	-.157	-.157	-.127
	Replacement	.017	-.029	.088	.020
	Restructuring	.073	.113	.093	.091

Table 6-86. The correlations between the process of note-taking and the interpreting scores in the professional group's (N=20) difficult segments of interpreting

		InfoCom	**FluDel**	**TLQual**	**Total**
Note quantity	Number of notes	.242	.171	.060	.209
Note form	Language notes	.050	-.039	.184	.057
	Symbol	.106	.090	-.019	.081
	Full word	-.009	-.057	.061	-.008
	Abbreviation	.066	.048	.116	.076
Note language	Chinese	-.015	.171	-.007	.038
	English	.041	-.138	.134	.013
Note-taking strategy	Ellipsis	.182	.070	.233	.173
	Replacement	.147	.242	.054	.163
	Restructuring	.115	.051	.078	.096

Results concerning the associations between the product of note-taking and interpretation quality are summarized in Table 6-87. Both groups of participants present significant positive correlations between note quantity and the interpreting scores in all aspects, but such correlations are only observed in the easy segments of interpreting. In terms of note forms, note languages and note-taking strategies, only the students present various significant correlations with the assessed aspects of interpretation quality.

Table 6-87. Significant associations observed between the product of note-taking and interpretation quality

		Students		Professionals	
Aspects	Category	Easy	Difficult	Easy	Difficult
Quantity	Note count	All aspects +		All aspects +	
Form	Language				
	Symbol		InfoCom +		
	Full-word		InfoCom & Total -		
	Abbreviation				
Language	Chinese				
	English		All aspects -		
Strategy	Ellipsis				
	Replacement	FluDel & Total +			
	Restructuring				

Note. "+" represents a positive relationship and "-" means a negative relationship.

6.6.4 Note-reading process and interpreting performance

There are two types of measures that point to the process of note-reading, with one indicating the overt visual attention that was paid to the note(s) (the sum and average TFD on the notes) and the other indicating the cognitive effort of note-reading in both early and late stages of note-processing (MFD, RVC, FFD, FPD and SPD). This section presents the correlation test results between the effort of note-reading and the quality of interpretation.

The two groups of participants show great differences in the correlation test results. In the easy segments of interpreting, the student group presents few significant correlations between the note-reading effort and the interpreting scores (Table 6-88), but the professional group shows many significant negative correlations between the two (Table 6-89). Specifically, the students' InfoCom score is negatively correlated with the overt visual attention they paid to each note (TFD_mean) and the cognitive effort of late-stage processing during note-reading (SPD). Their FluDel score presents a negative relationship with the overall cognitive effort of note-taking (MFD). In the professional group, consistent negative associations are observed between almost every aspect of interpretation quality, on the one hand, and the overt visual attention they paid to the notes (TFD_sum and TFD_mean), the overall cognitive effort of note-reading (MFD) and the specific cognitive effort of note recognition (FFD) and late-stage note-processing (SPD), on the other hand.

Table 6-88. The correlations between the process of note-reading and the interpreting scores in the student group's (N=29) easy segments of interpreting

Aspects	Sub-category		InfoCom	FluDel	TLQual	Total
	On noted areas	TFD_sum	.092	.204	.196	.210

Overt visual attention	On each note	TFD_mean	-.278*#	-.222	-.152	-.155
Cognitive effort	Overall	MFD	-.188	-.270*	-.158	-.202
	Incomplete processing	RVC	.011	-.193	-.063	-.035
	Note recognition	FFD	-.112	-.146	-.088	-.126
	Lexical-level integration	FPD	.008	-.134	.048	-.024
	Syntactic-and-textual level integration	SPD	-.291*#	-.217	-.176	-.166

Note. "#" means that the result is obtained through a Spearman correlation test rather than a Pearson correlation test.

Table 6-89. The correlations between the process of note-reading and the interpreting scores in the professional group's (N=20) easy segments of interpreting

Aspects	Sub-category		InfoCom	FluDel	TLQual	Total
Overt visual attention	On noted areas	TFD_sum	-.394*	-.447**	-.346*	-.415**
	On each note	TFD_mean	-.532**	-.545**	-.495**	-.550**
Cognitive effort	Overall	MFD	-.265	-.434**	-.326*	-.334*
	Incomplete processing	RVC	-.469	-.493	-.408	-.483
	Note recognition	FFD	-.126	-.318*	-.253	-.209
	Lexical-level integration	FPD	-.222	-.214	-.222	-.229
	Syntactic-and-textual level integration	SPD	-.549**	-.566**	-.508**	-.568**

In the difficult segments of interpreting, the students show three positive correlations between their interpretation quality and the cognitive effort of incomplete processing, as well as two negative correlations between their interpreting scores and the cognitive effort of note recognition and late-stage note processing (Table 6-90). The professionals only present significant negative correlations (Table 6-91). For instance, the average overt visual attention they paid to each note (TFD_mean) is negatively correlated with their InfoCom, FluDel, and total interpreting scores. The cognitive effort they expended on integrating the meaning of notes at a lexical level (FPD) is negatively correlated with their total interpreting scores; and the cognitive effort for integrating the meaning of notes at a syntactic and textual level (SPD) is negatively correlated with their FluDel and total interpreting scores.

Table 6-90. The correlations between the process of note-reading and the interpreting scores in the student group's (N=29) difficult segments of interpreting

Aspects	Sub-category		InfoCom	FluDel	TLQual	Total
Overt visual attention	On noted areas	TFD_sum	.111	.035	.040	.084
	On each note	TFD_mean	.066	.131	.194	.118
Cognitive effort	Overall	MFD	-.092	-.022	-.018	-.065
	Incomplete processing	RVC	.182	.280*	.354*	.263*
	Note recognition	FFD	-.289*	-.253	-.0128	-.267*
	Lexical-level integration	FPD	-.115	-.119	-.119	-.127
	Syntactic-and-textual level integration	SPD	-.291*#	.163	.230	.150

Note. "#" means that the result is obtained through a Spearman correlation test rather than a Pearson correlation test.

Table 6-91. The correlations between the process of note-reading and the interpreting scores in the professional group's (N=20) difficult segments of interpreting

Aspects	**Sub-category**		**InfoCom**	**FluDel**	**TLQual**	**Total**
Overt visual attention	On noted areas	TFD_sum	-.260	-.276	-.262	-.281
	On each note	TFD_mean	-.322*	-.327*	-.270	-.332*
Cognitive effort	Overall	MFD	-.236	-.210	-.216	-.239
	Incomplete processing	RVC	-.308	-.285	-.280	-.315
	Note recognition	FFD	-.217	-.193	-.235	-.315#
	Lexical-level integration	FPD	-.160	-.318	-.130	-.316*#
	Syntactic-and-textual level integration	SPD	-.301#	-.339*	-.276	-.308*#

Note. "#" means that the result is obtained through a Spearman correlation test rather than a Pearson correlation test.

In summary, compared with the students, the professionals presented many more significant correlations between their note-reading effort and interpreting performance (Table 6-92). Moreover, compared with the student group which shows some positive correlations in the tests, the professional group only presents negative correlations, suggesting that an increase in the cognitive effort of note-reading is associated with worse interpretation quality among the professionals. In addition, the significant correlations observed in the professional group's correlations cover almost all the assessed aspects of interpretation quality. From this perspective, the note-reading process presents a closer tie with the professionals' performance than that of the students. All in all, although all the observed correlations are at a weak level, many significant associations are observed between the note-reading effort and the interpreting performance. Such associations are more commonly found in the professional group than in the student one.

Table 6-92. Significant associations observed between the process of note-reading and interpretation quality

		Students		**Professionals**	
Aspects		Easy	Difficult	Easy	Difficult
Overt visual attention	Sum			All aspects -	
	Mean	InfoCom-		All aspects -	InfoCom, FluDel & Total -
Cognitive effort		FluDel -		FluDel, TLQual & Total -	
Incomplete processing			FluDel, TLQual & Total +		
Early processing			InfoCom & Total-	FluDel -	Total -
Late processing		InfoCom -	InfoCom -	All aspects -	FluDel & Total -

Note. "+" represents a positive relationship and "-" means a negative relationship.

6.6.5 Summary of RQ5

Results in this section, which are summarized in Tables 6-93 and 6-94, present four clear patterns. Firstly, although a variety of significant correlations are found between interpreters' note-taking behaviour and CI quality, most of the correlation coefficients are between 0.2 and 0.4, representing weak to medium levels of the correlations. Secondly, in terms of the product of note-taking, only the students present various correlations between the interpretation quality and their note choices, whereas the professionals show no significant correlations at all. Thirdly, as for the process of note-taking and note-reading, a greater number and a stronger level of correlations are observed in the professional group than in the student one, indicating that the professionals' interpreting performance is more closely associated with the note-taking behaviour than the students. Fourthly, the significant correlations are usually more frequent and stronger in the InfoCom and FlueDel scores than in the TLQual score.

Table 6-93. Significant correlations observed in the students' note-taking behaviour and interpreting performance

Stages	Aspects		Easy	Difficult
Note-taking process	Visual	Mean		TLQual+
	Cognitive			
	Temporal			
	Physical	Sum	FluDel+	
		Mean	TLQual & Total-	InfoCom & FluDel-
Note-taking product	Quantity		All aspects +	
	Symbol			InfoCom+
	Full-word			InfoCom & Total-
	English			All aspects-
	Replacement		FluDel & Total +	
Note-reading process	Visual	Mean	InfoCom-	
	Cognitive	Overall	FluDel-	
		Incomplete processing		FluDel, TLQual & Total+
		Early-stage		InfoCom & Total-
		Late-stage	InfoCom-	InfoCom-

Note. "+" represents a positive relationship and "-" means a negative relationship.

Table 6-94. Significant correlations observed in the professionals' note-taking behaviour and interpreting performance

Stages	Aspects		Easy	Difficult
Note-taking process	Visual	Mean	InfoCom, FluDel & Total-	InfoCom-
	Cognitive		All aspects-	All aspects-
	Temporal		FluDel-	
	Physical	Mean	InfoCom & Total-	

Note-taking product	Quantity Symbol Full-word English Replacement		All aspects+	
Note-reading process	Visual	Sum	All aspects-	
		Mean	All aspects-	InfoCom, FluDel & Total-
	Cognitive	Overall	FluDel, TLQual & Total-	
		Incomplete processing		
		Early-stage	FluDel-	Total-
		Late-stage	All aspects-	FluDel & Total-

Note. "+" represents a positive relationship and "-" means a negative relationship.

In summary, H9 and H11 are partially supported by the observed positive correlations between the students' note-taking effort and interpretation quality as well as the negative correlations between both groups' note-reading effort and interpreting performance. However, the collected data actually reveal much more complicated relationships between the two types of noting effort and the three aspects of interpretation quality in different participant groups and task conditions. This indicates that the observed significant associations need to be cautiously interpreted on the basis of taking the specific interpreter type and interpreting task into consideration. In terms of H10, on the one hand, it is confirmed by the positive correlations observed between interpretation quality and note quantity as well as the students' use of symbols and the SL in notes; on the other hand, it is rejected by the professionals' data where no significant correlations are found between note choices and interpretation quality at all. Therefore, all the three hypotheses are only partially supported by the data.

Chapter 7: Discussion

In the present study, the process of note-taking, the product of note-taking and the process of note-reading have been visualized and quantified with an eye-and-pen approach. The effects of interpreter work experience and source speech difficulty have been examined whenever possible to reveal the underlying cognitive mechanism of interpreters' note-taking behaviour in VRI. This chapter provides a thorough discussion of all the results obtained from the reported note-taking experiment with two groups of participants under two conditions of task difficulty. Section 7.1 focuses on the experience and difficulty effects on the process and product of note-taking. Section 7.2 is devoted to explaining the special reading patterns and cognitive demands observed in the note-reading process. Section 7.3 probes into the complex relationship between the effort of note production and the effort of note reception. Finally, Section 7.4 discusses the various associations between the participants' note-taking behaviour and interpretation quality. All of the discussion is conducted in the context of the research aims and research questions proposed in Chapter 1.

7.1 Experience and difficulty effects on the process and product of note-taking

7.1.1 Overt visual attention and physical effort

Contrary to the hypotheses, regardless of source speech difficulty, the professionals allocated a greater amount of overt visual attention and physical effort to conduct note-taking than the students; while, despite interpreter work experience, all the participants always distributed more overt visual attention and physical effort to complete note-taking in the easy segments than in the difficult ones. These observed experience and difficulty effects on the note-taking effort of the participants could be explained by referring to the participants' choice of notes, experience with VRI, and cognitive load of interpreting.

Firstly, in addition to the significant difference in the students' and the professionals' preference for full words and abbreviations during note-taking, the students and the professionals also presented a significant difference in their preference for language notes and symbols when the proportions of these note types were calculated on average across the four interpreting segments. As Table 7-1 presents, the students preferred symbols and abbreviations, whereas the professionals used many language notes and full words. It is found in the present study that, for Chinese-native interpreters, symbols entail significantly less note-taking effort than language notes in every measured aspect (i.e., visual, cognitive, physical and temporal). Between full words and abbreviations, the former demands more overt visual attention and physical effort during note-taking, while the latter entails greater cognitive effort during note-taking and longer time in note-planning. Combining these results, they suggest that the students took many of their notes in the forms that require less visual processing and handwriting, although they sometimes needed to bear more cognitive load during note-taking and spent longer time on note-planning. By comparison, the professionals did not prefer to abbreviate or transform the words in the source speech into symbols. They simply noted in language, mainly full words, regardless of a greater demand for note production and note reception. As a result, the students looked less at the virtual notepad on the screen during note-taking and invested less physical effort in note-writing than the professionals. This is also in accordance with the finding that professionals took more notes than the students in all interpreting segments (see Table 7-1), despite that this latter difference was not significant.

Table 7-1. Mann-Whitney test results on the two groups of participants' note choices

		Student		**Professional**		
Aspect	Category	M	SD	M	SD	Sig.
Note quantity	Number of notes	86.37	29.45	93.61	27.59	>.05
Note form	Language	53.29	13.39	59.88	13.28	<.01
	Symbol	35.40	11.10	29.89	12.27	<.01
	Number	4.00	3.69	3.96	3.67	>.05
	Full word	21.67	12.82	33.61	18.66	<.001
	Abbreviation	31.62	9.19	26.27	11.67	<.001
Note language	Chinese	21.58	13.86	22.97	11.74	>.05
	English	31.81	11.08	36.91	17.88	>.05

Secondly, the observed experience effects on the process of note-taking could also be attributed to the professionals' rich experience in VRI at work. It is found in the present study that, during the input stage of VRI, compared with the students, the professionals always paid more overt visual attention to the laptop screen and distributed a greater proportion of their overt visual attention to the noted area[22]. Since the two groups had similar working memory capacity and reported a similar amount of mental demand and frustration after interpreting, this professional-novice difference in cognitive resource allocation can suggest two things. Firstly, when facing interpreting

[22] Fixations on the 'noted area' include those on the blank space among notes, whereas fixations on 'the notes' only include those on the individual notes. For fixations on the 'notes', the two groups also present the same difference, with the professionals (*M*=54.33%, *SD*=11.92%) showing a significantly ($t(191.69)=-3.27$, $p<.01$) higher percentage of fixations on the notes than the students (*M*=47.91%, *SD*=15.51%).

difficulties, the professionals would allocate a larger amount of their processing capacity to address the difficulties than the students. Secondly, in terms of the allocated capacity, the professionals would distribute a larger proportion of this to conduct note-taking than the students. These findings appear to be contrary to the situation in on-site interpreting (e.g., Chmiel & Lijewska, 2019 on sight translation; Stachowiak-Szymczak & Korpal, 2019 on simultaneous interpreting; and Hu, 2008 on note-taking), where professional interpreters have been usually found to exert less cognitive effort during interpreting than student interpreters. One important reason could be the great difference between on-site and remote interpreting. It has been found that RI is less friendly to interpreters than on-site interpreting, because of its uncomfortable physical environment, poor ergonomics, high stress and quick burnout (Roziner & Shlesinger, 2010). However, professional interpreters could maintain their interpreting quality at a similar level across the two work conditions even though they themselves judge their performance as significantly poorer in the remote mode (Roziner & Shlesinger, 2010). According to Dong and Liu (2016), the cognitive advantage in interpreting is achieved through placing high cognitive demand on interpreters, even more than is needed, to push them to develop corresponding coping strategies. In the present study, the professionals had been adjusting themselves to working remotely during the COVID-19 pandemic and faced such real stress from clients and audience more often than the students, whereas the students who attended online classes within a relatively relaxing environment may have not developed coping strategies for RI. This could be the reason why the professionals were able to allocate more overt visual attention and physical effort to conducting note-taking than the students.

Thirdly, regardless of interpreting experience, the participants allocated more overt visual attention and physical effort to take notes for the easy segments than for the difficult ones, resulting in significant differences in note quantity and interpreting quality. The same effects of difficulty are found when comparing the two groups' fixations on the whole screen. These latter results could be attributed to the possible cognitive load of interpreting in the difficult segments. In that situation, increasing invested effort would not help interpreters to enhance interpreting quality (Phase B in Figure 7-1). Studies using pupil dilation to indicate participants' cognitive effort have also found that pupil dilation could drop below the baseline when the demand for information processing exceeds one's capacity (Poock, 1973). After conducting a Pearson correlation test between the interpreters' cognitive effort during the input phase of CI (MFD on the screen) and their interpreting scores, a negative correlation (Rho=-0.226, $p<.05$) is found between the two in the difficult segments. On the other hand, no such significant correlation is observed in the easy ones. This result confirms that the interpreters experienced cognitive overload while receiving the difficult segments as more effort devotion linked with undermined interpreting quality (Phase B in Figure 7-1). These findings suggest that interpreters' poor performance in VRI could be traced back to the note-taking process where cognitive resources are inadequate to conduct listening comprehension, video watching and note-taking at the same time.

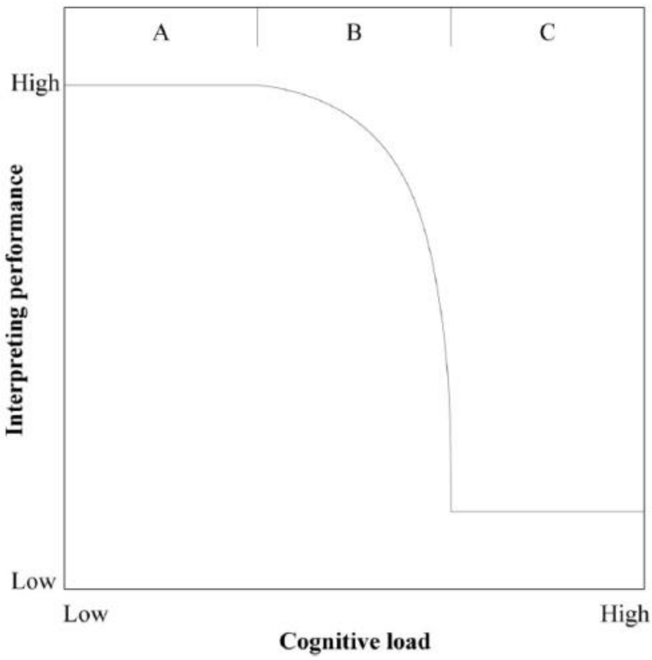


Figure 7-1. A hypothetical relationship between the cognitive load of interpreting and interpreting performance (S. Chen, 2017c, p. 650)

7.1.2 Cognitive effort

Overall, data concerning the cognitive effort (MFD) of note-taking present the same experience and difficulty effects as reported in Section 7.1.1. The easy segments of interpreting are frequently found to have longer MFD on the notes than the difficult ones. Meanwhile, the professional group usually presents longer MFD on the notes than the student one. However, since none of the observed differences reach a significant level, this means that there are no remarkable effects of interpreter work experience or source speech difficulty on the cognitive effort of note-taking. This result is not in line with those of Andres (2002) and Hu (2008) who found a lighter cognitive load of note-taking among professional interpreters than among student interpreters. Nonetheless, Andres (2002) drew her conclusion based on the EPS data, which in the present study were not used as an indicator of load but as a strategic choice of time distribution during note-taking. Hu (2008) asked interpreters to complete a secondary task during note-taking, which might misguide the interpreters to focus on secondary-task completion rather than on note-taking. Hence, with eye-tracking data obtained in easy and difficult task conditions, the present study concludes there are no significant experience or difficulty effects on the cognitive effort of note-taking.

The results of no significant experience or difficulty effects on the cognitive effort of note-taking are in accordance with the participants' note-taking product, which was dominated by an effort-saving strategy of note-taking: ellipsis. For both student and professional interpreters, they present a clear preference for the ellipsis strategy over non-ellipsis strategies in both easy and difficult segments of interpreting, corroborating Albl-Mikasa's (2006) finding with student interpreters and that of S. Chen (2022) with professional interpreters. In the meantime, the participants' preferences for different note forms and note languages also remained the same in different task conditions. With such a persistent preference for the ellipsis strategy and the preferred note categories, the participants' cognitive effort in note-taking does not show significant differences across different interpreter types and task conditions. Tracing back to the root cause of this inflexibility in the adoption of note-taking strategies, this cause could be the high cognitive demand for interpreting tasks. As Gile points out in the Tightrope Hypothesis, during the input phase of CI, "any increase in processing capacity requirements linked to source-speech features and any error in the way they manage available capacity … is likely to lead to saturation-based errors" (2001, p. 9). From this perspective, shifting among different note-taking strategies during the process of interpreting might increase the cognitive load of interpreters who are already working close to saturation. Therefore, instead of selecting and applying different note-taking strategies for source speech at variant difficulty levels, the participants chose to stay safe with the simple ellipsis strategy, which thus entailed a similar amount of cognitive effort of note-taking in different task conditions and interpreter groups.

In addition, the unique demand for information retention while adopting the ellipsis strategy could also lead to a saturated WM among the participants, which is a further possible cause for there being no experience or difficulty effects on the cognitive effort of note-taking. As discussed by Albl-Mikasa (2006), although the ellipsis strategy does not demand a deep level of language processing during the input stage of CI, it has high requirements for information retention during note-writing, especially when interpreters intend to pursue large note quantities. Specifically, the participants have to constantly refresh what they intend to write in their WM until the writing is finished (Piolat et al., 2005). However, humans' writing speed is only one-tenth of their speaking rate (Foulin, 1995), and the input phase of CI is speaker-paced. This means that, if the participants aim to store a great amount of source information in notes by applying an ellipsis strategy of note-taking, they have to significantly accelerate the SL processing, enhance their handwriting speed, rehearse what they intend to write in mind, and frequently update the source information in their WM. Compared with the difficult segments of interpreting which contained more complicated expressions and dense information, the participants could complete these language processing and note-planning procedures more quickly than in the easy ones. Therefore, they manage to produce more notes in the easy segments than in the difficult ones. However, at the same time, their WM could be saturated with the demand for subvocal rehearsal during note-writing, leading to no effects of source speech difficulty on the cognitive effort of note-taking. Similarly, compared with the students, the professionals could facilitate these procedures more easily and quickly, thus resulting in bigger note quantities and higher interpretation quality. Nevertheless, the professionals' WM could be saturated with the demand for information retention, thus resulting in no effects of interpreter work experience on the cognitive effort of note-taking.

It is worth mentioning that the result of no preference change in the note choices in the present study is not in line with Cardoen (2018) which is the only study that tested interpreters' note-taking preferences in different task conditions. Cardoen (2018) found that, as source speech difficulty increases, all the participants (five first-year masters' students, five second-year, master's students, and five professional interpreters and interpreter trainers) increased their note quantities and their use of abbreviations and symbols. One important reason for the inconsistency of the two sets of results could be the two studies' difference in task difficulty control and experimental design. In Cardoen (2018), task difficulty was induced by different speech durations and different levels of lexical and syntactical complexity; Part 2 of the source speech, which was designed to be more difficult to interpret than Part 1, had a longer speaking duration (382 s vs. 270 s), a greater syllable count (1,235 vs. 931), a more frequent (17 vs. 1), and diverse (years, units, percentages, etc.) use of numbers, a higher frequency of names (such as Ross Garnaut) and two long enumerations (e.g., residential buildings, commercial and industrial buildings and road and rail infrastructure, soil evaporation). It has been suggested repeatedly in note-taking guidelines that abbreviations and symbols should be adopted for numbers, proper names, parallel structures and long clauses in the source speech because of high information density and low predictability (e.g., Gillies, 2017). More importantly, the participants always interpreted the long and difficult Part 2 after the short and easy Part 1. All of these speech features suggest that the participants could face fatigue issues in Part 2 where a series of lexical and syntactical processing difficulties were designed. This could slow down SL processing and postpone note-taking, as found in the EPS data of the present study, leading to more frequent use of easy-writing note forms such as symbols and abbreviations to make up for the time loss in SL processing. Moreover, as they have already interpreted Part 1 and familiarised themselves with the speech topic, it would be easier for them to come up with abbreviations and symbols in Part 2 than in Part 1. Since Cardoen (2018) does not probe into the process of note-taking and the reports on the retrospective interview did not mention the participants' reflections on their note choices, further research can be conducted in this regard with source materials differing in lexical and syntactical complexity and with process-oriented research methods such as video-recording, eye-tracking, and pen-recording. All in all, in the present study, the participants did not change their note preferences in different task conditions, which contributed to there being no effects of source speech difficulty on the cognitive effort of note-taking.

7.1.3 Time spent on note-planning

In the present study, effects of source speech difficulty, but not effects of interpreter work experience, are found in the EPS data. Specifically, both groups of participants spent significantly longer time on note-planning in the difficult segments than in the easy ones. This is in accordance with the result on the readability indices of the speech segments. However, whether the prolonged EPS represents an increased cognitive load of note-taking (as interpreted in S. Chen, 2017b) needs to be further validated with other reliable measures of cognitive load. As Andres (2002) discovered from her EPS data, the participants usually present listening comprehension problems only when the EPS is longer than seven seconds. In the present study, MFD on the notes, which is adopted to indicate the cognitive effort of note-taking, is found to have no significant change across easy and difficult task conditions, presenting different results from the EPS data.

This means that EPS in this part of the investigation only points to the time distribution between note-planning and note-writing during the process of note-taking, and should not be over-interpreted from a cognitive load perspective. All in all, the present study finds that, regardless of interpreter work experience, the participants needed more time to plan for note-taking in difficult segments of interpreting than in the easy ones.

Secondly, the result that no significant group difference is observed in the EPS data of the present study is seemingly not in line with that in S. Chen (2022) who detected a significantly shorter EPS among professionals than among students in English-to-Chinese CI. However, both studies reveal the professionals' advantage in regulating EPS during the process of note-taking. S. Chen proved the professionals' expertise in this regard through findings comparing the average EPS of the two groups of participants. The present study found that the professionals were able to produce larger note quantities with a similar EPS as that of the students while using many full words in their note-taking. It is found in the present study that full words demand significantly more overt visual attention during note-taking and physical effort of note-writing than abbreviations. This means that the participants have to devote more effort to complete visual processing and handwriting while producing full words than creating abbreviations, which could postpone the composition of the next note, cause a long interval between source speech delivery and note-taking, and result in a small number of notes. Nonetheless, the professionals did not present significantly longer EPS than the students. They even produced larger quantities of notes and better interpretation quality than the students in every segment of interpreting. This indicates that the professionals successfully maintained their EPS within an acceptable range while ensuring an adequate number of notes and a high-level interpreting performance.

In summary, the EPS data in this part of the investigation only indicate the time spent on note-planning rather than the cognitive effort of note-taking. Effects of source speech difficulty are observed in the EPS data because the participants needed more time to process the source information in the difficult segments than the source information in the easy ones; while effects of interpreter work experience are not directly detected in the average EPS of the two groups of participants but are indirectly reflected in the professionals' advantage in EPS regulation while creating a large note quantity with frequent use of full words.

7.1.4 Sequence effects of source speech difficulty

For all the measures concerning the note-taking effort, significant differences are often observed between Segment 1 and Segment 2 in the difficulty-increase direction of interpreting but not between Segment 3 and Segment 4 in the difficulty-decrease direction of interpreting. Specifically, the participants decreased the total and average effort they expended on note-taking in Segment 2 (difficult) from Segment 1 (easy), but their note-taking effort turns out to be similar between Segment 3 (difficult) and Segment 4 (easy). This result can be explained by referring to a concept called sequential dependency, introduced by studies on human behaviour (e.g., Wolfe et al., 2003 on visual search; Bock & Griffin, 2000 on semantic judgment; Rogers & Monshell, 1995 on task switching). According to Mozer et al. (2007, p. 180), the essence of a sequential dependency is "an influence of one incidental experience on subsequent experience". For instance, when the previous task is easy, humans intend to perceive the current task as easy; and when the previous one is difficult, even if the current task is easy, humans would perceive the present task as difficult (see explanations in Table 7-2). Such a distorted perception of task difficulty can exert impacts on humans' cognitive control, which finally affects their performance in task completion. According to the participants' NASA-TLX scores, they rated Segment 2 to be significantly more difficult to interpret than Segment 1, but such gaps obviously narrowed when it came to Segment 3 and Segment 4. The students even performed worse in interpreting the last segment than in interpreting the previous one. All of these results indicate that their perception of task difficulty in Segment 4 was greatly affected by their experience in interpreting Segment 3. With such an altered perception of task difficulty, the participants devoted a similar amount of note-taking effort to complete interpreting Segments 3 and 4.

Table 7-2. Explanations for the sequential effects of human behaviour based on Mozer et al. (2007)

	Previous task (easy)	**Previous task (difficult)**
Current task (easy)	Perceive the current task as easy, react faster, and perform better	Perceive the current task as difficult, react slower, and perform worse
Current task (difficult)	Perceive the current task as easy, react faster, but perform worse	Perceive the current task as difficult, react slower, and perform worse

However, following the sequential effects presented in Table 7-2, the participants should perceive Segment 2 as being as easy as Segment 1. By contrast, such sequence effects of source speech difficulty are not observed between these two segments in the difficulty-increase direction of interpreting. One potential reason for this result could be related to test anxiety. According to H. Chen (2012, p. 328), “test anxiety will be most disruptive when a test is initially perceived as highly difficult”. This means that, when a test starts with difficult tasks or items, examinees would experience especially high test anxiety which results in undermined task performance. It is also found in class that students experience a higher anxiety level when the tested items are arranged in a difficult-to-easy manner than in the other direction (Tippets & Benson, 1989). From this perspective, a difficulty-decrease direction of interpreting can exert greater impacts on task difficulty perception than its reserve-direction counterpart. This is in line with what has been found about the participants’ self-reported *Frustration* levels while interpreting the four segments. However, since the connection between test anxiety and self-reported frustration cannot be established directly and arbitrarily, more comprehensive measurements of test anxiety can be included in future studies to examine the sequential effects of source speech difficulty on interpreters’ note-taking behaviour.

In addition, although both groups of participants present sequence effects of note-taking effort, the professionals appear to have been more resilient to such effects than the students. Firstly, despite that the professionals also present narrower score differences in the NASA-TLX ratings for Segments 3 and 4 than for Segments 1 and 2, as did the students, they did detect a marginally-significant difference in the *Mental demand* of the last two interpreting segments. Secondly, despite the decreased difficulty level of interpreting from Segment 3 to Segment 4, the professionals maintained their MFD (cognitive effort of note-taking) at a high level (over 500 ms) across the two segments, whereas that of the students decreased significantly from 515 ms in Segment 3 to 477 ms in Segment 4. Thirdly, the professionals achieved similar interpreting scores in Segments 1 and 4, both of which were considered easy to interpret in the present study. However, the students showed even worse interpreting performance in Segment 4 (easy) than in Segment 3 (difficult), indicating a detrimental effect of presentation sequence on their interpretation quality. All of these findings suggest that the professionals could perceive task difficulty more sensitively, maintain their effort of note-taking at a high level, and keep their interpretation quality within an acceptable range. These findings are in accordance with the results obtained in RI experiments (Roziner & Shlesinger, 2010), where interpreters succeeded in maintaining their interpretation quality at a high level across on-site and remote interpreting, although they experienced a significantly higher cognitive load of interpreting in the distant mode of interpreting. The present study further extends this finding by suggesting that the professionals' expertise in interpreting, when compared with the students, is not only reflected in higher interpretation quality but also in their greater sensitivity to task difficulty and stronger resilience to such changes in task difficulty.

7.2 Reading patterns and cognitive effort of note reception

7.2.1 Reading within and across note groups

Corroborating S. Chen et al. (2022), the present study found that note-reading is conducted in note groups. On the one hand, the participants frequently revisited the individual notes that belonged to the same note groups. Based on the content and layout of the notes, they could retrieve the source speech memory, generate integral meaning chunks and deliver sensible interpretations. On the other hand, the participants usually started to read the next note group when they were going to finish the interpretation of the current one. By such reading-ahead (Gillies, 2005; Jones, 2002) operations, the participants were able to limit the frequency and duration of pauses during the output stage of CI and ensure a fluent delivery of the target speech.

This group-based note-reading pattern is also in line with Chang (2015) where the process of note-reading is depicted by referring to social semiotics and visual grammar. Through recording and observing the note-reading process, Chang (2015) found that the participants retrieved information at micro and macro levels from, respectively, the narrative and visual structures in the notes. For instance, interpreters can retrieve lexis from calligraphic signs (e.g., "En" for "English), vectors (e.g., "→" for "lead to") and geometrical shapes (e.g., "○" for "earth") in the narrative structures. They can also elicit contextual information by observing visual structures such as the layout of the notes (e.g., taking notes vertically to represent the hierarchy of information value) and use of salience (e.g., using underlines to illustrate the importance of a note). One important technique that helps the interpreters to constantly shift between processing the micro and macro levels of information is the use of framing. Framing can be achieved through "actual frame lines, white space between elements, and discontinuities of colour"

(Kress & van Leeuwen, 2006, as cited in Chang, 2015, p. 187), with the first two commonly adopted by interpreters during note-taking (see Figure 7-2). This indicates "that elements deployed in the interpreter's notes can (either) be given separate identities or represented as belonging together" (p. 44). By setting boundaries among different note groups, interpreters can separate the meaning units from one another and produce a meaningful and accurate target speech. It has been found that the use of framing in note-taking is positively correlated with the number of meaning chunks that are accurately interpreted in CI and the fluency of target speech delivery (Li et al., 2022). Professional interpreters can adopt framing in note-taking five times more than student interpreters do (Li et al., 2022). From this perspective, the group-based note-reading pattern observed in the present study provides empirical support for Chang's (2015) argument that framing plays an essential role in note-taking as it realises a choice in the textual metafunction by creating "visual boundaries between different units of information" (p. 54).

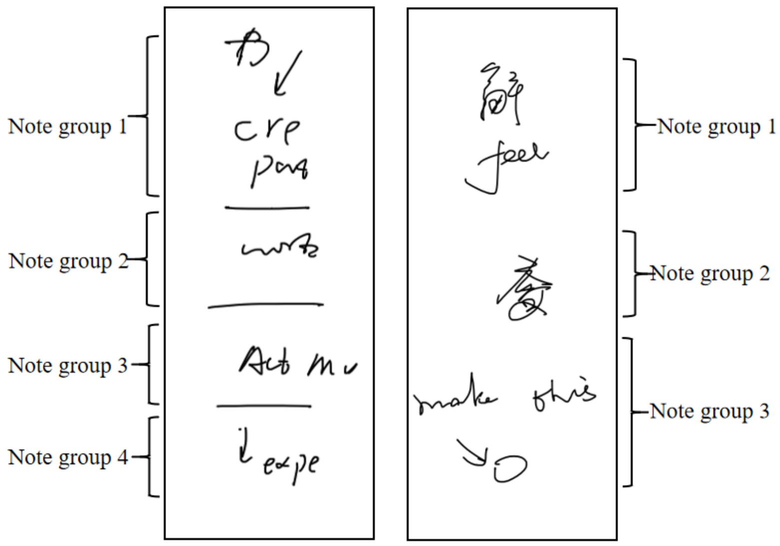


Figure 7-2. An example of frames in notes (actual frame lines on the left and the use of space on the right)

Moreover, this group-based note-reading pattern implies that note group can serve as an important unit of research for future note-taking studies to reveal the cognitive mechanisms underlying the note production and note reception processes. Looking into the composition of the note groups, it is found that, for the same source speech content, the participants can come up with note groups of different sizes, i.e., with different numbers of notes inside. As presented in Figure 7-3, P27 (a student interpreter) took three notes for the three content words in the first sentence in the source speech, whereas P32 (a professional interpreter) used six notes to indicate the same unit of the source speech. This result is not in accordance with what has been found in translation studies, that professional translators have bigger translation units than student translators (Jakobsen, 2003). Neither is it in accordance with Albl-Mikasa's (2006) speculation that professional interpreters are more willing to adopt note-taking strategies that depend on global-level source speech comprehension, and deprioritize note production. However, after checking the eye-tracking data (Figure 7-4), it is found that P27 spent 7,593 ms to read the first group of notes, while P32 only used 4,760 ms to finish the reading of the first note group. Moreover, an examination of the interpretation recordings shows that P27 did not complete the interpretation of the first note group ("*当我听到我要来* Ted *做一个演讲的时候我是很困惑的，因为我的父亲就*", meaning "when I heard that I was going to do this Ted talk, I was quite confused because my father"), and he misinterpreted "chuckled" as "confused" (*困惑*). By contrast, P32 finished this part of the interpretation accurately and fluently ("*当我被邀请参加* Ted *演讲的时候，我一听就笑了，因为我的父亲的名字就叫* Ted", meaning "When I was asked to do this TED Talk, I chuckled on hearing that because my father's name was Ted"). This result suggests that, while student interpreters attached more importance to the reception of notes, the professional interpreters paid more attention to the production of notes. By decreasing the cognitive demand for information retrieval during the process of note-reading, the professionals achieved higher interpretation quality than the students. From this perspective, the way interpreters decide the volume of source information in each note group may exert great impacts on their memory recall and speech production in the output phase of CI. With note groups as the unit of research, researchers can gain new insights into how the source speech contents are condensed into written notes, and how the message contained in the note groups is redelivered in the target speech.

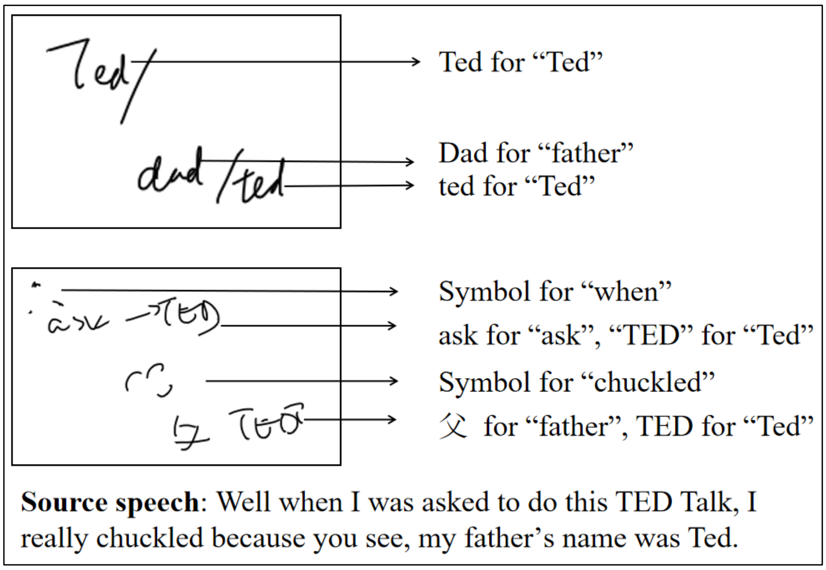
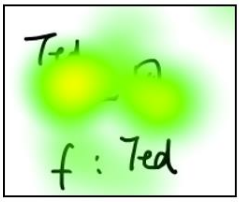
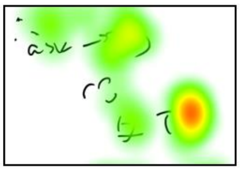


Figure 7-3. An example of different sizes of note groups for the same source speech unit (with the upper excerpted from P27's (student) notes and the lower excerpted from P32's (professional) notes)



AOI name (coded notes)	Identified note	Source speech unit	TFD
E%fw_en*Ted^1$1ted	Ted	Ted	3275
E%sym*/^1$ no_source	/		367
E%fw_en*ted^1$2ted	ted	Ted	1592
E%sym*/^1$ no_source	/		1175
Speaker window			325
RP%fw_en*dad^1$1father's	dad	Father's	859



AOI name (coded notes)	Identified note	Source speech unit	TFD
E%fw_en*ask^1$1asked	ask	ask	142
E%fw_en*TED^1$1ted	TED	Ted	867
E%sym*笑脸^1$1chuckled	smile face	Chuckled	200
E%ab_ch*父^1$1father's	父	Father's	358
E%fw_en*TED^1$2ted	TED	Ted	3193

Figure 7-4. The heatmaps and event logs of P27's (student) and P32's (professional) first note groups

7.2.2 Cognitive demand and visual conflicts

Overall, the cognitive effort of note-reading, which is indicated by the participants' MFD on the notes (over 300 ms among the students and over 290 ms among the professionals) during the output phase of CI, is found to be greater than that of many other reading tasks: 225 ms in silent reading for comprehension (Rayner, 1998), 275 ms in oral reading (Rayner, 1998), 205 ms (Jakobsen & Jensen, 2008) and 245 ms (Dragsted, 2010) in sight translation. Although these reading tasks cannot be compared directly as they vary considerably in the reading material and output modality, these results can still demonstrate that the MFD of note-reading is relatively high, suggesting a heavy cognitive load of note reception. Moreover, the MFD of note-reading found in the present study is longer than what is reported in S. Chen et al. (2021) (277ms). An important reason could be the extra cognitive demand of interpreting brought by VRI. In addition, looking into the MFD data on the individual notes (without the space among the notes) and the noted area (with the space included) in the present investigation, it is found that the former set of data always returns higher values than the second one, regardless of source speech difficulty and interpreter work experience. The difference between these two sets of MFD data reaches a significant level in the student group. These results indicate that, as long as note processing was involved along with target speech production, the cognitive load of interpreting increased substantially.

Tracing back to the cause of the high cognitive load of note-reading, three main reasons can be summarized. Firstly, since the notation language is featured with a "highly reduced or even fragmentary" surface that contains various "pictographic and iconic signs" (Albl-Mikasa, 2008, p. 211) in non-linear structures, this creates a variety of difficulties for interpreters to quickly recognize, integrate and translate the meaning of notes in different forms and languages within seconds. More often than not, student interpreters report that they cannot maintain a steady restitution speed because they do not understand their own notes and encounter memory retrieval problems during note-reading (Arumí Ribas, 2002). Even for professional interpreters, it is significantly more cognitively demanding for them to integrate the meaning of notes in the SL than in the TL, which further undermines the quality of their interpretations (S. Chen et al., 2021). Likewise, the present study detects a series of significant negative correlations between the participants' interpreting performance and the cognitive effort of early- and late-stage note-reading. All of these results demonstrate that note recognition and note-meaning integration during the process of note-reading can pose heavy cognitive demands on the participants' WM and impede them from generating high-quality interpretations.

Secondly, notes, on the one hand, serve as the only information source for interpreters to retrieve source speech memory during target speech production (Albl-Mikasa, 2008); while on the other hand, they compete for the interpreters' limited overt visual attention and cognitive effort (e.g., Gile, 2020), creating visual and cognitive conflicts during the output phase of CI. In the present study, over 30% of the notes were not fixated at all during the process of note-reading. This skip rate of note-reading is comparatively higher than that in S. Chen et al. (2021) (12%); and it basically approximates the situation of normal text reading (Rayner et al., 2011), where the skip rate is about one-third. This high skip rate of note-reading might not originate from the high predictability of notes. Instead, it can be caused by the fact that there are serious visual and cognitive conflicts for processing multiple notes at one time. To avoid cognitive overload, the participants chose to skip the reading of a number of notes at the risk of producing inaccurate and incomplete renditions. This argument can be further consolidated by consulting the participants' revisits to the notes during note-reading. In the present study, the students and the professionals revisited their notes, regardless of the note forms and note languages, more than twice while producing the target speech. Similarly, S. Chen et al. (2021) found that professional interpreters revisited their notes more than twice (except for symbols) during the process of note-reading. These results further imply that interpreters might not have adequate overt visual attention and cognitive effort to process all the notes, leading to a high skip rate of note-reading. In addition, it is found in the present study that the students' RVC to the notes was positively correlated with their FluDel, TLQual and total interpreting scores in the difficult segments of interpreting. This correlation suggests that revisiting notes is necessary during note-reading to help interpreters to generate a holistic understanding of the note groups and generate a complete and accurate target speech.

Thirdly, cognitive conflicts in verbal processing caused by reading the notation language and producing the TL at the same time can contribute to the participants' high cognitive load during note-reading. As mentioned, the skip rate of note-reading reached around 30% in the present study. Moreover, during the process of note-reading, the participants frequently looked at the blank space on the virtual notepad while delivering the target speech. These fixation patterns during the process of note-reading appear to suggest that the presence of notes could cause visual and cognitive interference during target speech delivery. Although no note-reading studies have reported such findings before, similar results can be found in studies on sight translation which resembles note-reading in visual input and auditory output. Chmiel and Lijewska (2019) found that both student and professional interpreters would look away from the source text during sight translation. Such a phenomenon is especially obvious when they interpret object-relative sentences which are assumed to be very cognitively demanding during target speech production because of additional syntactical processing. These results indicate that there is source text interference in target speech delivery during sight translation and that such interference increases as task difficulty increases. Applying these results to the present study, notes serving as the source text during the output stage of CI can entail such source text interference during target speech production. This interference is especially prominent when interpreters have difficulties in processing their notes. Therefore, they look away from the notes to decrease such verbal processing conflicts in delivering the target speech, which meanwhile, demonstrates the great cognitive demand for note-reading.

All in all, note-reading is a cognitively demanding task that differs from normal reading tasks in three aspects: 1) the reading material for note reception is comprised of a special notation language that requires much cognitive effort in note recognition and meaning integration; 2) processing multiple notes at the same time causes serious conflicts in visual and cognitive processing; and 3) reading the notation language and producing the TL at the same time causes cognitive conflicts in verbal processing.

7.3 Note production and note reception

7.3.1 More note-taking effort, more note-reading effort

By correlating the overall effort of note-taking with that of note-reading, the present study finds a positive relationship between the two, suggesting no trade-off between the total effort of note production and that of note reception. This result is seemingly not in line with Craik and Lockhart's (1972) levels of processing model which assumes a balance between the effort expended on language processing and the cognitive demand for memory retrieval. However, it in fact supports Craik and Lockhart's model after taking the participants' note-taking strategy and note quantity into consideration. In the present study, regardless of source speech difficulty, the participants mainly applied an ellipsis strategy to conduct note-taking. Moreover, on average, they took one note for every four words (for students: 1 note for every 4.60 words; for professionals: 1 note for every 4.26 words) in the source speech. However, people's writing speed is only one-tenth of their speaking rate (Foulin, 1995). This means that, while taking these large quantities of notes through the ellipsis strategy, the participants expend a considerable amount of cognitive effort on information retention for the purpose of finishing note-writing. With a great amount of effort devoted to note-taking, however, this process of note production does not involve deep-level information processing, which is a prerequisite for fast memory retrieval proposed by Craik and Lockhart. In a word, taking a large number of notes with an ellipsis strategy demands a great number of cognitive resources for information retention during the input stage of CI. At the same time, because of a low level of language processing during the input, this approach to note-taking also entails a heavy cognitive demand for memory retrieval during the output stage of CI. A positive relationship between the effort of note production and of note reception is thus established. In addition, the strongest correlations between the note-taking and note-reading effort are frequently observed in the measure of SPD, which indicates the cognitive effort of integrating the meanings of notes at a sentence and textual level. This result suggests that, the more notes that are produced, the heavier the cognitive demand is to integrate the meanings of these notes into a coherent whole, which further implies a low level of source speech comprehension during the production phase of notes.

Discussions above manifest that the adoption of note-taking strategy plays an important role in deciding the positive or negative relationship between the note-taking effort and the note-reading effort. As discussed earlier, the preference for the ellipsis strategy of note-taking is detected in other studies (e.g., Albl-Mikasa, 2006). Meanwhile, studies focusing on the EPS of note-planning suggest that there is a great percentage of notes that can be traced back to the source speech (e.g., 90% in S. Chen, 2020b with 26 professional interpreters and 83% in S. Chen, 2022 with 22 professional interpreters and 22 student interpreters), again indicating close connections between the choice of notes and the linguistic expressions in the source speech. These empirical findings appear to suggest that the ellipsis strategy is widely adopted by interpreters in interpreting practice. Since an ellipsis strategy can easily result in a larger note quantity, combining the results reported in the present study, it appears that there is a common positive relationship between interpreters' note-taking effort and note-reading effort in interpreting practice. According to Albl-Mikasa (2008), such a micropropositional approach to note-taking can be advantageous when it helps interpreters to better retrieve their memory of the source speech content. However, when the cognitive demands for information retrieval is too heavy during the output stage of CI, devoting a great amount of cognitive effort into note-reading can result in poor interpretation quality.

Interpreters' preference for the ellipsis strategy also seems to deny the conventional principle of note-taking: "(n)oting the idea and not the word" (Rozan, 1956, p. 15). To understand the reasons behind interpreters' deviation from this note-taking principle, two important clarifications should be made about the ellipsis strategy. Firstly, the ellipsis strategy can be easily misunderstood as a shorthand of the source speech, leading to the rejection of this text-based approach to note-taking (Albl-Mikasa, 2008). However, an ellipsis strategy of note-taking is different from that of shorthand, as the former approach requires interpreters to identify, select and transform the micropropositions in the source speech, whereas the latter is a simple sketch of the source speech. Shorthand is slow in handwriting and does not involve any nonverbal comprehension (Seleskovitch, 1975). Therefore, it "does not work as a means for note-taking and should be rejected" (Matyssek, 1989, as cited in Gillies, 2017, p.272). As for ellipsis, although its "result is not so much a detachment of the source text's surface structures, (it is) some kind of loosening of it" (Albil-Mikasa, 2008, p. 216), it demands speech analysis from the interpreters. From this perspective, "'note the idea rather than the words' does not mean that one has to give up the propositional form and move to a (deverbalised) level distant from the text" (p. 224). Instead, an ellipsis note-taking strategy could help interpreters to identify the micropropositions in the source speech, while avoiding the issues of slow handwriting by selectively omitting the source speech content.

Secondly, the interpreters' preference for an ellipsis strategy of note-taking is also decided by the inherently complex cognitive demands of note-taking and interpreting. As illustrated by Gile's Effort Model, interpreters have to simultaneously fulfil the cognitive demands for L (listening), M (memory), NP (note production) and C (coordination) during the comprehension phase of CI. Note-taking itself also contains note-planning and note-writing procedures, where cognitive and motor controls are required to be executed co-ordinately (e.g., Peverly et al., 2014). Under this situation, it is very difficult for interpreters to generate a thorough understanding of the source speech and note down the extracted sense of the speech without any linguistic references (Kirchhoff, 1979). For this reason, interpreters generally conduct note-taking through ellipsis; and this strategic move determines that they have a lot to make up during the process of note-reading because of their inadequate comprehension of the source speech during note-taking. Similar findings can be observed in Andres (2002) who finds that what hinders the student interpreters' listening comprehension and note-taking during the input phase of CI also impedes them from target speech production during the output phase of CI. With much effort devoted to both the note-taking and note-reading processes, a positive relationship between these two types of effort thus exists.

When comparing the note-taking effort and the note-reading effort in different note categories, it is found in the present study that, compared with symbols, language notes always demanded significantly more cognitive and physical resources from the participants in both the production and reception phases of note-taking. This result provides empirical evidence for Matyseek's (1989) symbol-based note-taking approach. On the one hand, symbols "are quicker and easier to write than words" and "can save space on the page, leaving the structure of (your) notes clearer" (Gillies, 2017, p. 101); on the other hand, symbols "are quicker and easier to read on the page than words" and they "represent concepts not words…so they help (us) avoid source language interference" during interpreting (p. 101). However, even though language notes were more cognitively and physically demanding in note production and note reception, the participants in the present study still preferred to take notes in language rather than in symbols, corroborating previous findings in Andres (2002), S. Chen (2017b, 2022) and Dam (2004a). The reason behind this preference for language notes could be explained by referring to Gile's (2020) notion of language availability, which was proposed to explain why interpreters differ in "the time it takes to find/understand a word/linguistic structure" (p. 19) in language comprehension and language production. In the notation language, compared with language notes, symbols require an additional procedure of SL decoding to extract the intended meaning in the source speech, leading to lower language availability in symbols than in language notes. Unless the participants could achieve an automatic use of symbols during note-taking, it would be easier for them to note in language rather than in symbols.

However, this explanation seems to be against the present study's finding, where a longer EPS was found in language notes rather than in symbols. These seemingly contradictory findings could be resolved through a further examination of the proportions of content-based and function-based symbols in the collected data. This study classified symbols by following the criteria proposed in S. Chen (2016, p. 11), that "(a) symbol is a representation of (1) the underlying meaning of a word or expression rather than the actual word or expression; or (2) the relationship(s) between two units". However, under the general category of symbols, there are symbols that have actual meaning and can be directly traced back to the source speech (e.g., "♪" for "music" in the source text) and those that do not have intended meaning or corresponding source units (e.g., "/" for separating different grammatical elements in a note group). In the present study, it is found that around 49.63% (*SD*=17.17%) of the students' notes and 51.61% (*SD*=16.66%) of the professionals' notes did not contain actual meaning. They simply exist as functional notes to indicate the logical relations among the individual notes. After excluding this part of the data, both the students (symbols: *M*=4.65 s, *SD*=1.74 s; language notes: M=4.42 s, *SD*=1.47 s) and the professionals (symbols: *M*=4.66 s, *SD*=1.70 s; language notes: *M*=4.27 s, *SD*=1.63 s) presented longer EPS in producing symbols than in taking language notes, with the difference in the student group reaching a significant level (Z=-2.259, p<.05). This data result indicates that, when creating content-based symbols, the participants had to expend considerably more time on note-planning than on producing language notes; while producing function-based symbols demands considerably less note-planning time than taking language notes. Therefore, when recording the actual contents of the source speech with notes, the participants still preferred language notes over symbols.

In summary, a positive relationship between the overall note-taking and note-reading effort is observed in the present study because of the participants' frequent adoption of an ellipsis strategy during note-taking. Breaking down the noting effort according to different note categories, symbols generally demanded less effort from the participants than language notes during the two noting stages because of fast handwriting and little language interference. However, content-based symbols could entail a longer EPS than language notes as the former took more time for the participants to decode the source information and transfer it into written notes.

7.3.2 More note-taking effort, less note-reading effort

Looking into language notes, the present study finds that abbreviations entailed greater cognitive effort of note-taking (only significant in easy segments) and longer time for note-planning (only significant in difficult segments) than full words during note production. As EPS can "potentially indicates the combined cognitive effort required to analyse the source speech unit, to decide whether or not to write a note, and if so, which form and language to use" (S. Chen, 2020b, p. 134), and as the EPS and MFD data point to the same direction in this part of the investigation, these results suggest that it is more cognitively demanding to take notes in abbreviations than in full words. When it comes to note reception, a reverse phenomenon is observed, that full words demanded a greater amount of cognitive effort from the participants for note-processing than abbreviations did in all measures. Taken together, there is a trade-off between the cognitive effort of note-taking and the cognitive effort of note-reading across full words and abbreviations. Craik and Lockhart's (1972) levels of processing model can provide explanations for this observed balance of cognitive effort. During the process of producing notes in abbreviations, interpreters usually follow five methods: 1) noting the first and last letters of a word (Gillies, 2017; Matyssek, 1989; Rozan, 2002); 2) writing a few initial letters to indicate the whole word (Becker, 1972); 3) borrowing affixes and suffixes from the working language (such as -ism) (Gillies, 2017; Matyssek, 1989); 4) taking advantage of phonetic spelling and misspelling (such as “thru” for “through”) (Gillies, 2017); and 5) following abbreviations that are widely adopted in daily life (such as “F” for “female” and “M” for “male”) (Matyssek, 1989; Wu, 2008). All of these abbreviating procedures require some extent of language processing, including letter selection, affixes/suffixes identification as well as LTM search and retrieval. Compared with full words which could be noted without further manipulation, abbreviations thus entailed a deeper level of language processing during note-taking and resulted in an easier retrieval of memory during the note-reading. In principle, as interpreters automatize the application of these five abbreviating rules, they can take notes in abbreviations very easily. However, since the present study found a higher cognitive load of note-taking in the abbreviations than in the full words, this indicates that the participants had not completely mastered the abbreviating procedures in note-taking.

The observed differences between full words and abbreviations in the note-taking and note-reading effort are partially in accordance with the findings in S. Chen (2017a). Specifically, S. Chen (2017a) found that reading notes in abbreviations is less cognitively demanding than reading notes in full words, whereas the cognitive load of taking notes in these two forms is at a similar level. However, since S. Chen (2017a) draws this conclusion by using EPS to indicate the cognitive effort of note-taking, which in the present study indicates the time spent on note-planning, the results concerning the cognitive effort of note-taking in the two studies cannot be compared directly. Simply looking at the EPS data, S. Chen (2017a) reports that abbreviations entailed longer EPS than full words but that the difference did not reach a significant level. This finding is in line with what is found for the easy segments of the present study. The fact that this difference in EPS between full words and abbreviations turned significant in the difficult segments of interpreting suggests that the gap between the two types of notes in the EPS data widens as source speech difficulty increases. Specifically, when the cognitive demand for SL processing increases, the participants need more time to identify and abbreviate the words in the source speech, thus widening the gap between full words and abbreviations in the time spent on note-planning. In summary, based on the data collected in the present study, compared with full words, abbreviations demand more note-taking effort and less note-reading effort from interpreters.

Between Chinese and English notes, this study found that the participants devoted more effort to taking Chinese notes, whereas they expended more effort on reading English notes (only close-to-significant or significant in the easy segments). This result is easy to explain when considering that Chinese served as the TL of the interpreting task and English as the SL. Noting in Chinese requires the participants to complete an additional procedure of language transfer during the process of note-planning; while reading notes in English while ensuring a fluent delivery of the target speech demands fast and accurate language transfer (Abuín González, 2012). Both operations include an additional procedure of language conversion, leading to increased cognitive demands of note processing in taking Chinese notes and reading English notes (Abuín González, 2012). However, the fact that significant differences in the note-reading effort between the two note languages only exist in the easy segments of interpreting suggests that, when the source speech is difficult, identifying and transforming the meaning of Chinese notes can be as difficult as for English notes.

It is worth mentioning that, for the measures of the cognitive effort of note-reading, close-to-significant or significant differences between the two types of note languages are only observed in TFD and SPD in the easy segments, which points to the cognitive effort of late-stage note-processing. This result indicates that notes in the two languages do not differ significantly in the cognitive effort of note recognition or lexical-level meaning integration in the early stage of note-reading. Instead, compared with Chinese notes, English notes require significantly more cognitive effort from the participants during the process of integrating the meaning of notes at a sentence and textual level. This result is in accordance with S. Chen et al. (2021), where the researchers reported significantly longer TFD and SPD in English notes than in Chinese notes in English-to-Chinese interpreting. Moreover, S. Chen et al. (2021) does not detect such significant differences in Chinese-to-English interpreting. Combining the findings of S. Chen and those of the present study suggests that it is only in the L2-to-L1 interpreting that English notes demand significantly more cognitive effort during late-stage note processing than Chinese notes for Chinese native interpreters.

These results provide important implications for note-taking training, where interpreter trainers should pay more attention to help student interpreters to organize their notes logically. As reported by Shen and Liang (2019, p. 11), many self-repairs during CI among student interpreters are caused by the fact that the students' "notes were not logically presented for long segments of the original speech, which forced them to restart once the sentence structure in the target version was no longer appropriate". If notes are organized in a way that can clearly indicate the chunks of information in the source speech, this can greatly facilitate the note reception process (Shen & Liang, 2019), which echoes the importance of note groups in note-taking as discussed in Section 7.2.1.

Overall, a trade-off between the cognitive aspects of note-taking and note-reading effort is observed between different note forms (full words and abbreviations) and different note languages (Chinese and English) in certain task conditions. By contrast, such a relationship is not commonly observed between the visual, physical and temporal aspects of the note-taking effort and the cognitive aspects of the note-reading effort.

7.4 Note-taking behaviour and interpretation quality

7.4.1 Weak correlations between note-taking behaviour and interpreting performance

Overall, the present study observes a few weakly (r<.50) significant correlations between the interpreters' note-taking behaviour and CI quality, which corroborates previous findings of no close or clear relationships between many aspects of note-taking and interpreting performance (e.g., S. Chen, 2017b; Dai & Xu, 2007; Wang et al., 2010). This suggests that note activities as subtasks in CI (e.g., Gile, 2009; Gillies, 2005) are not closely related to CI quality. After all, "note-taking is just a means, and not the end, of CI" (Viezzi, 2013, as cited in Russel & Takeda, 2015, p. 103). This loose connection between note-taking activities and interpreting performance has previously been reported in S. Chen (2017b), where she scored interpreters' notes according to whether they correctly represented source-speech units and were correctly translated in the target speech. S. Chen (2017b) found that high-quality interpretation always appeared with high note-taking scores, but that highly scored notes did not necessarily yield successful interpretation. Such results indicate that good note-taking is only a part of the conditions needed to produce high-quality interpretation. The following discussion explains how this weak association between note-taking and CI quality exists (does not exist) in different interpreter groups and task conditions.

7.4.2 The relationship between CI quality and the production and reception of notes

Many significant correlations observed in the present study are in the note-reading stage rather than in the note-taking stage. One primary reason for this is that note-reading is directly related to the production of the target speech, as the two tasks are conducted concurrently. According to the limited-capacity resource model (Kahneman, 1973), concurrent tasks usually induce attentional conflicts as more processing capacity devoted to one task will lead to less capacity available for other tasks. Therefore, smooth note-reading can facilitate information recall and leave more cognitive resources for target speech production and monitoring, but arduous note-reading can make interpreters get stuck in speech organization and stumble in speech delivery. As mentioned earlier, Shen and Liang (2020) discovered that student interpreters often failed to understand the logical relations in their notes, leading to many disfluencies in their renderings. This result supports that when too much effort is allocated to reading notes, little effort is available to ensure high interpretation quality.

Another reason for the common correlations between the note-reading effort and the participants' interpreting performance could be caused by the adopted note-taking strategy. It has been proven that the choices interpreters make during note-taking are influential in determining the effort of note-reading (e.g., S. Chen et al., 2021). In the present study, over 85% of the collected notes in each segment were created through the ellipsis strategy. In other words, only a superficial level of language processing was involved during note-taking in most situations. As discussed in Section 7.3.1, such a shallow level of SL processing during note-taking could lead to a greater cognitive demand for information retrieval during the process of note-reading. In fact, this was proved by a series of significant positive correlations between the note-taking effort and the note-reading effort observed in the present study. Since little listening comprehension was achieved during the inputs stage of CI, interpreters had to make up for the inadequate source speech comprehension during the process of note-reading. This procedure has to be completed quickly, accurately and fluently, as notes created through the ellipsis strategy usually entail large quantities (Albl-Mikasa, 2006). If interpreters are troubled by any stage of note processing during note reception, the speech production would be hindered and interpretation quality would drop directly. Taken together, the adoption of note-taking strategies can exert great impacts on the effort of note-reading; while the effort of note-reading further determines the amount of processing capacity that is available for target speech production and monitoring. The connection between note-reading and CI quality is consequently established.

For the same reason, with an ellipsis-dominated note-taking approach, both groups of participants only show positive correlations between their note quantity and interpretation quality in the easy segments but not in the difficult ones. Those aspects of note-taking that are linked with increased note quantity in the easy segments, such as shorter note-planning (EPS) among the professionals and more handwriting (CC_sum) among the students, also present the same kinds of correlations. This is because it was more difficult for the participants to make up for the inadequate comprehension of the source speech in the difficult segments than in the easy segments, which further affected their interpreting performance. This effect of task difficulty also explains why such a positive association between interpreters' note quantity and CI quality has sometimes been observed in previous literature (e.g., S. Chen, 2020b; Dam, 2007; Her, 2001) and sometimes not (e.g., Dai & Xu, 2007; Liu, 2010; Wang et al., 2010).

7.4.3 The relationship between note-taking and CI quality in students and professionals

Overall, the student group show many correlations between their note-taking behaviour and CI performance in the difficult segments, which mainly involved the product of note-taking and the process of note-reading. By comparison, the professional group present most of the correlations in the easy segments, which concerned both the processes of note-taking and note-reading.

One primary reason for the observed correlations among the student group is their inadequate note-taking expertise. In the difficult segments, the students show various correlations between their note choices and CI quality. They were troubled by the note form that required much physical effort of handwriting (full words) and the note language (English) that demanded much effort for note-reading. By comparison, no significant correlations are observed in the easy segments. Nor are these correlations observed among the professionals. This finding indicates that, when the demands of listening comprehension increased, the students who were occupied by SL processing could not retrieve the appropriate note forms and note languages from their LTM quickly and accurately. Another reflection of this effect of source speech difficulty on the relationship between their note-taking behaviour and interpretation quality is the change in their interpreting performance in different aspects. In the easy segments, the students' note-reading effort is only associated with their scores in one aspect of the assessed interpretation quality. However, in the difficult segments, significant correlations are witnessed between the cognitive effort of note-reading and any aspect of interpretation quality. This can be caused by the fact that the ellipsis strategy postponed the comprehension and interpreting tasks of the source speech from the process of note-taking to the process of note-reading (Albl-Mikasa, 2006). When it came to the difficult segments where the cognitive demand for comprehension and language transfer was especially high, if they could not recognize the notes in early processing (FFD) quickly and integrate the meaning of notes in late-processing (SPD) correctly, their information completeness could be directly affected. Extra effort was also needed in incomplete processing (RVC) to maintain their interpreting fluency and TL quality.

Unlike the student interpreters, the professional interpreters show most significant correlations between their note-reading effort and CI quality in the easy segments. The professionals' articulation rates, which are an important determinant of judged fluency, are calculated for both two task conditions by dividing the total duration of speech (apart from silent and filled pauses) by the number of syllables in the target speech (Yu & van Heuven, 2017). The results show that the professionals' articulation rates are similar in the two task conditions (easy: M=4.95, SD=0.67, difficult: M=4.85, SD=0.68, Z=-1.459, p>.05), and that their articulation rates are always significantly higher than those of the students (p<.01). This means that, compared with the difficult segments, the professionals had to read a greater number of notes at one time in the easy segments to ensure a continuous interpretation. With more processing capacity devoted to note-reading, less was available for target speech production and monitoring (Kahneman, 1973). This could be the reason why negative correlations are observed between the professionals' note-reading effort and CI quality in the easy segments. Despite their high interpreting scores in the easy segments, these negative correlations indicate that a large note quantity can cause note-reading problems, and that the effort of note-reading is associated with the quality of interpretation.

In the difficult segments, the professionals present few negative correlations between their CI quality, on the one hand, and their cognitive effort of note-taking and note-reading, on the other hand. One explanation for these results could be that the professionals experienced cognitive overload during interpreting, which led to a decreased role of note-taking in influencing interpreting performance. Despite their interpreting experience, the professionals reported even more mental demand (professionals: *M*=6.93, *SD*=1.46, students: *M*=6.59, *SD*=1.24) and a higher level of frustration (professionals: *M*=6.18, *SD*=2.16, students: *M*=5.95, *SD*=1.73) in the difficult segments than those reported by the students. Their performance dropped significantly in the difficult parts and was only slightly better than that of the students. These findings suggest that the professionals are highly sensitive to task difficulty and might experience overload during interpreting. Similar findings have been reported in Hu (2008), where professionals' cognitive effort of note-taking was especially sensitive to source speech difficulty, and in Cardoen (2018), where professionals did not outperform students in difficult CI tasks. Taken together, these results indicate that the role of note-taking can be very limited when professional interpreters perceive there to be much difficulty in CI. However, further investigation is needed to examine how this association is established during the process of interpreting.

Chapter 8: Conclusion

This thesis explored the process of note-taking, the product of note-taking and the process of note-reading in remotely-conducted and video-mediated CI, by collecting online eye-tracking and pen-recording data from two groups of interpreters (student interpreters and professional interpreters) within two task conditions (easy and difficult), and analysing their offline products of note-taking and interpreting through descriptive and summative methods. By triangulating the data obtained through these process- and product-oriented methods, this study addressed the following four research aims:

> *Aim 1*: to explore the effects of interpreter work experience and source speech difficulty on the process and product of note-taking, which includes the visual, cognitive, physical and temporal aspects of the note-taking effort and the distribution of note quantity, note form, note language and note-taking strategy in the note-taking product.

> *Aim 2*: to examine the process of note-reading by visualizing the interpreters' fixation patterns on the notes with the help of heat maps, gaze plots and event logs, and by quantifying the interpreters' note-reading effort with a range of fixation measures that point to both the early and late stages of note processing.

> *Aim 3*: to investigate the relationship between the effort of note-taking and the effort of note-reading, with the former being measured from four aspects (visual, cognitive, physical and temporal), and the latter being quantified with eye-tracking measures that reflect the cognitive effort of different stages of note processing.

> *Aim 4*: to explore the associations between the interpreters' note-taking behaviour on the one hand, namely their note-taking effort, note-taking product and note-reading effort, and the interpretation quality on the other hand, which includes information completeness, fluency of delivery and TL quality.

In response to these research aims, five research questions (RQs) were proposed and eleven corresponding hypotheses (H) were tested in the present study. The effects of interpreter work experience and source speech difficulty on the process of note-taking (RQ1 and H1-H2) were examined from the following aspects: overt visual attention that the participants paid to the notes, cognitive effort of note-taking, physical effort of note-writing, and time spent on note-planning. Compared with the cognitive and temporal aspects of note-taking, the visual and physical measures presented many more significant differences across different interpreter groups and task conditions. In addition, sequence effects were observed in almost all the measures, with significant differences mainly found in the note-taking effort between Segments 1 and 2 (in the difficulty-increase direction of interpreting) but not between Segments 3 and 4 (in the difficulty-decrease direction of interpreting). As for the experience and difficulty effects on the product of note-taking (RQ2 and H3-H4), these were investigated by comparing the two groups of participants' note quantities, note forms, note languages and note-taking strategies in the easy and difficult segments of interpreting. Note quantity was found to be significantly different across the students and the professionals, as well as between the easy and the difficult segments of interpreting. In addition, the professionals presented a clear preference for full words in their notes, whereas the students preferred to adopt abbreviations in note-taking. The focus of the second research aim, i.e., the process of note-reading (RQ3), was investigated from the perspectives of reading patterns (H5) and cognitive effort (H6). An examination of the heat maps, gaze plots and event logs revealed that the note group served as an important unit of reading during note reception. A comparison of the MFD data between the note-reading task and other types of reading tasks demonstrated that note reception was a very cognitively demanding process. Relationships between the note-taking and note-reading effort (RQ4 and H7-H8), as well as the associations between the participants' note-taking behaviour and interpretation quality (RQ5 and H9-H11), were explored through correlation tests. A generally positive relationship was found between the effort of note production and note reception. Comparisons were also conducted between various note forms and note languages in terms of the note-taking effort and the note-reading effort. It was found that there was a cognitive trade-off between full words and abbreviations, as well as between Chinese and English notes during note production and note reception. Meanwhile, a series of positive and negative significant correlations were observed between interpreters' note-taking behaviour and interpreting performance.

This chapter concludes the whole study with a summary of the key findings relating to the research questions and hypotheses (Section 8.1). Implications on how these findings could assist interpreting practice and interpreter training are presented in

Section 8.2. The strengths and limitations of the present investigation are discussed in Section 8.3. Finally, possible directions for future research and relevant suggestions are provided in Section 8.4.

8.1 Summary of the major findings

Overall, results obtained from the eye-tracking and pen-recording measures indicate that the effects of interpreter work experience and source speech difficulty on the process of note-taking are prominent in the overt visual attention that the participants paid to the notes and the physical effort they devoted to note-writing. This result can be attributed to the participants' note choices, experience with VRI, and cognitive load of interpreting.

Firstly, on the product of note-taking, the students preferred note forms that require less overt visual attention and physical effort in handwriting, such as symbols and abbreviations, whereas the professionals frequently used note forms that demand much of these types of note-taking effort, such as language notes and full words. Therefore, a significantly greater amount of overt visual attention and physical effort is detected in the professional group than in the student one. Moreover, the fact that the professionals always took more notes than the students in each segment of interpreting also contributes to the two groups' differences in the visual and physical aspects of note-taking effort. Secondly, the same effects of experience are observed in the average overt visual attention that the participants paid to the screen and to the noted area in the segments of interpreting. Since the two groups of participants had similar WM capacities and reported a similar amount of mental demand and frustration after interpreting, the observed group differences indicate that, when facing interpreting difficulty, the rich experience that the professionals had with VRI helped them to activate more cognitive resources to deal with the high cognitive load of interpreting and note-taking. Thirdly, a negative correlation test result between the participants' cognitive effort of interpreting and interpreting performance in the difficult segments suggests that the decreased note-taking effort in the difficult task condition could be caused by a cognitive overload of interpreting. This result suggests that poor interpretation quality in VRI can be traced back to the input stage of CI where interpreters have great difficulties in handling note-taking and video-watching at the same time.

On the other hand, no experience or difficulty effects are observed on the cognitive effort of note-taking. This result can be attributed to the participants' preference for an ellipsis-based note-taking strategy throughout the four interpreting segments. It is found that over 85% of the collected notes in each segment of interpreting were produced through the ellipsis strategy, and that the participants managed to create more notes through this strategy in the easy segments than in the difficult ones. This means that, in the easy segments, the participants' WM was heavily occupied by the need for information retention during the process of note-writing; while in the difficult segments of interpreting, their WM was again heavily demanded by SL processing and note-planning. As a result, they faced a high cognitive load of note-taking in both task conditions, leading to no difficulty effects on the cognitive effort of note-taking. Similarly, although the professionals managed to create more notes than the students whose WM was occupied with listening and analysis, the former group also experienced a high level of cognitive load during note-taking because of the great demand for information retention during note-writing. For this reason, no effects of experience are observed in the cognitive effort of note-taking.

In addition, significant differences in the note-taking effort are frequently observed between Segments 1 and 2 in the difficulty-increase direction of interpreting (from easy to difficult), but not between Segments 3 and 4 in the difficulty-decrease direction of interpreting (from difficult to easy). Such sequence effects of source speech difficulty on the process of note-taking are in accordance with the participants' NASA-TLX scores and interpreting performance. These results indicate that, when interpreting a difficult segment before an easy one, there are spill-over effects on task difficulty perception. With a distorted perception of interpreting difficulty, the participants maintained their note-taking effort at a similar level across Segments 3 and Segment 4. At the same time, compared with the students, the professionals showed greater resilience to such sequence effects of source speech difficulty, presenting interpreting expertise in task difficulty perception, effort devotion and interpreting performance.

On the product side, note quantity is found to be the dependent variable most sensitive to the two independent variables of interpreter work experience and source speech difficulty. This finding corresponds to the results concerning the process of note-taking in that significant differences are mostly observed in the total overt visual attention paid to the noted areas and the total physical effort devoted to note-writing. In terms of the choice of notes, the professional interpreters present a clear preference for full words over abbreviations, whereas the student interpreters show an opposite preference between these two note forms. One important reason for this group difference in the use of language notes is that the professionals pursued large note quantities in the interpreting task, which required them to adopt note forms that required less cognitive effort in note-taking and shorter time in note-planning. Since full words entail shorter MFD and EPS (significant in the difficult segments of interpreting) than abbreviations during note-taking, this implies that noting in full words involves less cognitive processing and faster note-planning. Therefore, although using more full words can lead to a heavier visual and physical load of note-taking, the professionals still prefer this form of notes throughout the segments of interpreting.

Moving to the process of note-reading, this study detects a group-based reading pattern among the participants. On the one hand, they frequently revisited the notes in the same note group to integrate the meaning of individual notes into a coherent whole and recall the meaning units in the source speech; on the other hand, when they were going to finish the interpretation of the current note group, they would glance at the next one to prepare for the following interpretation and ensure a fluent speech delivery. During this process, visual and cognitive conflicts happened frequently because the participants had to process multiple individual notes, even different note groups, at the same time. This means that, without an effective distribution of their limited overt visual attention and cognitive effort during the output stage of CI, this could easily result in interpreters experiencing a cognitive overload of interpreting and undermined interpretation quality.

In addition, a positive relationship is detected between the average effort that the participants expend on taking a note and the effort they use in reading a note, and this is decided by the participants' note-taking strategy. With an ellipsis-based note-taking approach, the participants create large quantities of notes without an in-depth understanding of the source speech. Therefore, during the process of note-reading, they have to quickly process multiple notes at the same time to make up for the inadequate source speech comprehension and deliver an accurate and fluent target speech. Hence, the more effort is devoted to note production, the greater the cognitive demand for note-reading will be. Meanwhile, examining the relationship between the note-taking effort and the note-reading effort in different note categories reveals that there is a cognitive trade-off between full words and abbreviations, as well as between Chinese notes and English notes. Specifically, noting in abbreviations entails a heavier cognitive load during note-taking than noting in full words, because of the extra procedure of abbreviating; while taking notes in the TL of the interpreting task demands more cognitive effort from the participants than noting in the SL due to the additional procedure of language conversion. On the other hand, with more in-depth SL processing during the process of abbreviating and language transferring, the participants could decode notes in abbreviations and the TL of the interpreting task with less cognitive effort during the process of note-reading. For this reason, a trade-off is detected between these note forms and note languages in the cognitive effort of note-taking and note-reading. However, it is worth mentioning that this cognitive trade-off mainly exists in the easy segments of interpreting but not in the difficult ones. This result suggests that, when the source speech is difficult to interpret, noting or reading in different forms (full words or abbreviations) or different languages (Chinese and English) can be similarly effortful.

Finally, this study examined the associations between the participants' note-taking behaviour and interpretation quality through a series of correlation tests. Although significant results are detected in many of the tests, the coefficients are mostly at a weak level (r<.5), suggesting a loose connection between note-taking and interpreting performance. Three general patterns can be summarized from the observed correlations. Firstly, note quantity is positively correlated with the two groups' interpreting scores in each assessed aspect of interpretation quality. However, such positive correlations only exist in the easy segments of interpreting but not in the difficult ones. Secondly, compared to the process of note-taking, the process of note-reading is more closely associated with interpretation quality, because it is concurrently conducted as the production of the target speech. Thirdly, the students present most of their correlations in the difficult segments of interpreting because of limited note-taking expertise, whereas the professionals show most of the correlations in the easy ones, which could be attributed to their large note quantity. Overall, the participants' note-taking behaviour is not closely related to their interpretation quality, indicating a limited role of note-taking in VRI from the performance perspective.

8.2 Implications for interpreting practice and pedagogy

A series of effects of interpreter work experience and source speech difficulty have been observed in the process and product of note-taking, which can provide some didactic implications for interpreting practice and interpreter training. Although both groups of participants distributed a large proportion of their cognitive resources to the noted area on the screen during the input stage of CI, the professionals always activated a greater number of cognitive resources during note-taking and interpreting than the students as indicated by their eye-tracking measures and self-reported *Effort* scores. Moreover, compared with the professionals, the students always devoted more visual attention and cognitive effort to process the visual information in the speaker window. On the one hand, the two participant groups present a similar WM capacity and a similar ellipsis-based approach to note-taking. This result implies that the group difference in the activated number of cognitive resources during interpreting can be traced back to their differences in their LTM, i.e., the number of note-taking and interpreting-related knowledge schemas at their disposal. In other words, professional interpreters can activate more cognitive resources to respond to the perceived difficulty because they possess the corresponding knowledge schemas, whereas student interpreters do not have these schemas stored in their LTM. Therefore, during interpreter training, apart from short-term memory training, the accumulation of interpreting skills and note-taking techniques in LTM is essential in helping students to increase the number of schemas they can utilize during interpreting practice. At the same time, the professionals always returned higher *Effort* scores than the students, even though they reported a similar amount of *Mental demand* during interpreting. This result implies a possibly higher motivation level among the professionals than among the students, suggesting the importance of the attitude towards the perceived task difficulty. It is found that student interpreters usually would not sustain their efforts if they experienced repeated failure during their interpreting training (Wu, 2016). Therefore, while training students' interpreting ability with a variety of exercises (e.g., short-term memory and note-taking skills), guiding them to maintain an adequate level of motivation is also important in developing interpreting ability.

On the other hand, as reported by the student interpreters in the present study, in interpreting class they were sometimes suggested to facilitate their listening comprehension during the input stage of CI by collecting visual information from the speaker and the meeting room. However, results in the present study indicate that this can be very challenging in a VRI environment where interpreters have to filter useful visual information based on what they are presented with. The extra effort expended on information searching can pose great challenges to their limited overt visual attention and processing capacity. This challenge is also reflected in the difficulty effects on the participants' note-taking effort.

In the present study, the participants decreased their note-taking effort for the difficult segments of interpreting because of a cognitive overload in interpreting. These findings suggest that interpreters' poor performance in VRI could be traced back to the note-taking process where cognitive resources are inadequate to conduct listening comprehension, video watching and note-taking at the same time. Therefore, when developing training modules for RI, which is an increasingly popular interpreting mode in the post-pandemic interpreting market, it is essential to take the features of video mediation into consideration and guide students to effectively distribute their attention to different areas on the screen and their notepads. For professional interpreters, this can be more challenging because the visual information presented on the screen can be disorganized and massive (e.g., the speaker's slides, the audience and the meeting room) in their actual interpreting work. Hence, developing the ability to quickly identify the visual information they need on the screen while completing note-taking at the same time becomes a new key requirement for professional interpreters. All in all, at this stage of the pandemic, this would be a good time for those involved in pedagogy to reflect on and learn from the fast-acquired skills in VRI, and for those working with VRI to equip themselves with conflict resolution abilities to balance their effort devoted to the multiple subtasks during interpreting.

Pronounced sequence effects of source speech difficulty are detected on the participants' task difficulty perception, note-taking behaviour and interpretation quality. When the interpreting task starts with a difficult segment, the participants' perception and response to the next interpreting segment are found to be 'distorted'. However, compared with the students, the professionals show greater resilience to these sequence effects during note-taking and interpreting. These findings suggest that, in interpreter training, teachers can design materials that are arranged in different sequences of interpreting difficulty to help students to develop such resilience to the change of task difficulty. In interpreting practice, interpreters should also pay special attention to the potential extra cognitive load caused by a 'distorted' perception of task difficulty.

On the product of note-taking, compared with the students who present some negative correlations between their note choices and interpreting scores, the professionals' interpretation quality is not associated with their choices of note forms or note languages at all. The two groups' great difference in this regard indicates that the students are still bothered by the selection and application of different note categories, especially when the source speech is difficult to interpret, whereas the professionals are not restricted by the form or language of notes anymore. Therefore, automatic use of note-taking resources (such as symbols) is essential in developing note-taking expertise, as it plays an essential role in deciding the amount available of processing capacity for listening analysis during the input stage of CI.

When it comes to note-reading, interpreters usually have to process several notes at the same time, leading to a high cognitive load of note reception. To regulate the cognitive load of note-reading in interpreting practice, it is essential to apply note-taking strategies flexibly and adjust note quantity appropriately. In the present study, the participants pursued large note quantities by following an ellipsis-based approach to note-taking. With a great number of notes produced with a shallow level of SL processing, the participants face heavy cognitive demands during the process of note-reading. As indicated by the correlation test results, note quantity is only positively correlated with the interpretation quality in the easy segments, where the participants could manage to make up for the inadequate source speech comprehension, but not in the difficult ones. Therefore, it is important for interpreters to be aware of the great risks of sticking to an effort-saving strategy and pursuing a large note quantity in note-taking. In other words, securing a great number of notes, which is easier to be achieved with an ellipsis strategy than with other strategies of note-taking, does not represent securing a considerable amount of source information. Without fast and accurate memory retrieval, processing many notes simultaneously can increase interpreters' cognitive load substantially and impede the production of the target speech.

Meanwhile, the observed cognitive trade-offs between different note forms (full words and abbreviations) and note languages (Chinese and English) in the note-taking and note-reading effort further demonstrate that, when making note decisions, it is important to take the cognitive demand for note reception into consideration. In existing note-taking guidelines, much attention has been paid to teaching interpreting learners to use note forms and note languages that are effort-saving in note production, but little has been discussed about the effort of note reception. After all, the ultimate goal of interpreting is not creating many notes with little effort but delivering an accurate and fluent rendering. To achieve that, the cognitive demand for note-reading, which directly decides the amount of processing capacity that is available for target speech production and monitoring, should always be considered in making note decisions. It is worth mentioning that these cognitive trade-offs between different note forms and note languages are more obviously observed in the easy segments than in the difficult segments in the present study. This result suggests that, when the source speech is difficult for interpreters, taking notes in full words or abbreviations, Chinese or English, can entail similar amounts of effort. All in all, the importance of the cognitive demand for note-reading should be emphasized in interpreter training and interpreting practice, for the purpose of helping interpreters to better allocate their limited cognitive resources across the two stages of note-taking. Moreover, when advising a certain type of note form or note language in textbooks, the difficulty level of the source speech should also be considered as an important variable.

Finally, the present study observes various significant negative correlations between the participants' average physical effort in writing one note and their interpretation quality in different aspects. These correlations demonstrate that the "effort of the hand" (S. Chen, 2020b, p. 134) is still an important issue that troubles interpreters during note-taking. Apart from "the mental effort" (p. 134) of note-taking which is emphasized repeatedly in existing textbooks and guidelines, the physical effort of handwriting should be attached more importance in interpreter training. Moreover, special attention should be paid to this manual effort of interpreters while examining their note-taking effort, note decisions and note-reading effort, in future studies.

8.3 Strengths and limitations

In order to provide a more complete account of interpreters' note-taking behaviour in VRI, the present research collected both online and offline data generated by eye-tracking and pen-recording methods to visualize and quantify the process of note-taking, the product of note-taking and the process of note-reading. Such an organic combination of cutting-edge methods and a full coverage of the three stages of note activities in interpreting are recognized as the major strengths of this study. On the one hand, the adoption of eye-tracking and pen-recording methods allows a more comprehensive examination of the topic; on the other hand, the triangulation of a variety of data types consolidates the reliability of the research findings. Especially, the proposal and application of a series of eye-and-pen measures allow a number of significant issues that have not been previously touched upon in related work to be addressed. These include the measurement of the note-taking and note-reading effort from visual, cognitive, physical and temporal aspects, and the exploration of the relationship between the effort devotion in different stages of note activities and the interpretation quality. Moreover, such a combined process-oriented and product-oriented methodology can also make an important contribution to the interdisciplinary development of studies on interpreting and note-taking for general purposes. It provides an innovative and reliable solution for researchers to probe into interpreters' or general note takers' 'black box' during note production and note reception, broadening the interface between Interpreting Studies and neighbouring subjects such as learning sciences and cognitive psychology.

In addition, in previous research on note-taking in CI, only one study (Cardoen, 2018) had taken interpreter work experience and source speech difficulty into consideration. Cardoen (2018) mainly puts her focus on the descriptive features of notes, and the recruited professional interpreters rarely worked in the interpreting market. The present investigation expanded the research scope to the process of the note activities and refined the recruitment of participants by specifying work experience criteria and balancing the interpreters' WM capacity. Findings yielded from the comparisons of two interpreter groups and two task conditions in the present study, which are backed up by solid empirical evidence, thus provide abundant implications for interpreting practice and interpreter training. Moreover, by manipulating the presentation sequence of the speech segments that varied in the difficulty levels of interpreting, the present study reveals a series of sequence effects on the participants' task difficulty perception, note-taking behaviour and interpreting performance. These findings provide important implications for future note-taking and interpreting studies to take the sequence effects of source speech difficulty into consideration, which marks another major strength of this investigation.

The present study is also the first investigation that looks into interpreters' note-taking behaviour in VRI, which is an increasingly popular mode in the post-pandemic era. Through exploring the proportions of the participants' eye fixations on different areas on the screen, the present study finds that, compared with the professional interpreters, the student interpreters attached more importance to collecting and processing visual information presented in the video. By comparison, the professional interpreters put note-taking as a higher priority during the input phase of VRI. Because of these differences in overt visual attention distribution, the professionals yielded larger note quantities and better interpreting performance than the students. Future studies can further look into the reasons behind these observed differences in attention distribution in VRI and explore the optimal attention distribution methods for interpreters to follow in conducting VRI tasks.

While it is the combination of eye-tracking and pen-recording measures that enabled the examination of the note-taking and note-reading processes in the present study, this methodology could add a limitation to the research as the interpreters were asked to take notes with a smart pen on a digital pad rather than with a normal pen on a notepad. The differences between these experimental settings and the actual practice of VRI could bring some impacts on the research findings. Future studies can try to present the video on full screen and adopt a Livescribe pen which is equipped with a built-in camera and a microphone, to film the writing scenes on the paper, record the voice of the interpreter, and upload the recorded videos to computers for further pen-movement analysis. Although this setting would not allow the desktop-based eye-tracker to collect interpreters' eye movements on the paper, it can still show how frequently interpreters refer to the screen during VRI and reveal some switch patterns between the screen and the notepad.

In addition, the duration of each speech segment in the present study was controlled to be less than 3 minutes, because the general duration of speech segments in CI is estimated to be between 45 seconds and 3 minutes (Setton & Dawrant, 2016). However, the overall task duration (around 20 minutes) was not as long as that in some real-life CI scenarios which could last hours or even a whole day. Therefore, the findings in the present study might not be applicable to especially long CI tasks. Since interpreters' fatigue level can directly affect their mental workload and interpreting performance (e.g., Moser-Mercer, 2003), the role of note-taking in these situations needs further investigation.

Finally, as Wang and Zou (2017) demonstrate in a corpus-based CI study, since Chinese and English are non-European and European languages, respectively, there is an issue of language specificity in Chinese-English interpreting. The "wide differences in linguistic structures and cultural conceptualization" (p. 65) between these two languages increase interpreters' cognitive load of interpreting and influence the organization of the target speech. These dramatic differences between the two languages can produce specific difficulties for interpreters to conduct note-taking and note-reading in Chinese-English interpreting. For instance, word order differences are frequently observed in Chinese and English languages (Ma et al., 2021). If interpreters intend to complete word order reorganization during the input phase of CI, they will have to wait for more source information, restructure the words, and postpone note-writing. In addition, as proven in this study, noting in Chinese generally involves more physical effort of handwriting and overt visual attention during note-taking. These differences in the note-taking effort between noting in Chinese and noting in English could also affect interpreters' note choices to a great extent. Since the present study only investigated English-Chinese CI with Chinese native interpreters, future research needs to re-examine these reported findings in other language pairs with interpreters from different language backgrounds.

8.4 Avenues for future research

As the present study focuses on note activities in RI, it would be of interest for future research to have these activities examined in on-site CI. By comparing how interpreters conduct note-taking in online and offline modes of interpreting, valuable suggestions can be provided for interpreter trainers to design their note-taking training sessions for online and offline interpreting tasks, and for professional interpreters to cope with a remote mode of note-taking.

Furthermore, to facilitate the analysis of the eye-tracking data on the screen, the present study did not allow the participants to consult external resources such as e-dictionaries, machine translation and terminology bank during the process of interpreting. However, this might deviate a little from the reality where interpreters can flexibly check their prepared materials and have easy access to online resources during interpreting. The involvement of consultation can certainly affect the way that interpreters allocate their time and effort to note activities during the input and output phases of CI. Therefore, in the extended study of this doctoral research project, consultation will be included into the experimental design to further explore how interpreters balance video-watching, note-taking and consulting during the input stage of VRI.

In addition, as observed in the present study, the participants conducted note-reading in note groups. This result suggests that the note group can serve as an important unit of research in future studies. Many interesting research questions can be raised from this perspective. For instance, it is worth investigating how interpreters decide the volume of source information in one note group, and how this group formation affects their memory recall and speech reconstruction in the second phase of CI. This can be affected by the interpreters' work experience, as in translation studies where professional translators are found with bigger translation units than student translators (Jakobsen, 2003). It might also be influenced by the difficulty level of the source speech, as Dam (1998) finds that interpreters opt for meaning-based interpreting in difficult CI tasks and form-based interpreting in easy ones. In summary, by investigating interpreters' note groups, researchers can gain new insights into how the source speech contents are transcoded into written notes, and how the meaning of the note groups is redelivered in the target speech.

Bibliography

Abuín González, M. (2012). The language of consecutive interpreters' notes: Differences across levels of expertise. *Interpreting*, *14*(1), 55-72.

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science*, *205*(4412), 1289-1292.

AIIC. (2020). *AIIC interpreter checklist: Performing remote interpreting Assignments from home in extremis during the Covid-19 pandemic*. Retrieved May 6, 2021, from https://aiic.ch/wp-content/uploads/2020/05/aiic-interpreter-checklist.pdf.

Albl-Mikasa, M. (2006). Reduction and expansion in notation texts. In C. Heine, K. Schubert & H. Gerzymisch-Arbogast (Eds.), *Text and translation: Theory and methodology of translation* (pp.195-214). Narr.

Albl-Mikasa, M. (2008). (Non-)Sense in note-taking for consecutive interpreting. *Interpreting*, *10*(2), 197-231.

Alexieva, B. (1994). On teaching note-taking in consecutive interpreting. In C. Dollerup & A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims, visions* (pp. 199-206). John Benjamins.

Andres, D. (2002). *Konsekutivdolmetschen und notation*. Peter Lang.

Angelelli, C. V. (2009). Using a rubric to assess translation ability: Defining the construct. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 13-47). John Benjamins. https://doi.org/10.1075/ata.xiv.03ang

Arumí Ribas, M. (2012). Problems and strategies in consecutive interpreting: A pilot study at two different stages of interpreter training. *Meta*, *57*(3), 812-835.

Baddeley, A. & Hitch, G. J. (1974) Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47-89). Academic Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in Cognitive Sciences*, *4*(11), 417-423.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575-589.

Baddeley, A., & Andrade, J. (1994). Reversing the word-length effect: A comment on Caplan, Rochon, and Waters. *The Quarterly Journal of Experimental Psychology*, *47*(4), 1047-1054.

Barbier, M.L., Piolat, A., Roussey, J.Y., & Olive, T. (2006). Note-taking in second language: Language procedures and self-evaluation of the difficulties. *Current Psychology Letters. Brain, Behavior and Cognition*, *20(*3). http://cpl.revues.org/document1283.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*(4), 553-573.

Baumeister, R. F., & Heatherton, T. F. (1996). Self-regulation failure: An overview. *Psychological Inquiry, 7,* 1-15.

Becker, W. (1972). *Notizentechnik*. BBK.

Bendazzoli, C., Sandrelli, A., & Russo, M. (2011). Disfluencies in simultaneous interpreting: A corpus-based analysis. In A. Kruger, K. Wallmach & J. Munday (Eds.), *Corpus-based translation studies: Research and applications* (pp. 282-306). Continuum.

Błaszczyk, P., & Hanusiak, D. (2010), The choice of language note-taking for consecutive interpreting: A Polish perspective. Electronic proceedings of the KäTu Symposium on Translation and Interpreting Studies 4. Retrieved July 20, 2022, from https:// www.sktl.fi/@Bin/40680/Blaszczyk%26Hanusiak_MikaEL2010.pdf.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*, 177-192.

Braun, S., & Taylor, J. (2012). *Videoconference and remote interpreting in criminal proceedings*. Intersentia.

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, *40*(2), 540-545.

Cai, R., Dong, Y., Zhao, N., & Lin, J. (2015). Factors contributing to individual differences in the development of consecutive interpreting competence for beginner student interpreters. *The Interpreter and Translator Trainer*, *9*(1), 104-120.

Campbell, S. & Hale, S. (2003). Translation and interpreting assessment in the context of educational measurement. In G. Anderman & M. Rogers (Eds.), *Translation today: Trends and perspectives* (pp. 205-224). Multilingual Matters.

Campos, P. P., & Salinas, M. J. V. (2017). La necesidad de una técnica de toma de notas en interpretación consecutiva: una experiencia. *HERMES-Journal of Language and Communication in Business*, *55*, 171-192.

Cardoen, H. (2018). *Efficient note-taking in consecutive interpreting: Qualitative notes as a polymorph tool* [PhD Thesis, University of Mons]. Mons.

Chang, C. C., & Schallert, D. L. (2007). The impact of directionality on Chinese/English simultaneous interpreting. *Interpreting*, *9*(2), 137-176.

Chang, L.W. (2015). *Investigating note-taking in consecutive interpreting-Using the concept of visual* grammar [PhD Thesis, The University of Manchester]. Manchester.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55-81.

Chen, H. (2012). The moderating effects of item order arranged by difficulty on the relationship between test anxiety and test performance. *Creative Education*, *3*(3), 328-333.

Chen, S. (2016). Note taking in consecutive interpreting: A review with special focus on Chinese and English literature. *The Journal of Specialised Translation*, *26*(1), 151-171.

Chen, S. (2017a). *Exploring the process of note-taking and consecutive interpreting: A pen-eye-voice approach towards cognitive load* [PhD Thesis, Macquarie University]. Sydney.

Chen, S. (2017b). Note-taking in consecutive interpreting: New data from pen recording. *Translation & Interpreting, 9*(1), 4-23.

Chen, S. (2017c). The construct of cognitive load in interpreting and its measurement. *Perspectives*, *25*(4), 640-657.

Chen, S. (2018). A pen-eye-voice approach towards the process of note-taking and consecutive interpreting: An experimental design. *International Journal of Comparative Literature and Translation Studies*, *6*(2), 1-8.

Chen, S. (2020a). The impact of directionality on the process and product in consecutive interpreting between Chinese and English: Evidence from pen recording and eye tracking. *Journal of Specialised Translation*, *34*, 100-117.

Chen, S. (2020b). The process of note-taking in consecutive interpreting: A digital pen recording approach. *Interpreting*, *22*(1), 117-139.

Chen, S. (2022). The process and product of note-taking and consecutive interpreting: empirical data from professionals and students. *Perspectives*, *30*(2), 258-274.

Chen, S., Kruger, J. L., & Doherty, S. (2021). Reading patterns and cognitive processing in an eye-tracking study of note-reading in consecutive interpreting. *Interpreting*, *23*(1), 76-102.

Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. *Written Communication*, *20*(1), 99-118.

Chmiel, A. (2010). How effective is teaching note-taking to trainee interpreters?. *The Interpreter and Translator Trainer*, *4*(2), 233-250.

Chmiel, A. (2021). Eye-tracking studies in conference interpreting. In M. Albl-Mikasa, & E. Tiselius (Eds.), *The Routledge handbook of conference interpreting* (pp. 457-470). Routledge.

Chmiel, A., & Lijewska, A. (2019). Syntactic processing in sight translation by professional and trainee interpreters: Professionals are more time-efficient while trainees view the source text less. *Target*, *31*(3), 378-397.

Chuang, L. (2008). Note-taking know-how: A processing perspective on consecutive interpreting. *Spectrum: Studies in Language, Literature, Translation, and Interpretation 2*, 93-101.

Clarke, R. J., Zhang, H., & Gamlin, P. D. (2003). Characteristics of the pupillary light reflex in the alert rhesus monkey. *Journal of Neurophysiology*, *89*(6), 3179-3189.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684.

Cui, Y., & Zheng, B. (2021). Consultation behaviour with online resources in English-Chinese translation: An eye-tracking, screen-recording and retrospective study. *Perspectives*, *29*(5), 740-760.

Da Silva, I. L., Schmaltz, M., Alves, F., Pagano, A., Wong, D., Chao, L., Leal, A. L. V. Quaresma, P. & Garcia, C. (2015). Translating and post-editing in the Chinese-Portuguese language pair: Insights from an exploratory study of key logging and eye tracking. *Translation Spaces*, *4*(1), 145-169.

Dai, W. 戴炜栋, & Xu, H. 徐海铭. (2007). 汉英交替传译过程中译员笔记特征实证研究——以职业受训译员和非职业译员为例 [An empirical study of the features of interpreters' notes in Chinese-English consecutive interpreting: The examples of professionally trained and unprofessional interpreters]. *Foreign Language Teaching and Research, 39*(2), 136-144.

Dam, H. V. (1998). Lexical similarity vs lexical dissimilarity in consecutive interpreting: A product-oriented study of form-based vs meaning-based interpreting. *The Translator*, *4*(1), 49-68.

Dam, H. V. (2001). On the option between form-based and meaning-based interpreting: The effect of source text difficulty on lexical target text form in simultaneous interpreting. *The Interpreters' Newsletter*, *11*, 27-55.

Dam, H. V. (2004a). Interpreters' notes: On the choice of language. *Interpreting*, *6*(1), 3-17.

Dam, H.V. (2004b). Interpreters' notes: On the choice of form and language. In G. Hansen, K. Malmkjær & D. Gile (Eds.), *Claims, changes and challenges in Translation Studies* (pp. 251-261). John Benjamins.

Dam, H.V. (2007). What makes interpreters' notes efficient? Features of (non-)efficiency in interpreter's notes for consecutive. In Y. Gambier, M. Shlesinger & R. Stolze (Eds.), *Doubts and directions in Translation Studies: Selected contributions from the EST Congress, Lisbon 2004* (pp. 183-197). John Benjamins. https://doi.org/10.1075/btl.72.19dam

Dam, H.V., Engberg, J. & Schjoldager, A. (2005). Modelling semantic networks on source and target texts in consecutive interpreting: A contribution to the study of interpreters' notes. In H.V. Dam, J. Engberg & H. Gerzymisch-Arbogast (Eds.), *Knowledge Systems and Translation* (pp. 227-254). Mouton de Gruyter.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466.

De Beaugrande, R. A., & Dressler, W. U. (1981). *Introduction to text linguistics* (Vol. 1). Longman.

Díaz-Galaz, S., Padilla, P., & Bajo, M. T. (2015). The role of advance preparation in simultaneous interpreting: A comparison of professional interpreters and interpreting students. *Interpreting*, *17*(1), 1-25.

Doherty, S., Martschuk, N., Goodman-Delahunty, J., & Hale, S. (2022). An eye-movement analysis of overt visual attention during consecutive and simultaneous interpreting modes in a remotely interpreted investigative interview. *Frontiers in Psychology*, 1182. doi: 10.3389/fpsyg.2022.764460

Dong, Y., & Cai, R. (2015). Working memory and interpreting: a commentary on theoretical models. In Z. Wen, M. B. Mota & A. McNeill (Eds.), Working memory in second language acquisition and processing (pp. 63-81). Multilingual Matters.

Dong, Y., & Li, P. (2020). Attentional control in interpreting: A model of language control and processing control. *Bilingualism: Language and Cognition*, *23*(4), 716-728.

Dong, Y., & Liu, Y. (2016). Classes in translating and interpreting produce differential gains in switching and updating. *Frontiers in Psychology, 7,* 1297. doi:10.3389/fpsyg.2016.01297.

Dragsted, B. & Hansen, I.G. (2009). Exploring translation and interpreting hybrids. The case of sight translation. *Meta*, *54*(3), 588-604. https://doi.org/10.7202/038317ar

Dragsted, B. (2010). Coordination of reading and writing processes in translation: An eye on uncharted territory. In G.M. Shreve & E. Angelone (Eds.), *Translation and cognition*. (pp. 41-62). John Benjamins. https://doi.org/10.1075/ata.xv.04dra

Drieghe, D., Pollatsek, A., Staub, A., & Rayner, K. (2008). The word grouping hypothesis and eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1552-1560.

DuBay, W.H. (2004). The principles of readability. Retrieved May 30, 2022, from http://www.impactinformation.com/impactinfo/readability02.pdf

Duman, D. (2014). Training in note taking for consecutive interpreting: A qualitative analysis. *Proceedings of the XIIth International Language, Literature and Stylistics Symposium*, 441-444.

Ericsson, K.A. & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8)*,* 725-747.

European Commission. (2020). *European language industry survey*. Retrieved August 16, 2021, from https://ec.europa.eu/info/sites/default/files/2020_language_industry_survey_report.pdf

Feng, J. 冯佳. (2018). *中英双向互译中翻译认知过程研究: 基于眼动追踪和键盘记录实证分析* [Cognition process in E-C and C-E translation: An eye-tracking and key-logging study of directionality]. Foreign Language Teaching and Research Press.

Fiske, S. T., & Neuberg, S. L. (1988). A continuum model of impression formation: From category-based to individuating processes as a function of information, motivation, and attention. *Advances in Experimental Social Psychology, 23,* 1-74.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221-233.

Foulin, J-N. (1995). Pauses et debits: Les indicateurs temporels de la production écrite. *L'année Psychologique*, *95*(3), 483-504. https://doi.org/10.3406/psy.1995.28844

Gao, B. 高彬. (2019). 英汉交替传译笔记中的语言选择发展规律—基于口译学习者的横向研究 [Evolvement of the choice of note language in English-Chinese consecutive interpreting—A Cross-sectional study on interpreting learners]. *Chinese Translators Journal*, *40*(1), 83-90.

Gile, D. (1991). Prise de notes et attention en début d'apprentissage de l'interprétation consecutive-une experience-démonstration de. *Meta*, *36*(2-3), 431-439.

Gile, D. (1995/2009). *Basic concepts and models for interpreter and translator training*. John Benjamins.

Gile, D. (1999). Testing the Effort Models' tightrope hypothesis in simultaneous interpreting: A contribution. *Hermes*, *23*, 153-171.

Gile, D. (2001). Consecutive vs. simultaneous: Which is more accurate. *Interpretation Studies*, *1*(1), 8-20.

Gile, D. (2008). Local cognitive load in simultaneous interpreting and its implications for empirical research. *Forum*, *6*, 59-77.

Gile, D. (2020, Sept. 29). *The effort models and gravitational model: Clarifications and update*. Retrieved May 6, 2021, from https://painel.translators101.com.br/wp-content/uploads/2021/03/The-Effort-Models-and-Gravitational-Model-Clarifications-and-update.pdf

Gile, D. (2021). 认知负荷模型在口译教学中的建构 [The effort models of interpreting as a didactic construct] (B. Lei Trans.). *Translation Horizons*, 11, 1-24.

Gillies, A. (2017). *Note-taking for consecutive interpreting: A short course*. St. Jerome.

Gran, L. (1982). *L'annotazione grafica nell'interpretazione consecutiva*. Università degli Studi di Trieste.

Gran, L. (1989). Interdisciplinary research on cerebral asymmetries: Significance and prospects for the teaching of interpretation. In L. Gran & J. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation* (pp. 93-100). Campanotto Editore.

Green, A. (2014). *Exploring language assessment and testing*. Routledge.

Gumul, E., & Łyda, A. (2007). The time constraint in conference interpreting: Simultaneous vs. consecutive. *Research in Language*, *5*, 165-183.

Hale, S., Goodman-Delahunty, J., Martschuk, N., & Doherty, S. (2021). The effects of mode on interpreting performance in a simulated police interview. *Translation and Interpreting Studies*. https://doi.org/10.1075/tis.19081.hal

Han, C. (2015). Investigating rater severity/leniency in interpreter performance testing: A multifaceted Rasch measurement approach. *Interpreting*, *17*(2), 255-283. https://doi.org/10.1075/intp.17.2.05han

Han, C. (2018). Using rating scales to assess interpretation: Practices, problems and prospects. *Interpreting*, *20*(1), 59-95.

Han, C. (2019). A generalizability theory study of optimal measurement design for a summative assessment of English/Chinese consecutive interpreting. *Language Testing*, *36*(3), 419-438.

Han, X. (2013). The skill-focused approach to interpretation teaching: An empirical exploration. *Open Journal of Modern Linguistics*, *3*(2), 161-165.

Harrison, C. (1980). *Readability in the classroom*. Cambridge University Press.

Her, H. (2001). Notetaking in basic interpretation class: An initial investigation. *Studies of Translation and Interpretation*, *6*, 53-77.

Hiltunen, S., Pääkkönen, R., Vik, G. V., & Krause, C. M. (2014). On interpreters' working memory and executive control. *International Journal of Bilingualism*, *20*(3), 297-314.

Hoffman, R. R. (1997). The cognitive psychology of expertise and the domain of interpreting. *Interpreting*, *2*(1), 189–230.

Hradilová, H. (2019). Výuka tlumočnické notace ve spolkové republice německo. [Master's thesis, Charles University]. Prague.

Hu, L. (2008). *Note-taking in consecutive interpreting: A psycholinguistic perspective* [PhD Thesis, Shanghai International Studies University]. Shanghai.

Hui, M. (2019). A focus-group study: students' perceived benefits from learning with practitioners' note-taking and consecutive interpreting demonstration videos. *Current Trends in Translation Teaching and Learning*, *6*, 105–159.

Hunter, D., Jones, R. & Randhawa, B. (1996). The use of holistic versus analytic scoring for largescale assessment of writing. *The Canadian Journal of Program*

Evaluation, 11(2), 61–85.

Hvelplund, K. T. (2011). *Allocation of cognitive resources in translation: An eye-tracking and key-logging study* [PhD thesis, Copenhagen Business School]. Copenhagen.

Hvelplund, K. T. (2014). Eye tracking and the translation process: Reflections on the analysis and interpretation of eye-tracking data. In Muñoz Martín, R. (Ed.), *MonTI special issue 1: Con la traducción en mente* (pp. 201-223). Publicaciones de la Universided de Alicante.

Hvelplund, K.T. (2019). Digital resources in the translation process–attention, cognitive effort and processing flow. *Perspectives*, *27*(4), 510-524.

Hyönä, J., Lorch, R.F., Jr & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyona, R. Radach & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313-334). Elsevier. https://doi.org/10.1016/B978-044451020-4/50018-9

Ilg, G. (1988). La prise de notes en interprétation consécutive. Une orientation générale. *Parallèles 9*, 9–13.

Ilg, G., & Lambert, S. (1996). Teaching consecutive interpreting. *Interpreting*, *1*(1), 69-99.

Jakobsen, A. L. & Jensen, K. T. H. (2008). Eye movement behaviour across four different types of reading task. In S. Gopferich, A.L. Jakobsen, & I.M. Mees (Eds). *Looking at eyes: Eyetracking studies of reading and translation processing*. (pp. 103-122). Samfundslitteratur.

Jakobsen, A. L. (2003). Effects of think aloud on translation speed, revision and segmentation. In F. Alves (Ed.), *Triangulating translation. Perspectives in process oriented research* (pp. 69-95). John Benjamins.

Jensen, K. T. H. (2009). Indicators of text complexity. In S. Göpferich, A. L. Jakobsen & I. M. Mees (Eds.), *Behind the mind: Methods, models and results in translation process research* (pp. 61-80). Samfundslitteratur.

Jiménez Ivars, A., Pinazo Catalayud, D., & Ruiz i Fores, M. (2014). Self-efficacy and language proficiency in interpreter trainees. *The Interpreter and Translator Trainer*, *8*(2), 167-182.

Johannsen, G. (1979). Workload and workload measurement. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 3-11). Plenum Press.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge University Press.

Jones, R. (2002). *Conference interpreting explained*. St. Jerome Publishing.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329-354.

Kahneman, D. (1973). *Attention and effort.* Prentice-Hall.

Kantowitz, B. H., & Knight Jr, J. L. (1976). Testing tapping timesharing, II: Auditory secondary task. *Acta Psychologica*, *40*(5), 343-362.

Kieras, D. E., & Bovair, S. (1986). The acquisition of procedures from text: A production system analysis of transfer of training. *Journal of Memory and Language*, 25, 507-524.

Kim, N. H. (2009). The role of note-taking and coordination effort during consecutive interpretation: a case study in German-Korean consecutive interpretation. *Journal of Interpretation & Translation Research Institute*, *12*(2), 51-78.

Kintsch, W. (1974). *The representation of meaning in memory*. Erlbaum.

Kirchhoff, H. (1979). Die notationssprache als hilfsmittel des konferenzdolmetschers im konsekutivvorgang. In W. Mair & E. Sallager (Eds.), *Sprachtheorie und ssprachpraxis* (pp. 121-133). Gunter Narr.

Klare, G. R. (1963). *The measurement of readability*. Iowa State University Press.

Ko, L. (2006). The need for long-term empirical studies in remote interpreting research: a case study of telephone interpreting. *Linguistica Antverpiensia*, *5*, 325-338.

Ko, L., & Chen, N. S. (2011). Online-interpreting in synchronous cyber classrooms. *Babel*, *57*(2), 123–143. https:// doi.org/10.1075/babel.57.2.01ko

Kohn, K. & Albl-Mikasa, M. (2002). Note-taking in consecutive interpreting: On the reconstruction of an individualised language. *Linguistica Antverpiensia*, *1*, 257-272.

Kolinsky, R. (1998). Spoken word recognition: A stage-processing approach to language differences. *European Journal of Cognitive Psychology*, *10*(1), 1-40.

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

Korpal, P., & Jasielska, A. (2019). Investigating interpreters' empathy: Are emotions in simultaneous interpreting contagious?. *Target*, *31*(1), 2-24.

Korpal, P., & Stachowiak-Szymczak, K. (2018). The whole picture: Processing of numbers and their context in simultaneous interpreting. *Poznan Studies in Contemporary Linguistics*, *54*(3), 335-354.

Korpal, P., & Stachowiak-Szymczak, K. (2020). Combined problem triggers in simultaneous interpreting: Exploring the effect of delivery rate on processing and rendering numbers. *Perspectives*, *28*(1), 126-143.

Kostopoulou, G. (2007). The role of coherence in text approaching and comprehension: Applications in translation didactics. *Meta*, *52*(1), 146-155.

Kress, G., & Van Leeuwen, T. (2006). *The grammar of visual design*. Routledge.

Ladas, H. S. (1980). Note taking on lectures: An information-processing approach. *Educational Psychologist*, *15*(1), 44-53.

Lee, J. & Choi, M. (2012). A pedagogy for effective notetaking for consecutive interpreting: A case study of notetaking for English-Korean consecutive interpreting. *Interpreting and Translation Studies*, *16*(4),169-195.

Lee, S. B. (2015). Developing an analytic scale for assessing undergraduate students' consecutive interpreting performances. *Interpreting*, *17*(2), 226-254.

Lee, T. H. (2002). Ear voice span in English into Korean simultaneous interpretation. *Meta*, *47*(4), 596-606.

Lei, V.& Li, D. (2019, July 4-6). *Tracking where the professional conference interpreter looks: A case study in naturalistic settings* [Conference session]. Second International Congress on Translation, Interpreting and Cognition, Germersheim, Germany.

Levy, C. M., & Ransdell, S. (1995). Is writing as difficult as it seems?. *Memory & Cognition*, *23*(6), 767-779.

Li J. 李晋, Zhang, W. 张威, & Chen, W. 陈文荣. (2022). 交替传译笔记的视觉语法对译员口译产出质量影响研究 [A study on the influence of the visual grammar of consecutive interpreting note-taking on the quality of interpreting]. *Foreign Languages and Their Teaching, 325*(4), 36-47.

Lin, Y., Lv, Q., & Liang, J. (2018). Predicting fluency with language proficiency, working memory, and directionality in simultaneous interpreting. *Frontiers in Psychology*, *9*, 1543. doi: 10.3389/fpsyg.2018.01543.

Liu, J. 刘建军. (2010). 英语专业本科学生汉英交传笔记特征——一项基于学生交传笔记的实证研究 [Note-taking characteristics of English majored undergraduates in Chinese-English consecutive interpreting: An empirical study based on students' consecutive interpreting notes]. *Foreign Language World*, *2*, 47-54.

Liu, M., & Chiu, Y. H. (2009). Assessing source material difficulty for consecutive interpreting: Quantifiable measures and holistic judgment. *Interpreting*, *11*(2), 244-266.

Liu, M., Schallert, D. L., & Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting*, *6*(1), 19-42.

Loch, B. I. (2005 November 22-26). *Tablet technology in fifirst year calculus and linear algebra teaching* [Conference session]. Kingfifisher delta '05: 5th southern hemisphere Conference on Undergraduate mathematics and statistics teaching and learning, Fraser Island, Australia. http://eprints.usq.edu.au/archive/000000185

Locke, J. L., & Fehr, F. S. (1972). Subvocalization of heard or seen words prior to spoken or written recall, *85*(1), *The American Journal of Psychology*, 63-68.

Lockhart, R. S., & Craik, F. I. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology, 44*(1), 87-112.

Lowenstein, O., & Loewenfeld, I. E. (1962). The eye. In *Muscular mechanisms* (pp. 231-267). Academic Press.

Lu, Z. 卢植, & Zheng, Y. 郑有耀. (2021). 英汉隐喻视译过程中注意资源分配的眼动实验研究[An eye-tracking study on attention resources allocation in the process of sight-translation of English to Chinese metaphors]. *Foreign Language Research*, *222*(5), 72-79.

Lung, R. (1999). Note-taking skills and comprehension in consecutive interpretation. *Babel*, *45*(4), 311-317.

Lung, R. (2003). Taking "notes" seriously in the interpretation classroom. In Á. Collados Aís, M. M. Fernández Sánchez & D. Gile (Eds.), *La evaluación de la calidad en interpretación: investigación* (pp. 199-205). Comares.

Lv, Q., & Liang, J. (2019). Is consecutive interpreting easier than simultaneous interpreting? A corpus-based study of lexical simplification in interpretation. *Perspectives*, *27*(1), 91-106.

Ma, X., Li, D., & Hsu, Y. Y. (2021). Exploring the impact of word order asymmetry on cognitive load during Chinese-English sight translation: Evidence from eye-movement data. *Target*, *33*(1), 103-131.

Ma, X., Li, D., Tsai, J. L., & Hsu, Y. Y. (2022). An eye-tracking based investigation into reading behavior during Chinese-English sight translation: The effect of word order asymmetry. *Translation & Interpreting*, *14*(1), 66-83.

Mackintosh, J. (1985). The Kintsch and Van Dijk model of discourse comprehension and production applied to the interpretation process. *Meta*, *3*0(1),37-43.

Marquardt, T. P., & Gillam, R. B. (1999). Assessment in communication disorders: Some observations on current issues. *Language Testing*, *16*(3), 249-269.

Matyssek, H. (1989). *Handbuch der notizentechnik für.* Julius Groos.

Mead, P. (2000). Control of pauses by trainee interpreters in their A and B languages. *The Interpreters' Newsletter*, *10*, 89-102.

Mead, P. (2005). Directionality and fluency: An experimental study of pausing in consecutive interpretation into English and Italian. *Communication and Cognition. Monographies, 38*(1-2), 127-146.

Meijman, T. F., & O'Hanlon, J. F. (1984). Workload. An introduction to psychological theories and measurement methods. In P. J. D. Drenth, H. Thierry, P. J. Willems & de Wolff, C. J. (Eds.), *Handbook of work and organizational psychology* (pp. 257-288). Wiley.

Meshkati, N. (1988). Toward development of a cohesive model of workload. In P. A.

Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 305–314). North–Holland.

Moser-Mercer, B. (2003): *Remote interpreting: Assessment of human factors and performance parameters.* Retrieved July 21, 2021, from <http://www.aiic.net/ViewPage.cfm?page_id=1125#methods.

Moser-Mercer, B. (2005). Remote interpreting: issues of multi-sensory integration in a multilingual task. *Meta*, *50*(2), 727-738.

Moser-Mercer, B. (2008). Skill acquisition in interpreting. *The Interpreter and Translator Trainer*, *2*(1), 1–28. doi:10.1080/1750399X.2008.10798764

Mouzourakis, P. (2006). Remote interpreting: a technical perspective on recent experiments. *Interpreting*, *8*(1), 45-66.

Mozer, M. C., Kinoshita, S., & Shettel, M. (2007). Sequential dependencies in human behavior offer insights into cognitive control. *Integrated Models of Cognitive Systems*, *1*, 180-193.

Murray, W. S. (2000). Sentence processing: Issues and measures. In A. Kennedy, R. Radach, D. Heller & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 649-664). Elsevier. https://doi.org/10.1016/B978-008043642-5/50030-9

Napier, J., Skinner, R., & Braun, S. (2018). *Here or there. Research on interpreting via video link*. Gallaudet University Press.

Nimmo, L. M., & Roodenrys, S. (2002). Syllable frequency effects on phonological short-term memory tasks. *Applied Psycholinguistics*, *23*(4), 643-659.

North, R. (1977). Task functional demands as factors in dual task performance. Proceedings of the 21st Annual Meeting of the Human Factors Society. Retrieved May 20, 2022, from https://doi.org/10.1177/107118137702100504

O'Donnell, C. R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. II: Cognitive processes and performance* (pp. 42.41–42.49). John Wiley & Sons.

Olalla-Soler, C. (2018). Using electronic information resources to solve cultural translation problems. Differences between students and professional translators. *Journal of Documentation*, *74*(6), 1293–1317.

Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high-and low-level production processes in written composition. *Memory & Cognition*, *30*(4), 594-600.

Orlando, M. (2010). Digital pen technology and consecutive interpreting: Another dimension in note-taking training and assessment. *The Interpreters' Newsletter 15*, 71-86.

Orlando, M. (2014). A study on the amenability of digital pen technology in a hybrid mode of interpreting: Consec-simul with notes. *Translation & Interpreting*, *6*(2), 39-54.

Orlando, M. (2015). Digital pen technology and interpreting training, practice and research: Status and trends. In S. Erlich & J. Napier (Eds.), *Interpreter education in the digital age: Innovation, access and change* (pp.125–152). Gallaudet University Press.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371. https://doi.org/10.1007/BF02213420

Padilla, P. (1995). *Procesos de memoria y atención en la interpretación de lenguas* [PhD thesis, Universidad de Granada]. Granada.

Pavlović, N. & Kristian T. H. J. (2009). Eye tracking translation directionality. In A. Pym & A. Perekrestenko (Eds.), *Translation research projects 2* (pp. 93-109). Intercultural Studies Group.

Peverly, S. T., Garner, J. K., & Vekaria, P. C. (2014). Both handwriting speed and selective attention are important to lecture note-taking. *Reading and Writing*, *27*(1), 1-30.

Peverly, S. T., Ramaswamy, V., Brown, C., Sumowski, J., Alidoost, M., & Garner, J. (2007). What predicts skill in lecture note taking?. *Journal of Educational Psychology*, *99*(1), 167-180.

Piolat, A. (2007). Effects of notetaking technique and working memory span on cognitive effort and recall performance. In M. Torrance, & L. van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 109–124). Elsevier.

Piolat, A., Kellogg, R. T., & Farioli, F. (2001). The triple task technique for studying writing processes: on which task is attention focused?. *Current Psychology Letters. Brain, Behavior and Cognition*, *4*, 67-83.

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology*, *19*(3), 291-312.

Piolat, A., Olive, T., Roussey, J. Y., Thunin, O., & Ziegler, J. C. (1999). SCRIPTKELL: A tool for measuring cognitive effort and time processing in writing and other complex cognitive activities. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 113-121.

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review*, *31*(2), 339-359.

Plevoets, K., & Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting: A corpus-based regression analysis. *Translation and Interpreting Studies*, *11*(2), 202-224.

Pöchhacker, F. (2004). *Introducing interpreting studies*. Routledge.

Poock, G. K. (1973). Information processing vs pupil diameter. *Perceptual and Motor Skills*, *37*(3), 1000-1002. https://doi.org/10.1177/003151257303700363.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., Slattery, T.J. & Drieghe, D. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514–528.

Rogers R. D., & Monsell S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207-231

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019–1034. https://doi.org/10.1037/0278-7393.28.6.1019

Rosiers, A., Eyckmans, J. & Bauwens, D. (2011). A story of attitudes and aptitudes? Investigating individual difference variables within the context of interpreting. *Interpreting*, *13*(1), 53-69. https://doi.org/10.1075/intp.13.1.04ros

Rozan, J. (1956). *La prise de notes en interprétation consécutive*. Georg.

Rozan, J. (2002). *Note-taking in consecutive interpreting* (A. Gillies, Trans.). Tertium Society for the Promotion of Language Studies. (Original work published 1956).

Roziner, I., & Shlesinger, M. (2010). Much ado about something remote: Stress and performance in remote interpreting. *Interpreting*, *12*(2), 214-247.

Russell, D. & Takeda, K. (2015). Consecutive interpreting. In H. Mikkelson & R. Jourdenais (Eds.), *The Routledge handbook of interpreting* (pp. 96-111). Routledge.

Russell, D. (2005). Consecutive and simultaneous interpreting. In T. Janzen (Ed.), *Topics in signed language interpreting* (pp. 135-164). John Benjamins.

Russell, I. J., Hendricson, W. D., & Herbert, R. J. (1984). Effects of lecture information density on medical student achievement. *Journal of Medical Education*, *59*(11), 881–889.

Sakamoto, A. (2011). An empirical study on the use of note-taking for Consecutive Interpreting in the teaching of written translation. *The Journal of Specialised Translation*, *16*, 203-223.

Salaets, H., & Theys, L. (2016). Link words in note-taking and student interpreter performances: An empirical study. *International Journal of Interpreter Education*, *8*(2), 20-33.

Santorini, B. (1995). *Part-of-speech tagging guidelines for the Penn Treebank Project.* University of Pennsylvania. Retrieved March 27, 2008, from ftp://ftp.cis.upenn.edu/pub/treebank/doc/tagguide.ps.gz.

Scardamalia, M., & Bereiter, C. (1982). Assimilative processes in composition planning. *Educational Psychologist*, *17*(3), 165-171.

Schnotz, W. (1994). *Aufau von wissensstrukturen: Untersuchungen zur kohärenzbildung beim wwissenserwerb mit ttexten*. Beltz.

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories-new models. *Interpreting*, *13*(2), 176-204.

Seeber, K. G. (2013). Cognitive load in simultaneous interpreting: Measures and methods. *Target*, *25*(1), 18-32.

Seeber, K. G., & Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism*, *16*(2), 228-242.

Seleskovitch, D. (1975). *Langage, langues et mémoire: Étude de la prise de notes en iinterprétation consécutive*. Minard Lettres Modernes.

Setton, R. & Dawrant, A. (2016). *Conference interpreting: A trainer's guide*. John Benjamins.

Setton, R. (2002). Seleskovitch: A radical pragmatist before her time. *The Translator*, *8*(1), 117-124.

Shen, M., & Liang, J. (2021). Self-repair in consecutive interpreting: similarities and differences between professional interpreters and student interpreters. *Perspectives*, *29*(5), 761-777.

Sherman, J. L., & Kulhavy, R. W. (1978). Abstractness and prose comprehension. *Acta Psychologica*, *42*(1), 59-65.

Sjørup, A. C. (2013). *Cognitive effort in metaphor translation: An eyetracking and key-logging study* [PhD thesis, Copenhagen Business School]. Copenhagen. http://hdl.handle.net/10419/208853

Slotte, V., & Lonka, K. (1999). Review and process effects of spontaneous note-taking on text comprehension. *Contemporary Educational Psychology*, *24*(1), 1-20.

Someya, Y. (2017). A propositional representation theory of consecutive notes and notetaking. In Y. Someya (Ed.), *Consecutive notetaking and interpreter training* (pp. 147-190). Routledge.

Stachowiak-Szymczak, K., & Korpal, P. (2019). Interpreting accuracy and visual processing of numbers in professional and student interpreters: An eye-tracking study. *Across Languages and Cultures*, *20*(2), 235-251.

Su, W., & Li, D. (2020). Exploring processing patterns of Chinese-English sight translation: An eye-tracking study. *Babel*, *66*(6), 999-1024.

Sun, S., & Shreve, G. M. (2014). Measuring translation difficulty: An empirical study. *Target*, *26*(1), 98-127.

Szabó, C. (2006). Language choice in note-taking for consecutive interpreting. *Interpreting*, *8*(2), 129-147.

Tabachnick, B. G., & Linda S. F. (2007). *Using multivariate statistics (5th edition)*. Pearson.

Timarová, S., Čeňková, I., Meylaerts, R., Hertog, E., Szmalec, A., & Duyck, W. (2014). Simultaneous interpreting and working memory executive control. *Interpreting*, *16*(2), 139-168.

Tippets, E., & Benson, J. (1989). The effect of item arrangement on test anxiety. *Applied Measurement in Education*, *2*(4), 289-296.

Tiselius, E., & Sneed, K. (2020). Gaze and eye movement in dialogue interpreting: An eye-tracking study. *Bilingualism: Language and Cognition*, *23*(4), 780-787.

Turner, A. (1987). The propositional analysis system Version 1.0 (Tech. Rep. No. 87-2). Boulder, CO University of Colorado, Institute of Cognitive Science.

Turner, A., & Greene, E. (1977). *The construction of a propositional text base*. University of Colorado Press.

Tzou, Y. Z., Eslami, Z. R., Chen, H. C., & Vaid, J. (2012). Effect of language proficiency and degree of formal training in simultaneous interpreting on working memory and interpreting performance: Evidence from Mandarin-English speakers. *International Journal of Bilingualism, 16*(2), 213-227.

United Nations. (2001). United Nations General Assembly, fifty-sixth session A/56/188, 16 July 2001. Remote interpretation. Report of the Secretary-General. *The Interpreters' Newsletter*, *11*, 163-180.

Van der Schoot, M., Vasbinder, A. L., Horsley, T. M., Reijntjes, A., & van Lieshout, E. C. (2009). Lexical ambiguity resolution in good and poor comprehenders: An eye fixation and self-paced reading study in primary school children. *Journal of Educational Psychology*, *101*(1), 21-36.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Van Gerven, P. W., Paas, F., Van Merriënboer, J. J., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, *41*(2), 167-174.

Vanroy, B., Tezcan, A., & Macken, L. (2019). Predicting syntactic equivalence between source and target sentences. *Computational Linguistics in the Netherlands Journal*, *9*, 101-116.

Viezzi, M. (2013). Simultaneous and consecutive interpreting (non-conference settings). In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of Translation Studies* (pp. 377-388). Routledge.

Volpe, M. (2015). Notizentechnik und persönlichkeit. Ein versuch mit 12 sudierenden im masterstudiengang konferenzdolmetschen [Master's thesis, Johannes Gutenberg-Universität Mainz]. Mainz.

Wan, H. 万宏瑜, & Qian, Y. 钱仪雯. (2020). 基于眼动数据的英汉视译文本干扰研究[Textual interference in sight translation based on eye-tracking evidence]. *Technology Enhanced Foreign Language Education*, *193*(3), 88-112.

Wang, B. & Zou, B. (2018). Exploring language specificity as a variable in Chinese–English interpreting. A corpus-based investigation. In M. Russo, C. Bendazzoli & B. Defrancq (Eds.), *Making way in corpus-based interpreting studies* (pp. 65-82). Springer.

Wang, B., & Lei, M. (2009). Interpreter training and research in mainland China: Recent developments. *Interpreting*, *11*(2), 267-283.

Wang, J. 王家义, Li. D. 李德凤, Li, L. 李丽青, & He, Y. 何妍. (2018). 译员阅读加工的认知机制——基于眼动追踪技术的实证研究[Cognitive mechanism underlying reading tasks among interpreters—An empirical study based on eye-tracking]. *Technology Enhanced Foreign Languages*, *182*, 84-90.

Wang, S. (2018). Note-taking in consecutive interpreting: A comparative study of professional and novice interpreters. *Translation Quarterly*, *88*, 45-83.

Wang, W. 王文宇, Zhou. D. 周丹丹, & Wang, L. 王凌. (2010). 口译笔记特征与口译产出质量实证研究[An empirical study of note-taking characteristics and output quality in interpreting]. *Foreign Language World*, *4*, 9-18.

Weaver, C. A., Ill, & Kintsch, W. (1990). Expository text. In R. Barr, M. L. Kamil, P. B. Rosenthal, & P. D. Pearson (Eds.), *Handbook of reading* (pp. 230-245). Longman.

Wen, H., & Dong, Y. (2019). How does interpreting experience enhance working memory and short-term memory: A meta-analysis. *Journal of Cognitive Psychology*, *31*(8), 769-784.

Weng, Y., & B. Zheng. (2020). A multi-methodological approach to studying time-pressure in written translation: Manipulation and measurement. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *19*, 218-236.

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology, 51,* 59–91.

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, 34(1), 205-223.

Wickens, C. D. (1976). The effects of divided attention on information processing in tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(1), 1-13.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical*

Issues in Ergonomics Science, *3*(2), 159-177.

Wickens, C. D. 1984, Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63-101) Academic Press.

Wilson, D. & Sperber, D. (2004). Relevance theory. In L. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 607-632). Blackwell.

Winkler, S. (2015). *Symbole in der notation beim konsekutivdolmetschen*. [Master's thesis, University of Vienna]. Vienna.

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 483-502.

Wu, ZM. 吴钟明. (2008). *英语口译笔记法实战指导* [*Practical guide of English interpretation note-Taking*]. Wuhan University Press.

Wu, ZW. (2016). Towards understanding interpreter trainees'(de) motivation: An exploratory study. *Translation & Interpreting* 8 (2), 13-25.

Xu, H. 徐海铭, & Chai, M. 柴明颎. (2008). 汉英交替传译活动中译员笔记困难及其原因的实证研究——以国际会议职业受训译员和非职业译员为例 [Difficulties perceived by professional trainee interpreters and non-professional interpreters in note-taking when doing consecutive interpreting: An empirical inquiry through stimulated recall. *Foreign Language Research*, *1*, 122-127.

Yamada, H. (2018). Validity of note-taking for new consecutive interpreting learners: An empirical study of university interpretation courses. *Theory and Practice in Language Studies*, *8*(11), 1387-1396.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459-482.

Yu, W., & Van Heuven, V. J. (2017). Predicting judged fluency of consecutive interpreting from acoustic measures: Potential for automatic assessment and pedagogic implications. *Interpreting*, *19*(1), 47-68.

Zhang, A. 张爱玲, & Ding, N. 丁宁. (2021). 抗疫背景下的远程专业口译教学 [Distance interpreter training in the pandemic era]. *Chinese Translators Journal*, 1, 81-88.

Zhao, X. 赵雪, & Xu, H. 徐晗宇 (2018). 逻辑连词对汉英视译过程中认知负荷影响研究——一项基于眼动的研究[Effects of logical connectives on cognitive efforts in Chinese-English sight translation: An eye-tracking experiment]. *Foreign Languages Research*, *5*, 7-11.

Zhao, Y. (2022). *Take it or leave it? Notationstechnik beim Konsekutivdolmetschen Chinesisch Deutsch* [*Take it or leave it? Note-taking in Chinese-German* *consecutive interpreting*]. Franck & Timme.

Zheng, B. H., & Zhou, H. (2018). Revisiting processing time for metaphorical expressions: An eye-tracking study on eye-voice span during sight translation. *Foreign Language Teaching and Research*, *50*(5),744-759.

Appendices

Appendix 1: Transcripts of the source speech

(From: TED.com. 09/05/2019. “Music and emotion through time” by Michael Tilson)

Segment 1:

Well when I was asked to do this TED Talk, I really chuckled because you see, my father’s name was Ted. And much of my life, especially my musical life, is really a talk that I’m still having with him, or the part of me that he continues to be. Now Ted was a New Yorker, an all-around theater guy, and he was a self-taught illustrator and musician. He didn’t read a note. And he was profoundly hearing impaired. Yet, he was my greatest teacher. Because even through the squeaks of his hearing aids, his understanding of music was profound. But he was tough when it came to music. He said, “There are only two things that matter in music: what and how. That was his passion for the music. Both my parents really loved it. They didn’t know all that much about it. But they gave me the opportunity to discover it together with them. And I think inspired by that memory, it’s been my desire to try and bring it to as many other people as I can, sort of pass it on through whatever means. One day in New York, I was on the street. And I saw some kids playing baseball between stoops and cars and fire hydrants. And a tough, slouchy kid got up to bat. And he took a swing and really connected, and then he went, “Da dada da.” And he ran around the bases. How did this piece of 18th- Austrian aristocratic entertainment turn into the victory crow of this New York kid? How was that passed on? How did he get to hear Mozart? Well when it comes to classical music, there’s an awful lot to pass on. Much more than Mozart, Beethoven or Tchaikovsky. Now the raw material of it, of course, is just the music of everyday life. It’s just a design of pitches and silence and time. And the pitches, the notes, as you know, are just vibrations. They’re locations in the spectrum of sound. And whether we call them four hundred and forty per second, A, or three thousand seven hundred and twenty nine, B flat. Trust me. That’s right. They’re just phenomena. But the way we react to different combinations of these phenomena is complex and emotional and not totally understood. And the way we react to them has changed radically over the centuries as have our preferences for them.

Segment 2:

So for example, in the twenty first century (*the speaker played several notes*), Now your twenty first century ears are quite happy with this last chord. Even though a while back it would have puzzled or annoyed you or sent some of you running from the room. And in classical music we can follow these changes very, very accurately because of the music's powerful silent partner: notation. In two hundred B.C., a man named Sekulos wrote this song for his departed wife and inscribed it on her gravestone in the notational system of the Greeks. And a thousand years later, this impulse to notate took an entirely different form. In the tenth century, little squiggles were used just to indicate the general shape of the tune. And in the twelfth century, a line was drawn, like a musical horizon line, to better pinpoint the pitch's location. And then in the thirteenth century, more lines and new shapes of notes locked in the concept of the tune exactly. And that led to the kind of notation we have today. Now inspired moves of improvisation could be recorded, saved, considered, prioritized, made into intricate designs. And from this moment, classical music became what it most essentially is, a dialogue between the two powerful sides of our nature: instinct and intelligence. And there began to be a real difference at this point between the art of improvisation and the art of composition. Now an improviser senses and plays the next cool move, but a composer is considering all possible moves. Testing them out, prioritizing them out, until he sees how they can form a powerful and coherent design of ultimate and enduring coolness. But every musician strikes a different balance between faith and reason, instinct and intelligence. And every musical era had different priorities of these things, different things to pass on, different 'whats' and 'hows'. So in the first eight centuries or so of this tradition the big 'what' was to praise God. And by the fourteen hundreds, music was being written that tried to mirror God's mind as could be seen in the design of the night sky. And perhaps its tremendous intellectual perfection and serenity meant that something new had to happen -- a radical new move, which in sixteen hundred is what did happen.

Segment 3:

This, of course, was the birth of opera, and its development put music on a radical new course. The what now was not to mirror the mind of God, but to follow the emotion turbulence of man. And the chords, it turned out, were capable of representing incredible varieties of emotions. And the basic chords were the ones we still have with us, the triads, either the major one, which we think is happy, or the minor one, which we perceive as sad. But what's the actual difference between these two chords? It's either E natural, and six hundred and fifty-nine vibrations per second, or E flat, at six hundred and twenty two. So the big difference between human happiness and sadness? Thirty seven freaky vibrations. So you can see in a system like this there was enormous subtle potential of representing human emotions. And in fact, as man began to understand more his complex and ambivalent nature, harmony grew more complex to reflect it. Turns out it was capable of expressing emotions beyond the ability of words. And it turned out the symphony could be used for more complex issues, like gripping ones of culture, such as nationalism or quest for freedom or the frontiers of sensuality. But whatever direction the music took, one thing until recently was always the same, and that was when the musicians stopped playing, the music stopped. To me this is the intimate, personal side of music. It's the passing on part. It's the 'why' part of it. And to me that's the most essential of all. Mostly it's been a person-to-person thing, a teacher-student, performer-audience thing, and then around eighteen eighty came this new technology that first mechanically then through analogy then digitally created a new and miraculous way of passing things on, albeit an impersonal one. And technology democratized music by making everything available. And technology pushed composers to tremendous extremes, using computers and synthesizers to create works of intellectually impenetrable complexity beyond the means of performers and audiences. At the same time technology, by taking over the role that notation had always played, shifted the balance within music between instinct and intelligence way over to the instinctive side. The culture in which we live now is awash with music of improvisation that's been sliced, diced, layered and, God knows, distributed and sold. What's the long-term effect of this on us or on music? Nobody knows.

Segment 4:

The question remains: What happens when the music stops? What sticks with people? Now that we have unlimited access to music, what does stick with us? Well let me show you a story of what I mean by “really sticking with us.” I was visiting a cousin of mine in an old age home, and I spied a very shaky old man making his way across the room on a walker. He came over to a piano that was there. nd he balanced himself and began playing something like this. And he said something like, “Me ... boy ... symphony ... Beethoven.” And I suddenly got it, and I said, “Friend, by any chance are you trying to play this?” And he said, “Yes, yes. I was a little boy. The symphony: Isaac Stern, the concerto, I heard it.” And I thought, my God, how much must this music mean to this man that he would get himself out of his bed, across the room to recover the memory of this music that, after everything else in his life is sloughing away, still means so much to him? Well, that’s why I take every performance so seriously, why it matters to me so much. I never know who might be there, who might be absorbing it and what will happen to it in their life. That’s what drives my interest in projects that explore the potential of the new performing arts centers for both entertainment and education. And the exciting thing is all this is just a prototype. There’s just a role here for so many people -- teachers, parents, performers -- to be explorers together. Sure, the big events attract a lot of attention, but what really matters is what goes on every single day. We need your perspectives, your curiosity, your voices. And it excites me now to meet people who are hikers, chefs, code writers, taxi drivers, people I never would have guessed who loved the music and who are passing it on. You don’t need to worry about knowing anything. If you’re curious, if you have a capacity for wonder, if you’re alive, you know all that you need to know. You can start anywhere. Ramble a bit. Follow traces. Get lost. Be surprised, amused inspired. All that ‘what’, all that ‘how’ is out there waiting for you to discover its ‘why’, to dive in and pass it on.

Thank you.

Appendix 2: Instructions for the formal experiment (glossary list included)

Step 1:

Please copy the following words with the digital pen.

- All the following contents are about the interpreting task you are going to perform.
- Feel free to write more (except for copying) for the preparation of the interpreting task.

In this epic overview, Michael Tilson Thomas traces the development of

classical music through the development of written notation, the chords,

and the technology.

Speech Brief

Conductor Michael Tilson Thomas is an all-around music educator --

connecting with global audiences, young musicians and concertgoers

in San Francisco and London.

Speaker Brief

Step 2:

Please familiarize with the vocabulary.

- You can copy or write anything you want on the page.
- Please inform the researcher when you finish.

English	Chinese
Note/Notation [not]/[noʊˈteɪʃən]	n. 笔记；音符；注解；纸币；便笺；调子/n. 符号；乐谱；注释；记号法
Pitch [pɪtʃ]	n. 沥青；音高；程度；树脂；倾斜；投掷；球场
Instinct and intelligence	本能与智慧
Improvisation and composition	即兴创作与作曲
Polyphony [pə'lɪf(ə)nɪ]	n. 复调，复位音乐
Harmony ['hɑːmənɪ]	n. 协调；和睦；融洽；调和；和声

Vocabulary

English	Chinese
Chord [kɔrd]	n. 弦；和弦；香水的基调
Triad [traɪæd]	n. 三和音；三个一组；三价元素；三合会
E natural & E flat	E本位音和降E大调
Symphony ['sɪmf(ə)nɪ]	n. 交响乐；谐声，和声
Synthesizer ['sɪnθəsaɪzɚ]	n. 合成器；合成者

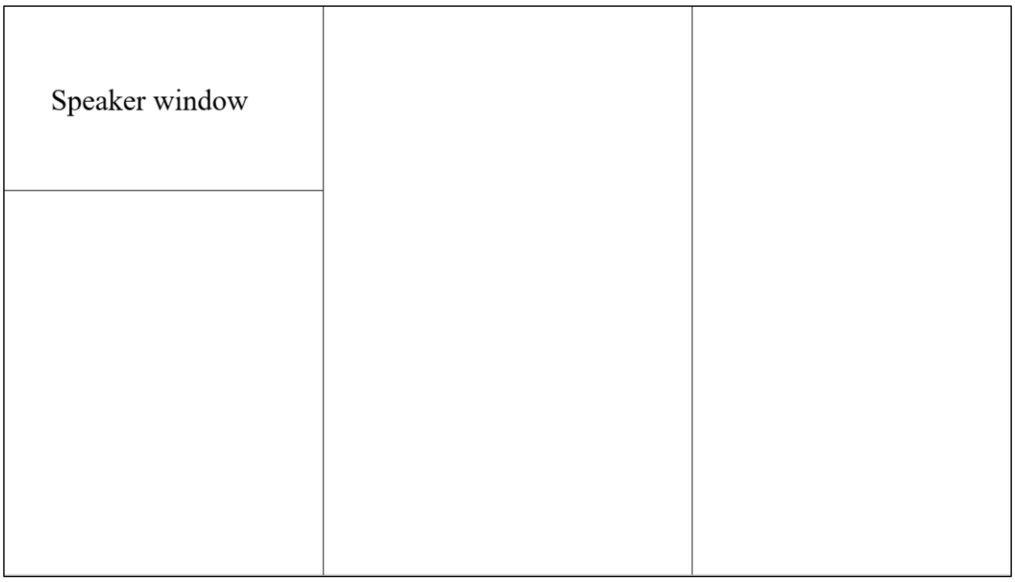
Vocabulary

Step 3:

Warm-up exercise

Step 4:

Interpreting



(In this example, the note-taking area was divided into three parts as the participant preferred)

Appendix 3: Consent form

CONSENT TO PARTICIPATE IN RESEARCH

Investigator Information & Research Purpose

You are going to participate in an experiment conducted by Huolingxiao Kuang, a research postgraduate from School of Modern Languages and Cultures at Durham University. This study will contribute to the researcher's completion of her Ph.D.

Research Procedures

This research aims to investigate note-taking behaviours in English to Chinese consecutive interpreting (CI). In this experiment, you will be asked to perform a memory test and an interpreting task. During interpreting, you will take notes with a smart pen on the tablet while looking at the computer screen where everything you note will be shown simultaneously. In each task, there are four parts, each lasting approximately three minutes. There will be no time limit for interpreting. During this process, your eye movement, pen movement and screen will be recorded. After interpreting, you will be asked to explain your notes. Then you will complete a questionnaire about your educational background, professional experience, and evaluate task difficulty. During interpreting, I will sit in the same room and help you to play the source material. After that, you will be asked to have a retrospective interview if there is anything unclear.

Privacy and Confidentiality

The result of this research will be coded in a way in which respondents' identity will not be attached to the final presentation of the study. The researcher retains the right to use and publish non-identifiable data. While individual responses are confidential, the overall result and data will be presented representing averages or generalisations of each group of participants as a whole. All the data will be stored in a secure place and only accessible to the researcher.

Participation and Withdrawal

Your participation is fully on a voluntary basis. You will get a supermarket gift card. If you would like to withdraw your consent, you are free to inform me before, during or after the experiment. If your eye tracking data is invalid due to subjective factors, you will only receive half of the participation fee.

Right as Research Subjects

You are not waving any legal claims, rights or remedies because of your participation in this study.

Questions about the study

If you have questions or concerns about the study, please contact:

Researcher's name: HUOLINGIXIAO KUANG

Department: School of Modern Languages and Cultures, Durham University

Email: huolingxiao.kuang@durham.ac.uk

Telephone: +44(0)7422588181 / +8615501123377

Giving of Consent

Appendix 4: NASA-TLX questionnaire and questions about the role of note-taking in interpreting

1. Mental Demand

How mentally demanding was the task?

0 1 2 3 4 5 6 7 8 9 10

2. Effort

How hard did you have to work to accomplish your level of performance?

0 1 2 3 4 5 6 7 8 9 10

3. Frustration

How insecure, discouraged, irritated, stressed, and annoyed were you?

0 1 2 3 4 5 6 7 8 9 10

4. Performance

How successful were you in accomplishing what you were asked to do?

0 1 2 3 4 5 6 7 8 9 10

5. Note-taking

To which extent does note-taking help you in completing the interpreting task?

0 1 2 3 4 5 6 7 8 9 10

6. Note-reading

To which extent does note recognition or note interpretation impede your interpreting?

0 1 2 3 4 5 6 7 8 9 10

Note recognition — Note interpretation

Appendix 5: Basic information questionnaire

For professional interpreters:

Background Information

Contact Information:

Name: ________________ Email: ____________________ Date: ____________

Please answer the following questions to the best of your knowledge.

1. Age (in years):

2. Sex (circle one): Male / Female

4. Have you ever received any interpreting training? If yes, for how long?

5. Have you ever received any training about note-taking? If yes, in which way did you receive that training?

6. For how many years have you been working as a professional interpreter? (You rely on interpreting as your main income)

7. What is your biggest concern in terms of note-taking in consecutive interpreting?

(and how about your students?)

For student interpreters:

Background Information

Contact Information:

Name: ________________ Email: ____________________ Date: ____________

Please answer the following questions to the best of your knowledge.

1. Age (in years):

2. Sex (circle one): Male / Female

3. What is your university and major? Which year are you in? (current year/full years)

4. Have you ever received any interpreting training? If yes, for how long?

5. Have you ever received any training about note-taking? If yes, in which way did you receive that training?

6. What is your biggest concern in terms of note-taking in consecutive interpreting?

7. Your TEM-8 score:

8. If you have taken a standardized test of proficiency for English, please indicate the scores you received for each.

Test	**Scores**				
	Total	Listening	Speaking	Reading	Writing
IELTS					
TOEFL					
Others (please specify)					